Bits and Atoms: Proactive data activism and social change from a critical theory perspective

Miren Gutiérrez

PhD Programme in Leisure, Culture and Communication, University of Deusto

Supervisors: Dr. Pilar Rodríguez and Dr. Stefania Milan

March 3, 2017

Index

Tables and illustrations

Abbreviations

CDKN Climate and Development Knowledge Network
CKAN Comprehensive Knowledge Archive Network
CSO Civil society organisation
EJC European Journalism Centre
DDoS Distributed denial of service
FFS Fossil fuel subsidies
FOIA Freedom of Information Act (US)
GANs Global Action Networks
GCF Green Climate Fund
GEN Global Editors Network
GHG Greenhouse gas
GSI Global Subsidies Initiative
GST Goods and services tax
ICT Information and communication technologies
IEA International Energy Agency
IMF International Monetary Fund
IPS Inter Press Service
IRE Investigative Reporters and Editors
ITU International Telecommunication Union
NGO Non-governmental organisation
NPO Non-for-profit organisation
NSA National Security Agency (US)
NSO National statistical organisation
OCI Oil Change International
OECD Organisation for Economic Co-operation and Development
ODI Overseas Development Institute
PAH Platform Against Evictions (in English)
SMS Social Movement Studies
TANs Transnational advocacy networks
UNSC United Nations Statistical Commission
UPM Technical University of Madrid (in English)
VAT Value Added Tax
WFP World Food Programme (UN)

Acknowledgements

I am grateful to a number of people for their key inputs into this study. First of all, to Dr. Pilar Rodríguez (University of Deusto) and Dr. Stefania Milan (University of Amsterdam), my supervisors, for their invaluable guidance and constant support. Thanks are also due to international reviewers Dr. Amaia Arizaleta (University of Toulouse) and Dr. Lisa Cuklanz (Boston College); to Guillermo Gutiérrez (Bunt Planet), who helped me with the data analysis charts of this study; to Dr. Xabier Barandiaran (University of Deusto), who provided early methodological advice for this dissertation; and to Dr. José Luis Orihuela (University of Navarra), who guided me at the very beginning of this adventure. I am grateful to my family, and especially to my parents, who offered crucial practical support.

Finally, I would like to thank the people who generously offered time of their busy schedules and valuable insights in interviews for this dissertation, including Luis Hernando Aguilar, Hisham Almiraat, Sergio Álvarez Leiva, Esteban Beltrán, Fernando Blat, David Cabo, Mar Cabra, Alberto Cairo, Jordi Carrión, Duncan Clark, Mariluz Congosto, Sandra Cucianelli, John Downing, Gustavo Faleiros, Aidy Halimanjaya, Usman Haque, Mel Hogan, Daniel Innerarity, Sebastian Mitchell, Oscar Marín Miró, Craig Mills, Angela Oduor, Oluseun Onigbinde, Matthew O'Reilly, Emma Prest, Rachel Rank, Victor Sampedro, Jan Schwochow, Jan Willem Tulp, Cesar Velasco, Daudi Were, Katie Whipkey and Shelagh Whitley.

1. Introduction

> This realm, reality, world, whatever we choose to call it, is very diverse and complex and people in society are constantly striving by trial and error to come to terms with it; to map it; to coordinate their maps of it. Living in an unmanageably large and changing society permits neither perfect mapping, nor perfect coordination maps. This means that the members of the society are constantly learning about it; both the society and its members are in a constant process of self-discovery and self-making (Jarvie 2015, 165).

Big data are a manifestation of humanity's search for a mapping system, a compass that helps us navigate these tumultuous and complex times. In fact, big data are the expression of our times, and as such they represent post-modern complexity: from the positive (i.e. proactive data activism) to the negative (i.e. massive data-enabled surveillance) (Bollier 2010, 40). The management of complexity in this multifaceted post-modern world requires an information system that is capable of encompassing and comprehending it. In pre-modern and modern times, an information system based mainly on direct communications, first, and mass media, afterwards, was sufficient. In contrast, our post-modern, complex, advanced, 'network society' (Castells and Cardoso 2005, 3-6) needs tools that are on a par with its level of sophistication; and in this context, an information system based on big data infrastructures can, among other things, help us understand it, generate diagnoses and provide solutions to social problems.

As a data activist –that is, someone that employs data as a catalyst for activism—, the main motivation behind this dissertation is remedying the absence of systematic analyses of data-based activism that can frame, explain and capture the novel alliances and tactics that make it possible, and the transformations that they have brought about. An example of the changes and

new alliances that data are helping to forge is DataKind,[1] an initiative that brings together data scientists and social organisations 'to collaborate on cutting-edge analytics and advanced algorithms to maximise social impact' on complex issues such as poverty eradication, health, human rights, the environment and cities (DataKind 2015). These are the kind of issues that data activists care about, and the kind of dynamics and communities that data can generate.

To be precise, the central object of study in this dissertation is 'proactive data activism,' defined as 'new social practices rooted in technology and data,' which 'take a critical approach to big data' and use digital technology and data politically and proactively to foster social change (Milan and Gutiérrez 2015, 125). Proactive data activism is an 'emerging' phenomenon in the 'field of action' that combines 'communicative practices,' technology and information 'at its outermost complexity' (that is, big data), and 'collective organising' (ibid., 133).

In this introduction, I review briefly the main concepts framing reactive and proactive data activism –that is, respectively, the data activism that employs data infrastructures in a reactive way as an answer to massive data gathering by governments and corporations—, as well as the data activism that employs data infrastructures proactively. These concepts include 'digital technologies,' 'big data' and 'big data society,' 'activism,' social practice' and their role in 'social change' (basically, in logical consecutive order, from data to social change through activism). I also lay out the relevance of this study, my initial hypotheses and research questions, set up my objectives and how I am going to achieve them, and finally describe the different parts of this dissertation. What follows is meant to be a definitions toolkit, which frames the research and outlines general and preliminary concepts that are mobilised in this research.

[1] DataKind is funded by donations and supported by companies like Microsoft, Tableau or IBM. The business models of some of these companies have been criticised as discriminatory (Sandoval 2014); nonetheless, DataKind's objectives and project remain good examples of positive proactive data activism.

1.1 Definitions

Data activism is activism that utilises mostly data infrastructures, and in some occasions big data, as an enabling method. But what exactly are big data? Big data are examined more exhaustively later, but for the purpose of this dissertation, big data are considered a profusion of digital objects and user-generated online content stemming from users' digital activity, mass-interception and metadata,[2] as well as the *datafication* of human and non-human activity, which are so large in volume, can be processed with such velocity, are so varied, have so much economic and development potential when rendered useful, and show so much accuracy and complexity that can be considered really *big* and, therefore, can only be extracted, managed, analysed using new infrastructures and methods.[3] Big data, through machine learning, can produce new insights, new knowledge and 'new forms of value in ways that change markets, organisations, the relationship between citizens and governments, and more' (Mayer-Schönberger and Cukier 2013, 5). The analysis and visualisation of big data often produces 'small data,'[4] which can be easily comprehended and handled by humans, as opposed to machines, as we will see later on.

Big data are the result of the development of digital technology, which is originally based on an Enlightenment mathematical concept formulated by the 17th century mathematician

---

[2] Mayer, Mutchler and Mitchell –who have studied the privacy properties of telephone metadata— consider that metadata are even more sensitive than data, since they are subject to less protection. 'Since 2013, a stream of disclosures has prompted reconsideration of surveillance law and policy. One of the most controversial principles, both in the US and abroad, is that communications metadata receives substantially less protection than communications content' (2016, 5536). These authors conclude that telephone metadata are 'densely interconnected, can trivially be reidentified, and can be used to draw sensitive inferences' (ibid.). This matter will be tackled later.

[3] Elaboration by the author based on numerous sources, which are detailed later together with a more complete definition.

[4] In any case, 'small data are being made more big data-like through the development of new data infrastructures that pool, scale and link small data in order to create larger datasets, encourage sharing and re-use, and open them up to combination with big data and analysis using big data analytics' (Kitchin and Lauriault 2014, 463).

Gottfried Wilhelm Leibniz, who proposed a binary computing system. This innovation is the foundation of modern computing.[5] Digitised information is recorded in binary code of combinations of the digits 0 and 1, called *bits* –the substance of the virtual world—, which denote words and images. Digital technology permits for vast amounts of information to be condensed on ever smaller and more mobile, even wearable, devices. The digitalisation also accelerates data transmission velocities. Henceforth, a 'big data infrastructure' is to be understood as a digital organisation and structure that enables data sharing, management, storage, analysis and usage, which can include software and the platforms that allow the transfer and employment of data (Russom 2013, 4-20). Digital infrastructures have completely transformed how people learn, communicate, work, campaign, and engage in politics and social life.

Post-modernity –the historical period when big data infrastructures emerge— is described with more detail later to contextualise the emergence of the big data society. Post-modernism is to be understood hereafter merely in chronological terms as the economic, social and cultural state which is said to exist after modernity. As Harvey notes, 'no one exactly agrees as to what is meant by the term, except, perhaps, that post-modernism represents some kind of reaction to, or departure from, modernism. Since the meaning of modernism is also very confused, the reaction or departure known as post-modernism is doubly so' (1991, 7).

Activism is to be recognised henceforth as an endeavour, individual or collective,[6] designed and planned to foster, obstruct or guide social, political, economic or environmental

[5] Of course, others would later contribute to the progress of computing and digitalization: George Boole, who invented Boolean algebra in the 19th century; Charles Babbage and Ada Lovelace, who pioneered in computing; John Bardeen, William B. Shockley and Walter H. Brattain, who invented the transistor in the 20th century, to mention a few.

[6] Bennet and Segerberg distinguish between the logic of collective action and 'the less familiar logic of connective action based on personalised content sharing across media networks' (2012, 3). The emphasis

change with the final intention of making improvements in society or correcting social injustice. Activism and campaigning can combine a series of online and offline elements and strategies in the following areas (from the most proactive to the most reactive): research and knowledge creation around social, political, economic or environmental problems; communication and dissemination of information and knowledge, and campaign messages (i.e. writing letters to decision makers, sending press releases to media organisations, giving speeches); advocacy and direct lobbying (i.e. one on one negotiations and influence) (Barker 1995, 12); mobilisation (i.e. boycotts, rallies, street marches, strikes, hunger-strikes, sit-ins, hashtag advocacy,[7] 'digital activism'[8] and 'smart mobs')[9] (Joyce 2010, ix; Jordan 2004, 55; Rheingold 2005); and direct action, conceived as a prominent and public deed designed to bring attention to or reveal a problem, demonstrate a possible solution, obstruct a negative outcome or highlight an alternative to an already taken decision (Jordan 2004, 53). Direct action can include civil disobedience, and

---

is in the 'personal' or individual connected versus the collective. Although this distinction is helpful to understand how world protests have been organised based on personal action frames that spread through social media, I am not going to apply to the study of data activism.

[7] 'If enough people share campaign messages on Twitter or Facebook with a hashtag, the message may *trend* and receive even more attention via app interfaces that highlight trending topics for users and widely followed major media outlets that amplify these memes' (Frauzel & Collinson, 2015).

[8] Joyce distinguishes between 'cyber-activism' and 'online advocacy' –that is, 'the use of the internet for activism'— and 'digital activism,' which can include 'all instances of social and political campaigning practice that use digital network infrastructure' (2010, ix).

[9] A smart mob, contrary to a regular mob, behaves efficiently because of its exponentially increasing network links, enabling people to connect to information and others, and to perform a task allowing a form of social coordination. 'Smart mobs consist of people who are able to act in concert even if they don't know each other. The people who make up smart mobs cooperate in ways never before possible because they carry devices that possess both communication and computing capabilities. Their mobile devices connect them with other information devices in the environment as well as with other people's telephones. Dirt-cheap microprocessors are beginning to permeate furniture, buildings, and neighbourhoods; products, including everything from box tops to shoes, are embedded with invisible intercommunicating *smartifacts*' (Rheingold, 2002).

activists often put themselves in danger of being arrested in order to make a political statement.[10] Activism can also be expressed through art. In this study, data-enabled activism is examined.

The internet, mobile technologies and lately big data infrastructures have totally transformed activism. Earl and Kimport claim that the internet poses two key advantages to activism: clearly cheaper costs for the coordination, mobilisation and participation in protest; and a decreasing requirement for protesters to be physically together in one space in order to act collectively (2011, 197-201).[11] Big data infrastructures are also changing activism in a radical way. The rise of big data infrastructures and the mass collection of data by governments and corporations initially caused the phenomenon known as 'reactive data activism,' which comprises the practices of resistance to the threats to civil and human rights that derive from digital intrusion. The bypassing of surveillance and control in activism is not a new phenomenon and predates datafication, for example via strict control of membership (della Porta and Diani 2006, 203). But reactive data activism stands out as an activity that is enabled by socio-technical artefacts[12] to work with data or to shelter online interactions from intrusion and automatized collection (Milan and Gutiérrez 2015, 127). The infrastructure that made it possible emerged among elitist *technorati*, in the periphery of society, 'in the realm typically associated with grassroots activism and civic engagement,' to become more accessible to the common person, 'rapidly evolving from a peripheral, elitist form of (reactive) activism to a diffused one, whereby

[10] At Greenpeace, for example, *activist* is a label that only applies to elite militants that take part in direct action tactics and get specialist training, for example, in order to be able to board vessels at sea in extreme weather conditions. Ryan and Runswick Cole differentiate between advocacy and activism, which they consider implies a higher degree of involvement that the former. Activists describe themselves as such take a leadership role in the campaigning organisation or effort, and expend a substantial amount of time and energy on the cause (Ryan and Runswick Cole 2009). I am not going to make this distinction.

[11] See examples of massive cyber activism and mobilisation via mobile phones in Annex II.

[12] Although originally *socio-technical artefacts* referred to physical objects, in the context of this dissertation the definition will include also software and data infrastructure. This is a notion that comes from science and technology studies (known as STS).

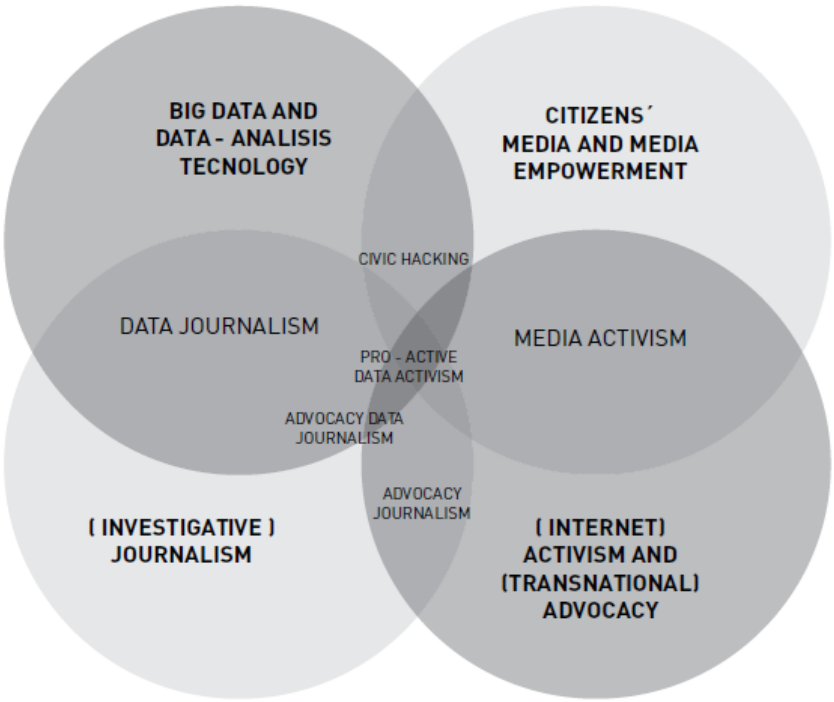
also non-skilled users take part in the game' (ibid., 127). In other words, data activism involves a series of practices 'at the intersection of the social and the technological dimension of human action,' with the aim of either 'resisting massive data collection' (reactive data activism) or 'actively pursuing the exploitation of available data for social change' (proactive data activism) (ibid.).

Proactive data activism –the kind of *propositive* activism that is enabled by data– is currently being harnessed to understand, analyse and solve a range of entwining, complex social problems, from climate change and biodiversity loss to inequality and human rights abuses (Hogan 2015). Armed with data-based knowledge and digital technology –as 'extensions of man' (McLuhan 1964, 103)–, proactive data activists get involved in awareness campaigns, advocacy, mobilisation and lobbying tactics, in order to change policy or practice,[13] and in this process, not only their lifeworlds,[14] but also data activists and their behaviours are transformed as well. Proactive data activism results from the juncture of several areas of the human action, including what Habermas has called 'communicative action' (1984, 278-279). Campaigning, awareness raising, the visualisation of information and other associated interactions ultimately seek a 'common understanding' in a process of the 'communicative action' (ibid., 11). Other relevant areas of action related to data activism include endeavours associated with processes of information and data communication. These areas overlap and mingle, as Figure 1 shows.

Figure 1: Areas of human action and proactive data activism

[13] More accurate distinctions will be made later on in this dissertation about other key concepts such as 'advocacy,' 'campaigning,' 'lobbying,' 'direct action' and 'awareness raising,' among others.

[14] The lifeworld 'is formed from more or less diffuse, always unproblematic, background convictions,' or sets of beliefs taken for granted. 'In their interpretive accomplishments the members of a communication community demarcate the one objective world and their intersubjectively shared social world from the subjective worlds of individuals and (other) collectives' (Habermas 1984, 70).



Source: (Milan and Gutiérrez 2015).

Based on this conceptualisation, in this study I explore the core of proactive data activism, as well as the immediate peripheral areas, such as other social uses of data (i.e. data journalism), or other change promoting activities (i.e. transnational activism), in order to dip into their commonalities and dissimilarities. I also classify data-enabled action according to its purposes and strategies, as many concrete cases cross the lines dividing these areas.

In this dissertation, I focus specifically on data-enabled proactive activism. The subject of the case study is Ushahidi platform, which, based on data infrastructures, is specialised in supporting humanitarian efforts in cases of crisis, armed conflict, emergency or disaster. By doing so in such innovative ways, Ushahidi has generated a new paradigm in humanitarianism. Paraphrasing Aguado and Martin, it shows how, in times of crisis and heightened systemic inequality, a group of people endeavour to find ‘alternatives to how things came to be,’ and through ‘hard questioning of the *status quo*,’ create ‘better methodologies of analysis and

critique in precisely those terrains that are conductive to social change' (2016, 165). Ushahidi does exactly that in the field humanitarianism.

Humanitarianism in this study is not be considered as a philosophy that puts humans in the centre, but *stricto sensu* as the efforts involved in providing material and logistic assistance to people, and responding to disasters and crises, mediating conflicts, including armed conflicts, and undertaking peacekeeping operations. The first objective of any humanitarian operation is saving lives, alleviating suffering and maintaining human dignity in the face of catastrophe.[15] However, increasingly humanitarian effort is being directed to saving also livelihoods and enhancing the resilience[16] of vulnerable communities (Humanitarian Policy Group 2006). Apart from the loss of lives, many of the people affected by disasters get injured, become homeless or vulnerable to transmissible diseases, experience violence and famine, and at least suffer the disruption of normal life in the aftermath, including the interruption of schooling, work, and limited, or lack of, access to food, services and a home. These are increasingly taken into account in humanitarian assistance. Humanitarian action has evolved into a highly specialised endeavour, with specific policies (i.e. the Hyogo Framework for Action addressing disaster reduction) and specialised institutions dedicated to tackling issues such as emergency response, disaster risk reduction (DRR) and management (DRM), and conflict management. In fact, the organisations dedicated to the betterment of humanity could be grossly divided in two types (whether intergovernmental, governmental or non-governmental): those responding to immediate emergencies, whose action is typically short-term assistance until other institutions replace them,

[15] The principles of humanitarian work are formally enshrined in two General Assembly resolutions. The initial three principles –humanity, neutrality and impartiality— were endorsed in General Assembly resolution 46/182, adopted in 1991. General Assembly resolution 58/114 (2004) added independence as a fourth key principle underlying humanitarian action.

[16] In this context, resilience is the ability to bounce back to the original situation, or even improve it, after being hit by a disaster [Elaboration by the author based on ODI (2011)].

and those organisations dealing with the underlying socioeconomic factors which may have led to or caused a crisis (i.e. working on development issues or on climate change).[17] The lines dividing them, though, are not so straightforward in reality. For the latter, a key concept is long term *change*. Campaigning organisations usually seeks long term changes in policy and practice. As said, campaigns can include a repertoire of tactical tools. Some of these organisations are good at direct action (i.e. Greenpeace) others are better at close-door negotiation (i.e. WWF). But what is a campaign? Basically, a campaign is a collection of actions, combined strategically during a certain period of time, that are implemented sequentially or simultaneously in order to change either policy and practice, or public attitudes and behaviour, or in order to change decision making processes so that affected communities benefit from it, or else in order to empower affected communities so they can change decisions or situations that affect them.[18] In all cases, the sustained effort of a campaign is about achieving some positive social change, not just raising awareness about a problem. Campaigns and campaigning organisations can be classified by their target, their geographic scope, their tactics, the desired outcomes and their cause (Chandler 2012, 4). But campaigning organisations are all characterised for outlining a 'theory of change.' Ideally a result of a collective effort with stakeholders, the ToC identifies the

[17] The first group includes organisations such as Action Against Hunger (AAH), World Food Programme (WFP), CARE, Caritas Internationalis, Catholic Relief Services (CRS-USCC), Doctors Without Borders, International Federation of Red Cross and Red Crescent Societies (IFRC), International Rescue Committee (IRC), Oxfam, Refugees International, Save the Children, UN High Commissioner for Refugees (UNHCR), UN Office for the Coordination of Humanitarian Affairs (OCHA), and US Committee for Refugees (USCR) (ReliefWeb 2015). The second group includes environmental organisations, such as Greenpeace, Friends of the Earth International (FoEI), WWF, SEO BirdLife International, Climate Justice Now!, Conservation International, EarthWatch, Forest Stewardship Council, Global Environmental Facility (GEF), International Union for Conservation of Nature (IUCN), Intergovernmental Panel on Climate Change (IPCC), UN Environment Programme (UNEP). And International development organisations include: HOPE International Development Agency, Oxfam, Plan International, World Vision International, World Accord - International Development Agency and UN Development Programme (UNDP), among many others. Some can be included in several spaces.

[18] Elaboration by the author based on *Advocacy and campaigning* (Chandler 2012, 2), which uses 'advocacy' and 'campaigning' practically as synonyms, as I do.

problem; then, detects the decision-makers who have the power to change, improve the situation and even solve the problem; finds the entry points in order to reach them and describes how to access them; lays out a plan with concrete steps within a timeline; envisages and spells out the expected wider effects of the plan; lays out the co-benefits and other externalities; and finally states the final goal (Hivos 2015).

Many organisations are hybrid: some development organisations have humanitarian programmes on the ground, and the other way around. In the context of the financial crisis that started in 2007, for example, some humanitarian organisations in Spain understood that they also had to campaign for causes in order to subsist (Gutiérrez 2012, 225-230). Although many of them share this hybrid nature, in general humanitarian organisations react to disasters, get in and out quickly, responding to an emergency, with the objective of saving lives and increasingly also livelihoods. They normally have to work in collaboration with local, regional or national authorities in order to be able to work undisturbed and assist the *victims*,[19] even in much polarised situations such as armed conflicts. That is why they do not usually have a position about those conflicts, or at least they do not trumpet it, and do not usually point fingers at culprits. Conversely, campaign organisations generally work towards social change by campaigning in favour of a good cause or against a case of abuse. They can go from being reactive (i.e. direct action or protest) to proactive (i.e. lobbying and negotiating), depending on how open the door for cooperation is with decision makers in each situation. When dialogue is not possible, this type of organisation can be forced to work from outside the concerned community. This is a key concept, as it is the reason why organisations have been divided into

[19] Many human rights organisations no longer talk about 'victims' but about 'rights-holders,' in order to highlight that *victims* have rights and are entitled to protection by virtue of being human. Although being an important distinction, it is not related to the core issue of this dissertation, therefore I will not use 'right-holder' when I am referring to someone who is enduring a crisis, abuse, disaster or violence. However, I am aware about the negative connotations around the term.

insiders and outsiders (Chandler 2012, 4), and even *beyonders* (Milan and Hintz 2013, 22), and it will be useful to observe data activist organisations as well.[20] Table 1 shows the differences between both types.

Table 1: The difference between humanitarian and campaigning

	Humanitarian	Campaigning
Perspective	Responsive	Generally pre-emptive
Time span	Short-term goals	Medium-long term goals
Objectives	Deals with emergencies	Deals with causes
Goals	Saving lives and livelihoods	Social change
Links with authority	Has to work with authorities	Can work from outside
Protest	Doesn't point fingers at culprits	Can oppose and point fingers at culprits

Source: Elaboration by the author based on Chandler, Chopard and Lusser, and Dijkzeul and Moke (Chandler 2012), (Chopard and Lusser 1997), (Dijkzeul and Moke 2005).

Social change is what campaigning organisations working on causes seek. I is understood hereafter as a significant transformation in the social configuration of a given society or part of it, which may include changes in social institutions, organisations, structures and norms, in social behaviours and practice, or in social relations. In data activism, social change starts with an

[20] 'Institutional policy arenas provide windows; that is, temporary opportunities for civil society activists and advocates to raise their policy concerns and influence the political environment' (Milan and Hintz 2013, 20). How activists react to such opportunities rests on their cultures and values, as well as on whether they perceive them as opportunities for gains or as threats (ibid.). From this point of view, campaigning organisations have been classified traditionally as using 'insider' or 'outsider' strategies (ibid.), although the reality is complex. Some campaigns require more coercive methods than others, even if the organisations leading them are best at insider strategies. This happens when, faced with closed doors for negotiation, organisations are sometimes forced to use outsiders' tactics. In other cases, alliances are the best option to make the most of out the different participant organisations' skills. One example is the environmentalist platform formed by Friends of the Earth, Ecologistas en Accion, WWF and Greenpeace in Spain. The alliance benefits from Greenpeace's and Ecologistas en Accion's capacity for mobilisation and direct action, as well as from WWF's and FoE's political clout. Other campaigns evolve with time, i.e. 'Armas bajo control.' This campaign was launched in 2004, and is still active, taking many forms over time, with periods of 'peaceful' negotiation and of tumultuous arm-twisting. Marta Arias, from UNICEF, defined it as the 'mutant' campaign, as it evolved from focusing on just asking for transparency in Spanish weapons exports to demanding that weapon exports were prohibited to countries in turmoil to avoid their use against civilians. This campaign grew over its 'quick wins,' including for example, the transformation of Spain from a producer of cluster bombs to a signatory of the Oslo Convention of 2008, which prohibits them.

empowerment process[21] that can potentially alter authority dynamics, placing muscle in the hands of groups or individuals of the 'low levels' of the power structure (Linstone and Mitroff 1994, 100). Through this process, a data activist becomes a fire-stealing Prometheus, 'the Greek demi-god, who saw that the gods had fire and regular people did not. He saw this injustice, so he stole the flames and taught any other to make fire. Stealing the fire we think it is a metaphor for the democratisation of technology' (Milan 2014b).

Social change can happen in social policy or practice, or both. Social policy refers to a set of ideas, norms or a plan of what to do in particular situations 'that has been agreed to officially by a group of people, a business organisation, a government or a political party' (Cambridge 2015).[22] 'Social practice' is used in this dissertation with two distinct meanings. The first one connotes an action that 'promotes development, social cohesion and the empowerment and liberation of people' (International Federation of Social Workers 2014). Social practice in that sense can be individual or collective, and involves an intervention in the context of a 'lifeworld' –that is, the socio-cultural structures, moral and practical principles, discourses and world-views, attitudes, personal value commitments, technical and other embodied skills, or roughly what Bourdieu called 'habitus' or socialised subjectivity (2004, 49), or what Habermas called the 'lifeworld' (1984, 70).[23] The International Federation of Social Workers (IFSW) says that this

[21] As is explored later on, 'within the communication sphere, empowerment is the process through which individuals and groups, by taking active part in the actions that reshape their communicative processes, exercise control over their communication resources' (Milan and Gutiérrez 2015, 128).
[22] Habermas would put it in relation to a set of accepted 'social norms' (1996, 50).
[23] Socio-cultural structures, moral and practical principles, discourses and world-views, attitudes, personal value commitments, technical and other embodied skills – roughly what Bourdieu called 'habitus' or socialised subjectivity, and corresponding with what Habermas, following the phenomenological tradition, called the 'lifeworld.' Many of these structures are not flexible or easily altered. Bourdieu defined *habitus* as a system of socially established configurations that, as structured and structuring structures, are the engendering and unifying principles of the collective practices and ideologies characteristic of a group of agents (Bourdieu 2004, 97-119). For Habermas, culture, society and

type of social practice's guiding values are 'social justice, human rights, collective responsibility and respect for diversities,' and social practice should 'address life challenges and enhance wellbeing' (2014). Social practice can also be directed to change policies 'in order to achieve the goal of social and economic justice' (Cummins, Byers and Pedrick 2011, 2). Most social practice, in this sense, involves knowledge generation and theorisation, since, before implementing any programme, an organisation usually sets up its mission, defines what it understands by notions such as 'social justice,' 'collective responsibility,' 'wellbeing' (i.e. institutional theorisation), and establishes what its programmes' objectives are and how it will implement them (i.e. intervention practices theorisation). The other meaning of 'social practice' used in this research is more general and refers to simply actions that happen in a social context or have social impacts. For example, enabled by digital technologies, social practices can happen simultaneously in the 'space of flows' (Castells 2010b, 407-453), as they are not anchored in real three-dimensional physicality –the world of atoms or the 'space of places' (ibid.)—, but in the *virtuality* of the digital world or the world of bits. The meaning of the expression will be clear from the context.

---

personality are the three components of the lifeworld. Habermas tried to integrate structure and agency with his theory of the 'colonisation of the lifeworld' by the official systems (the economy and the state), which leads to a perversion of public life (Ritzer 2005, 6). Habermas fears the advance of what he calls the 'system on the lifeworld,' the realm of formal rationality (using Weber's terms). The colonisation of the lifeworld –the realm of substantive rationality—, therefore, involves an increase in formal rationality to the detriment of substantive rationality. This restraint leads to a growing differentiation between culture, society and personality. 'The system is an external perspective that involves the way an outside actor not involved in society would view things. Although the system is rooted in the lifeworld, it has its own characteristics separate and distinct from the lifeworld. As these components grow and become strengthened through the maintenance-oriented actions of the lifeworld, they become more distant from and impose themselves on the lifeworld. This distancing, in turn, weakens the functions of the system (corresponding to those of the lifeworld) of cultural reproduction, social integration, and personality formation. Overall, the move toward agency-structure integration in Europe has become what many there consider the major issue in modern social theory' (ibid.).

Another relevant issue here is the use of human emotions in campaigning and how social change happens. For the same reason that there is no collective action without emotions (della Porta and Diani 2006, 13), no successful campaigning can ignore them either. Castells, for example, attaches great importance to emotions when studying the dynamics of power, political behaviour and social movements (2009, 202-210). And in fact, the study of emotions is increasing in political, media, social movement and communication scholarship. Goodwin, Jasper and Polletta proclaim emotions are back in the study of politics, after a period dominated by rationalistic, structural and organisational approaches (2001). There is no doubt human faces sell products, convey campaign messages and mobilise people.[24] According to Carreras, social change can only happen when there are high doses of widespread indignation and emotion, together with an attractive proposal for change and a viable plan of implementation –abiding by the formula C=I+M+P, where C is change, I is indignation, M is a model for change and P a plan (Carreras 2012). Emotions are essential to invoke some campaign elements, such as the creation of comunities and the mobilisation of *crowds* needed in crowdsourcing. Goodwin, Jasper and Polletta justify the study of emotions such as anger, indignation, fear, disgust, joy and love in the research on politics and social protest. 'The prospect of unexpected and sudden changes in one's surroundings can arouse feelings of dread and anger. The former can paralyse, but the latter can become the basis for mobilisation' (2001, 16). Emotions are also prominent in the study of how social media propagate *sentiment*. This means data activists face a challenge, because numbers are poor substitutes for emotions and empathy in campaigning. As Cairo puts it: 'I am just very sceptical to the idea that data visualisation is a medium that can convey (or even care about conveying) or increase *empathy'* (Zer-Aviv 2015). Empathy –the capacity to place oneself in

[24] In 2010, a team of psychologists of the University of Hertfordshire in Edinburgh planted 240 wallets in order to see what contents made people return a wallet. They found out that 88% of wallets with baby photos, 54% of wallets with dog photos and 48% of family photo portraits were returned (Bliss 2010).

another's place— is needed in campaigning. 'It was not the statistics and charts that reframed the global debate about the Syrian refugee crisis. Rather it was the horrifying images of baby Aylan Kurdi's body washed up to shore' (ibid.). However, if a campaign is not sustained by data and research, it can be easily dismissed as something anecdotal for 'bleeding hearts.' Data analysis brings the big picture into matters.[25] Cabra, interviewed for this dissertation, tells how she left television for data journalism precisely because data allow journalists to move away from the individual to find more encompassing stories. Policy making can also benefit from this approach, as Bloom states.

> Our public decisions will be fairer and more moral once we put empathy aside. Our policies are improved when we appreciate that a hundred deaths are worse than one, even if we know the name of the one, and when we acknowledge that the life of someone in a faraway country is worth as much as the life a neighbour, even if our emotions pull us in a different direction (2014, 3).

Not everything is so clear cut, nonetheless. Zer-Aviv notes the emotion transmitted by a dramatic visualisation of gun deaths in the US by Periscopic[26] —whose lemma is 'do good with data'—, which is set on a black background, with many visual elements that make it a gruesome display, including a score window showing the 'amount of stolen years' (2015). 'The empathetic and poetic framing of this chart is definitely the source of its power and effectiveness' (ibid., 3). Whitley, interviewed for this dissertation, talking about the campaign against massive subsidies

[25] In fact, big data analysis and visualisations allow for a macro and micro approaches, as Tufekci notes, since they are the 'telescopes, microscopes and scalpels' of the 21st century (Tufekci 2012, 6).
[26] See guns.periscopic.com/?year=2013 [accessed on May 4, 2016]. Periscopic is a for-profit organisation explicitly dedicated to convert data into 'visual and interactive experiences that allow people to empathise and understand' (Periscopic 2016). They also have 'a very strong desire to work with organisations involved in environmental issues, human rights, peace and equality, education and the arts' (ibid.).

funnelled towards the fossil fuel industry, says: 'we mixed things. Did you see the animation?[27] I got emotional watching that, but I am bought into the issue. This is not about crying people and things dying; it is not dramatic. But… we used lots of infographics that are quite striking visually, and rhetorically quite powerful.' The conclusion is that both data and emotion need each other in effective data activism. The doses of empathy and data should be strategically combined in a campaign. There is also a call for more data and statistics by institutions like the UN,[28] which is trickling down on other spheres of action. 'This renewed interest by multilateral organisations and governments alike for data and statistical evidence is likely to alter the advocacy strategies of a myriad of civil society organisations, who will move away from moral arguments into statistical evidence in advocating for their causes' (Milan and Gutiérrez 2015, 5).

Some notions developed by the social movement studies (SMS) help root data infrastructures in today's field of action and understand the shifts in power data promote in the sphere of the organised civil society,[29] as well as explore the trans-border nature of information and data flows. I will lean on key ideas from SMS in order to observe data activism. This does

---

[27] See odi.org/publications/10058-production-subsidies-oil-gas-coal-fossil-fuels-g20-broken-promises [accessed on April, 16].

[28] As we will see when the uses of data by governmental and intergovernmental agencies are examined, UN Department of Economics and Social Affairs Under-Secretary-General Wu Hongbo called for more data in development. 'Statistics is shaping our understanding of the world,' he said addressing the UN Statistical Commission (UNDESA 2014).

[29] The term civil society refers to 'the wide array of non-governmental and not-for-profit organisations that have a presence in public life, expressing the interests and values of their members or others, based on ethical, cultural, political, scientific, religious or philanthropic considerations. Civil Society Organisations (CSOs) therefore refer to a wide of array of organisations: community groups, non-governmental organisations (NGOs), labour unions, indigenous groups, charitable organisations, faith-based organisations, professional associations, and foundations' (World Bank 2013). The World Bank says that the number of international NGOs was reported to have increased from 6,000 in 1990 to more than 50,000 in 2006. CSOs have also become significant players in global development assistance, with the Organisation for Economic Cooperation and Development (OECD) estimating that, as of 2006, CSOs provided approximately US$15 billion in international assistance (ibid.). Civil society is also known as 'the third sector,' different from government (public) and business (private). Hereafter, I will use 'civil society,' 'NGOs' and 'third sector' as synonyms, in spite of the fact that they are not perfect substitutes.

not mean I am embracing the entire theoretical body of SMS work; only that I am using some of its concepts as a heuristic tool to examine data activism.

With the purpose of studying a new 21st century phenomenon, I draw also upon the *old* knowledge of 20th century critical social theory as well, with a particular emphasis on the work of Habermas. His theories of communicative action and the public sphere are particularly useful to illuminate reflections on the notion of data-based activism. I also rely on post-Habermasian criticism to make those notions more applicable to the post-modern phenomenon of data activism. Finally, the concepts of mass self-communication and the network society are also fundamental here (Castells 2009). Additionally, some concepts emerging from other branches of sociology that deal with collective and communicative action are cherry-picked as well, when needed. A heterogeneous body of literature is taken into consideration in the initial part of this research, as I utilise concepts that help me understand proactive data activism, the society that has created it and its impacts. Its heterogeneity only reflects the complexity and confluence of trends from which this new phenomenon has emerged. The methodological section substantiates more in detail these choices.

The next section outlines the reasoning behind this piece of research. More specific definitions are included across this dissertation whenever required.

1.2 Relevance and objectives

The starting point of this dissertation is an article entitled *Citizens' media meet big data: The emergence of data activism* (Milan and Gutiérrez 2015), in which we set the scene, link data activism to the tradition of citizens' media and lay out the fundamental questions surrounding the new phenomenon of data activism. This research also brings into play my personal experience as

an activist with Greenpeace and as a Research Associate at The Overseas Development Institute (ODI) of London.[30]

This dissertation is based on two premises that I use as points of entry to the field. The first one is that, as said, our complex society needs to utilise big data infrastructures with the purpose of both understanding itself and responding to its challenges. On the one hand, big data are a by-product of our network society, a society of highly connected individuals and machines, through mobile and Bluetooth technology, conventional broadband internet and the Internet of Things, which, combined, have knitted a compact coat of connections over almost everybody and everything in this planet. On the other hand, for the first time, through big data infrastructures, social science is closer than ever to painting a full picture of society.[31] My position is far from Anderson's (2008), as it will be discussed, but there is no doubt that big data allow us to analyse our society as comprehensively and quickly as never before (i.e. not based on surveys and what people think or say they do; but on what people really do).

The second premise is that big data infrastructures can allow tackling social problems from a democratic, participatory perspective, as data can empower people and be a social equalizer (but also create inequality and be a tool in oppressive practices, as it will be discussed). Democracy is all about equal access, opportunities and voice, as well as real participation.

---

[30] See odi.org/publications/10459-western-africas-missing-fish-impacts-illegal-unreported-and-unregulated-fishing-and-under-reporting [accessed on July 18, 2016]. At ODI, I have published data-based reports, including setting up a big data-based project, which involved analysing terabytes of data on illegal or irregular fishing activities in the territorial waters of western Africa for a year. Based on the new knowledge acquired in the process, ODI published a report in 2016, supported by Kofi Anan's led African Progress Panel (APP), and launched an international advocacy campaign aimed at changing policy and practice in the area of fisheries, and at generating social change for millions of people whose livelihoods and nutrition depend on fishing in western Africa. That is, a full-blown proactive data activist project.

[31] Anderson calls it 'the end of theory' (Anderson 2008). Rather than testing a theory by analysing relevant facts in a deductive way, new data analytics and artificial intelligence seek to attain insights 'born from the data' in an inductive way. This approach has far-reaching consequences to how knowledge is produced, business conducted and governance enacted. 'Petabytes (of data) allow us to say: "Correlation is enough"' (ibid.).

Precedents are, for example, civic hackers.[32] Empowered by code and data, they can match up with bankers, government officials and businesspeople by spotting and exploiting weaknesses in computer networks and systems. Occupy Wall Street's slogan 'We are the 99%,' possibly the first democratic movement to have data as a political motto.[33] Equipped with data, in 2011, the Occupy Wall Street mobilisation marshalled tens of thousands in New York around issues such as income inequality and unbalanced wealth distribution. The strength of the 'better argument' (Habermas 1984, 25) got people out of their homes and summoned them in Zuccotti Park. The global protest wave known simply as Occupy had some significant impacts on legislation. For example, in 2011, in Spain, 15M protesters[34] forced politicians reform the mortgage law. And in the US, the demonstrations inspired a group of senators to call for the constitution to be amended to prohibit corporate money in campaign funding (Ted Deutch 2011). Occupy Research and Occupy Data[35] –spinoffs of the Occupy movement—are a symptom of how central data and information is for this movement. The initiatives seek to support the protests by gathering, analysing and visualising data relevant to economic inequalities, explicitly addressing 'the politics of data, questioning data availability and the associated mainstream narratives,' interrogating data agency, inviting people to produce their datasets and to shape issues,

[32] 'Civic hackers are technologists and open-data activists who engage with datasets to address challenges relevant to their community' (Milan and Gutiérrez 2015, 133). Civic hackers seek to improve institutional output and democratic governance by means of software and data, requesting, processing, generating, modelling and contesting data (Milan and van der Velden 2016).

[33] The origin of the slogan is commonly believed to come from an article by Stiglitz entitled 'Of the 1%, by the 1%, for the 1% (2011), in which he criticises the inequality present in capitalist US. Nobel Prize winning economist Paul Krugman retorted that Stiglitz had set the bar too low and that the slogan should be 'We are the 99.9%,' since the income of the top richest 0.1% of the US population had risen 400% between 1979 and 2005, citing a Congressional Budget Office report (Krugman 2011).

[34] The origins of the 15M Movement (Movimiento 15M) can be traced to social networks such as Democracia Real YA. The demonstrations began on May 15, 2011, and expanded across Spain and beyond. Millions took part, at one time or another, in demonstrations across the whole country.

[35] See occupyresearch.net/ [accessed on September 28, 2016]. Occupy Research is an open space for research focused on the Occupy Movement to share ideas, research methods, tools and datasets.

ultimately generating alternative epistemologies in a datafied environment (Milan and van der Velden 2016).

However, when big data infrastructures are employed with undemocratic purposes, they inevitably become tools in the hands of the powerful, coveting the perpetuation of power relations and privileges. Gangadharan argues, for example, that vulnerable groups systematically are discriminated against by new practices and data techniques of online surveillance, and that, while online surveillance activity affects all users, members of excluded communities become potentially more vulnerable to the harmful effects of data-based snooping technologies, (2012). Another example is the British Regulation of Investigatory Powers Act 2000 (known as RIPA), an act of the UK parliament setting powers of public bodies to carry out surveillance, including the gathering of private data (Kampmark 2015). For David Anderson, this framework is 'undemocratic, unnecessary and – in the long run – intolerable' (Naughton 2015). RIPA, among other powers, enables certain public bodies to request that internet service providers (ISP) offer access to a customer's communications surreptitiously to facilitate surveillance. The dilemma 'security versus freedoms' is deeply embedded in data activism –that is, the current state of the social contract between citizens (the governed) and rulers (the governing) is directly linked with how big data infrastructures are being utilised by different actors. However, the emphasis here is data infrastructures' galvanising and emancipating powers –that is, how data infrastructures can be used in proactive activism. Social uses of big data (as well as small data) have big implications and deserve to be studied.

WikiLeaks is perhaps one of the most prominent cases of data activism. WikiLeaks uses cryptographic technologies to provide a secure system for whistle-blowers to leak information. Combining data infrastructures and journalistic techniques, WikiLeaks also publishes

information based on the analysis of the data provided by unidentified sources. For example, in 2012, WikiLeaks released millions of classified documents (leaked by former intelligence analyst Manning), including US diplomatic cables and reports from the Afghan and Iraqi wars, revealing thousands of unreported deaths (Brevini, Hintz and McCurdy 2013, 3; Tavernise 2010).[36] The disclosure spurred public outrage. This dissertation would provide an explanation of how an initiative such as WikiLeaks fits in the field of action.

More concretely, the research questions (RQs) that are a starting point for this dissertation are: RQ1 What is proactive data activism? RQ2 How has proactive data activism emerged? RQ3 What are people's data-based tactics to foster social change? RQ4 How can proactive data activism be effective? In the data gathering exercise included in this dissertation, the idea is to let the data speak for themselves in order to generate a model for effective proactive data activism at the end. This deserves a clarification. All observation is subjective and can be arbitrary, and data gathering is both enabled by my own experience as a data activist immersed in some global struggles and limited by my own shortcomings, and in no case aims at being positivist or normative. Other research questions have emerged from this piece of research, which are addressed in the section dedicated to areas for further research.

The objectives behind the present research questions are:

[36] The WikiLeaks cable revelations had social consequences too, to the point that some consider them the catalyst of the 2010-2011 Tunisian uprising. In an article for *Foreign Policy*, Dickinson says: 'we might also count Tunisia as the first time that WikiLeaks pushed people over the brink' (2011). Other accounts, such as one by *The New York Times*, indicated that 'the protesters… found grist for the complaints in leaked cables from the US Embassy in Tunisia, released by WikiLeaks, that detailed the self-dealing and excess of the president's family' (Kirkpatrick 2011). It is believed that the Tunisian revolution then spread to the Middle East, and Tunisia is considered Arab Spring's birthplace (Lynch 2014).

- O1: To describe in-depth a new phenomenon in the field of action in the era of the network society –that is, proactive data activism— as an autonomous phenomenon in its own right, different from other sorts of activism and data applications;
- O2: To illustrate how big data infrastructures, used politically and proactively, are strengthening democracy by integrating sophisticated digital technologies in social practices; and
- O3: To describe how proactive data activism can be effective and propose a viable model for proactive data activism, as well as to suggest new research questions.

Ultimately this study is intended to result in a contribution to the practice of activism in general. As seen in this dissertation, data activism does not spring out of nothing: it is part of an evolving social practice and landscape, and owes a debt of gratitude to preceding social phenomena. Data analysis and visualisation are increasingly becoming part of many activists' skillset. And currently, they are being utilised in combination with other elements available in advocacy efforts. In order to understand how data activism emerges and describe how it works within activism, this study is rooted in both academy and practicum, based on the principle that practice can generate new knowledge, which in turn informs practice, and so on. As 'critical theory seeks practice informed by theory' (Froomkin 2003, 872), I also endeavour to both test theory in practice and explore the theoretical frameworks of practical applications of big data infrastructures, especially in the case study included in this dissertation, which is focused on how the Ushahidi organisation uses its platform to both assist emergency *victims* in disasters and, in doing so, empowers them and produces social change. The purpose is to both narrow the study down to the particulars of a relevant case, as well as providing wide-ranging insights that can be useful to the general practice of activism.

1.3 Roadmap

This dissertation is divided in nine sections designed as concentric circles, organised from the most peripheral to most central. The first one contains the introduction; the second, the methodology; the third depicts the socioeconomic and technological context in which big data emerge and defines them; the fourth describes the social uses of big data infrastructures; the fifth dwells on the uses of big data infrastructures by individuals and groups of the civil society; the sixth describes in-depth proactive data activism as a data-enabled social practice; the seventh is dedicated to the case study –Ushahidi—as an class phenomenon to illustrate proactive data activism; the eighth is devoted to the analysis of the data; and finally the ninth includes the conclusions and a proposed model for effective proactive data activism, as well as suggestions about areas for further study.

Big data are 'social facts'[37] of the first magnitude and their social uses happen in a culture, value system and socioeconomic context. The theoretical foundations of this dissertation lay in the socioeconomic, cultural and political landscape of 'our datafied times' (Kennedy, Poell and van Dijck 2015, 1) and networks, in which the 'big data society' thrives today. This dissertation starts with a historical perspective on how the ideal Habermasian public sphere has been renewed and updated with big data infrastructures, in a way that make alternative digital public spheres for data activism feasible. The social, scientific and technical transformations that resulted from the emergence of big data and data activism are reviewed, beginning with key notions that have their origin in the Enlightenment, including computing, evidence-based knowledge as the engine of social progress, the scientific method, the consent between

[37] In sociology, social facts are the cultural norms, values and structures which go beyond the individual. Durkheim argued that social facts 'consist of manners of acting, thinking and feeling external to the individual, which are invested with a coercive power by virtue of which they exercise control over him' (1982, 52). Durkheim proposes two central theses, without which sociology would not be a science: its specific objects of study are social facts; and it must apply a recognised objective scientific method and avoid subjective judgment, bringing it close to the exact sciences (ibid.).

government and governed, democracy, the market economy and communitarianism. It does not intend to be a comprehensive historical review, but a selective account of the crucial notions and social facts that have made big data infrastructures possible today, and how these social facts relate to the advent of data activism. A technical definition of big data and a reflexion of the differences between data, information and knowledge, and how big data are not perfect or complete (Gitelman 2013) follow. The different theoretic debates, ideologies and epistemological interpretations that surround big data are gathered here too. Concretely, I review current epistemological approaches to big data, and how big data and big data infrastructures represent an important challenge to sociology itself as 'disruptive innovations'[38] that are reconfiguring how we see the world and rocking the foundations of some essential modern ideals, such as the scientific method, the paradigms of sociology and the social contract between citizens (the governed) and rulers (the governing).

The subject of the case study selected for this dissertation is an organisation called Ushahidi, which is an example of a proactive data activist organisation. It was created in Africa, where mobile technologies currently flourish (The Economist 2015a, 1; Hesse 2010; International Telecommunication Union 2015). Communications infrastructures –especially mobile phones— combined with big data infrastructures and interactive mapping make the Ushahidi platform technically possible. That is why this initial part devotes some space to explaining the proliferation of mobile technologies and the social changes they are producing in our society. Besides, mobile technologies are both key for the birth of the network society (Castells 2010) and the space of flows –which allows for simultaneous social practices such as

[38] A 'disruptive innovation' is 'a process by which a product or service takes root initially in simple applications at the bottom of a market and then relentlessly moves up market, eventually displacing established competitors.' This theory developed by Christensen was initially focused on businesses, but over time has been used to explain kinds of disruptive innovations in other sectors (Christensen, 2015; Markides, 2006).

data activism, without 'territorial contiguity' (Castells and Cardoso 2005, 3-6)—, and also the source of vast amounts of data about ordinary citizens' behaviour, to the point that the can be considered 'digital census-takers' (The Economist 2015a).These data, brought together with others collected via the Internet of Things (IoT)[39] and digital surveillance, are the basis of the 'computational politics' and the renaissance of the 'post-modern *panopticon'* (Foucault 1984, 1995; Tufekci 2012) –what Braman calls the *panspectron* of the 'informational state' (2007, 314). 'Computational politics' and the 'post-modern *panopticon*' are explored here too, as they are the reason behind the emergence of reactive data activists, who can be considered proactive data activists' forerunners (Milan and Gutiérrez 2015, 127). Massive data collection and exploitation by governments and corporations has been interpreted as a threat to civil rights by reactive activists and, as a consequence, some of them have become 'reactive data activists' (ibid.).

However, big data can also offer new opportunities for collective action,[40] emancipation, and social and political change –what Milan and I have called 'proactive data activism' (2015). Some reactive data activists may have experienced a similar journey as the reactive organisation Anonymous, which went from being purely confrontational to being more proactive in a *de-radicalisation* process (Gorenstein-Massa 2013, 63), capturing the power of collective work, volunteering, the open movement ethos and the hacker culture. Examples of this process towards

[39] The Internet of Things (IoT), or 'those uniquely identifiable objects that can be tracked' (International Telecommunication Union 2014, 173) resulted from the interconnection of identifiable embedded computing devises within the internet infrastructure. It is called the Internet of Things becomes this internet infrastructure interconnects 'things' (i.e. televisions, refrigerators, clocks, ovens, cars and urban furniture).

[40] The terminology around collective action is unclear. However, 'collective behaviour' could be considered the largest category including all social collective phenomena; 'social movement,' a sub-category of collective behaviour; and 'collective action' would mean a certain event of action. Although there is no consensus, in some cases the concepts are used almost as synonyms. Della Porta and Diani distinguish between 'collective behaviour' –focused on observable actions— and 'collective psychology' –focused on motivation of individuals (2006, 12).

proactiveness range from organisations such as HackForGood[41] and Code for Africa[42] to the certification as Ethical Hacker[43] and the project called 'Vidas Contadas,'[44] headed by Labarga, a hacker transformed into a proactive data activist focused on uncovering buried truths left behind by the Spanish Civil War.

The globalisation process and the new forms of exclusion that globalisation and big data have brought about are explored too. These reflexions are relevant as they are the backdrop of proactive data activists' campaigns and projects, which generally try to address injustice and exclusion, or work in situations of injustice and vulnerability, by creating digital public spheres that facilitate equal participation and decision making. The post-modern phenomena of individualism and consumerism –as by-product of a globalised, capitalist world whose values clash those of proactive data activism— are also briefly discussed here as well.

Once the socioeconomic and technological context and the main concepts are addressed, the data collection and analysis include a look at who is a subject in big data generation, namely governments, corporations and organisations. Governments and companies have been the first to realise the potential of big data infrastructures, and have driven innovation efforts. However, they have also created a coat of surveillance over every citizen in this planet. Some examples of both useful and manipulative use of big data infrastructures by governments and corporations are included here. Within civil society, journalists are among the first to make use of these infrastructures. This is an especially relevant sector for data activism, since many data activists,

---

[41] See hackforgood.net/ [accessed on August 23, 2016].
[42] See codeforafrica.org/ [accessed on August 23, 2016].
[43] See eccouncil.org/programs/certified-ethical-hacker-ceh/ [accessed on August 23, 2016].
[44] An expression that plays with the meaning of *contadas*, which denotes both 'told,' 'revealed' as well as 'tallied.' The project publishes information about the people disappeared and killed during the Spanish civil war and the mass graves left behind (Vidascontadas 2013).

in the absence of good and relevant journalism, have filled the gap by producing data journalistic content (i.e. Civio[45] in Spain). This section delves also on the uses of big data infrastructures within ONGs and non-profits to produce data-enabled activism. Previous analysis (Milan and Hintz 2013, 22; Milan and Gutiérrez, 2015, 127), and the empirical observation of concrete cases allow for an initial classification in reactive and proactive data activists. I include here an expansion of the description of proactive data activist included in Milan and Gutiérrez (2015): from the critical theory perspective, proactive data activism is a communicative action, with empirical, hermeneutic and critical elements, which generate digital public spheres for action; from a journalistic perspective, it is similar to investigative journalism because of it analytical nature, to advocacy journalism for holding a political viewpoint, and to citizen journalism for its all hands on deck, rand and file character; from an alternative media perspective, it involves the politics of the quotidian and changes the relationship between citizens and data, placing them at the vanguard of production; from an international relations perspective, it is transnational and collaborative; and from a social movement perspective, it works towards long-term norm change and gathers people around a specific use of technology to form a shared identity, communicate, mobilise and act. Proactive data activism takes many shapes, and from observation, I have classified the organisations that practice it in four types: the skills transferers (focused on transferring data or social skills, forging opportunities for alliances and generating digital platforms for data activism); the catalysts, which mainly fund this type of projects; the activist organisations that produce data journalism; and the actual data activists, most of them geoactivists. Being the most numerous group, the geoactivists deserve their own space here. They engage largely in creating alternative mapping systems and geoactivism –the most extended form of proactive data activism observed so far. This last group is also divided in

[45] See civio.es/en/ [accessed on August 25, 2016].

several subgroups, depending on how they specialise in gathering and generating data: they can a) rely on whistle-blowers for data; b) resort to open-source datasets; c) use crowdsourcing tools to generate their own data; d) turn to appropriating data; and e) get data from primary research that can be datafied and analysed, or generate them (i.e. via sensors). Here I also define key related concepts, including 'map,' 'geoactivism,' 'crisis mapping' and 'activist mapping,' provide examples of these activities and the action repertoires, and describe the usual technological platforms employed in proactive data activism and the associations that have been fashioned by data. In this section, I also test and illustrate the validity of big data infrastructures to build a democratic society, to conclude that they have empowering properties and can be instruments of emancipation 'through the diffusion of scientific knowledge' (Habermas 1984, 147) and creation of alternative digital public spheres. Habermas hypothesised that humans had a fundamental interest in emancipation, or the liberation from unnecessary constraints and distortions. As a result of data infrastructures' potentialities, data activism is indeed changing the way both ordinary people and organisations advocate for social change, raise awareness, mobilise, influence, lobby and foster social change, and free themselves from constraints. This section includes categorisations of concrete cases, a table showing groupings and subcategories and a chart to illustrate pairwise relations between types of organisations.

The selected case for further study is Ushahidi. Using communications and data infrastructures and visualisations, it literally facilitates the mapping of conflicts and disasters. Ushahidi deployments inform in-depth about crises, empowering *victims* and transforming them in reporting witnesses,[46] generating awareness, aiding humanitarian efforts, and changing humanitarian practice on the ground as well as the way it is theorised about it. I examine the

[46] In Braman's classification of data activists' roles, witnesses providing data could be considered *sensors*, as opposed to *routers, servers, network providers* and others (2016).

repertoires of action of Ushahidi as well. Ushahidi is a proactive data organisation that utilises crowdsourced data, which is an especially empowering process for the people who contribute the data. This last property –data agency— is examined in detail as well here. The case study includes a look at the origins and current situation of the organisation, including how it works and how it is funded; the same polyhedral analysis employed to look at proactive data activism in general is used here to examine the case of Ushahidi. The case study ends with an examination of the controversies that surround Ushahidi and how the organisation has learnt from mistakes and evaluates its performance. A few Ushahidi deployments have experienced problems in their ability to generate communities –the basis of their data gathering method— and to sustain already established initiatives; to set up processes of verification of the data and protection of reporters; to engage other key actors and adapt to local realities; and to introduce and implement evaluation and monitoring systems. These cases are scrutinised and conclusions are drawn and incorporated in the model for an effective proactive data activism included at the end of the dissertation.

The analysis of the data follows, summarising the interpretations of the data gathered in the empirical observation of relevant cases, the interviews and the case study. It includes a table showing most of the organisations and initiatives mentioned, and how several attributes of data activism (i.e. works in alliance with others) apply to them specifically, and with how much intensity (from an absence of the attribute in question to 3, maximum score). I then visualise this in a chart showing correlations between attributes to find, for example, that organisations that focus on transferring skills so others can develop data projects do often generate connections and networks among organisations.

The final section includes the conclusions of the study, as well as proposed model for effective proactive data activism. The dissertation ends with a series of related interesting areas that have emerged during this study, but could not be tackled because they would have meant a diversion from the main subject, for further research.

2. Methodology and ethical considerations

   Both as a candidate and now as President, Donald Trump has championed discriminatory policies. Among the groups his policies have most frequently targeted are Muslims and immigrants… Companies must make sure that they do not help governments, or other private actors, violate fundamental human rights. This means that data brokers, data analytics companies or others in the industry were to provide data or services that could be used to construct a "Muslim registry" or carry out mass deportations, they would be in breach of their human rights responsibilities (Center for Democracy & Technology et al. 2017).

This dissertation aims at demonstrating that, beyond a heuristic conceptual tool, data activism is indeed a social phenomenon its own right, separate from other forms of data usage or activism. In practice, data-enabled proactive social initiatives respond to some of today's most pressing social challenges (i.e. Shahada's data-based humanitarian platform is employed in the face of disaster and conflict), and enhance democratic participation by allowing people to generate data, and create their own alternative public spheres and empowering them (i.e. the data analysis that is both at the heart of the political messaging and the mobilisation of the Occupy movement).

The theoretical prism that I have used to observe proactive data activism is critical thought, among other relevant branches of sociology, for its usefulness in explaining communicative action and the public sphere. When needed, I resort to some concepts originated in other disciplines, such as SMS, which I have employed as tools to understand some relevant aspects of how collective action happens in data activism. The methodology that I have used to gather data and analyse them is qualitative. Semi structured interviews grant access to the interviewees' subjectivity in order to understand the world from their points of view (Packer

2010, 8), and to discover their *lifeworld* before venturing any scientific explanations. However, some qualitative methods are used to visualise the data, as detailed later on.

Most translations included in this dissertation have been done by the author, while the originals are included in footnotes.

Next I detail the disciplines employed to observe big data social uses –and more precisely, proactive data activism—, their social context and their impacts, the qualitative methods employed in the data analysis section, and the important ethical considerations around the social uses of big data infrastructures.

2.1 Disciplines

Given the importance for proactive data activism of diverse factors, concepts and dimensions, such as communication and media, knowledge, journalism, democracy, power dynamics, ICTs, data infrastructures, networks, activism, social movements and collective action, several disciplines that do not usually interact have been mobilised in this dissertation. As della Porta notes, studies around these topics have tended to be siloed: Research on democracy has focussed on representative institutions; research on the mass media has isolated media 'as a separate power,' concentrating on 'the technological constraints and opportunities for communication;' while social movement studies 'have mainly considered democratic characteristics as setting the political opportunities social movements have to address and – more rarely – looked at the constraints that mass media impose upon powerless actors' (2013, 86). But recently, 'some opportunities for reciprocal learning and interactions developed, prompted by some exogenous, societal changes as well as disciplinary evolutions' (ibid.). I intend to profit from this new encompassing approach to explore how democracy, media, technology, activism and data infrastructures intersect in proactive data activism.

Some concepts found in SMS are helpful to root data-based advocacy in the field of action. In spite of the fact that SMS usually plays down the importance of media and technologies –two key elements in proactive data activism—, SMS is a valuable tool to study data-based action, especially how proactive data activists use of unconventional means (della Porta and Diani 2006, 7,159, 180). Social movement organisations have been evolving since the 1970s by the continuous expanse of new forms of protest and action, as well as the constant refocussing on contentious, moving-target issues (i.e. in recent times, from the global questions tackled by the World Social Forum to the more domestic concerns of the 15M). SMS have evolved with them as well, beginning with the Marxist perspectives looking for a unique and global social movement determined by class to refocus attention towards strategic behaviour, structures, collective identities and individual commitment (ibid., 58-152). Since the 1990s, the underlying mechanisms that arise between structure and action have attracted the interest of SMS. The globalisation, the financial crisis starting in 2007 and the strengthening of neoliberal policies, together with the mass self-communication phenomenon, the omnipresence of ICTs, and the emergence of big data infrastructures present stark challenges to traditional conceptions of representative democracy, political action and protest. At the same time, new opportunities appear that offer alternative visions and venues for action, and are providing the grounds for novel insights. The so called Arab Spring, and the Indignad@s (15M) and Occupy movements represent a test to SMS, which are now challenged to explain how waves of mobilisations spread from one country to another. Data activism proposes another challenge as a subject, and its study benefits from the concepts proposed by SMS, in spite of being neither a social movement nor aimed mainly at dissent. From this perspective, proactive data activism can be seen as an instance of collective action deeply rooted in and caused by existing conflicts, which then

deploys unconventional means –data infrastructures used critically in alternative ways – for social change.

Habermas's work and critical theory have been useful to look at the social context in which big data have thrived (1971, 1984, 1996, 2006; Adorno 2004). The 'critical' in critical theory is interpreted in this dissertation as a critique of power, inequality, ideology and political economy (Froomkin 2003; Gitelman 2013; Kitchin 2014; Leonelli 2014; Trottier 2014). Helpful has been Calhoun's analysis of Habermas and his public sphere (1992). Besides, I have taken into account Webster's appreciation of Habermas's public sphere, even if he deems it (too) 'ideal' (2006, 163-168); Flyvbjerg's comparison of Habermas' 'ideal theory' with Foucault's and Nietzsche's 'real rationality' (2000, 4-6); Roderick's lectures on Habermas and the main representatives of the Frankfurt School (2012); and Terry's views on Habermas's perspectives on knowledge, communication and discourse (1997). In order to grasp the Frankfurt School's main relevant concerns, I have also resorted to Hegel's phenomenology (2008); Marx's theories on social classes (Marx and Engels 2000); Kant's views on the Enlightenment (1994); and Foucault's musings on 'daring to know' and the idea of the *panopticon* (1984, 1995). In addition, I have considered other scholars' attempt to capture, from a critical theory perspective, the worlds of bits, code, big data and their social uses, including Froomkin's work on *cyberspace* (2003); Berry's description of code's democratising properties and the 'computational turn' (2014, 49; 2011, 8); Fuchs's critique of social media as a false emancipatory public sphere (2013); Burns's theories on data-enhanced digital humanitarianism (2014a, 2014b); and finally, Berendt's, Büchler's and Rockwell's reflections on whether data scientists should cooperate in enhancing data infrastructures' capabilities to spy on people (2015). I have also reviewed Plage's interesting comparison of two key concepts for this dissertation: Castell's 'network society' and

Habermas's 'public sphere,' which this author equates as two of 'the most significant metaphors' to explain the media in contemporary society (2009, 1).

At the beginning of this dissertation, I travel back to the Enlightenment in order to find the origins of big data and today's society, considering the foundations of the social contract theory (Hobbes of Malmesbury 1651) and of the classic free market economy (Smith 2007); the substance of modern democracy and what causes it (Alemán and Kim 2015; de Tocqueville 2002), as well as the birth of the concept of emancipation (Israel 2006, 5); the emergence of 'the media' (Briggs and Burke 2002; Castells 2009), and their associations with modern democracy (Alger 1998).

Big data analysis can generate information, and information can be the basis of knowledge. To explore knowledge and how knowledge has been valued across time, I have delved into the different approaches to the knowledge pyramid, looking also at its hierarchy and components, starting with data and how they can be transformed into information, and ending with how valuable knowledge is in today's big data society. Through Ackoff (1999), Bellinger, Castro and Mills (2004), Cleveland and Eliot (1934), Habermas (1971), Liew (2007), Prensky (2001), Trottier (2014), Zeleny (1987) and Zins (2007), I discover that data are not valuable, useful or meaningful by themselves, that 'raw data' is an oxymoron (Gitelman 2013), that big data analysis can render meaningless data and metadata –previously considered *garbage*— into usable data, that information is data analysed, and that knowledge is information in action.

As technology's influence on society is crucial for this dissertation, I explore some concepts laid out by technological determinism approaches and the 'doctrine of progress' (Anderson 2008; Innis 1986, 2008 [1951]; McLuhan 1964, 8-9; Ogburn 1949; Rheingold 2002). However, technologic innovation does not explain all social change, and social and cultural

circumstances in which a technology develops, as well as how people use them, can be decisive. Therefore, I also borrow ideas from critics of determinism and followers of cultural materialism (Huesemann and Huesemann 2011; Fischer 1994; Williams 1990).

The globalisation and technology advancement have left many people behind, even if some reject the idea of the digital divide (Rossetto 1998). I also review the theories of several authors who have explored this issue to come up with a classification of three types of post-modern marginalisation: Exclusion today can be technological, economic and structural (Gangahar and Jones 2007; Krugman 2011; Lessig 2006; Oxfam 2015; Prensky 2001; Rosling 2011; Stiglitz 2011; Touraine 1971). I also revise the notion of the 'space of flows,' which plays a central role in Castells's vision of the network society (Ball, Haggerty and Lyon 2012; Berry 2011; Castells 2010; Gibbs 2014).

The big data society is also closely watched. I resort to Bernays (1947), Braman (2007), Berry (2011), Brevini, Hintz and McCurdy (2013), Bauman (2007), Bauman and Lyon (2013), Deibert (2010), Lyon (2014) and Tufekci (2012), who describe and explore massive data gathering by governments and corporations, the dangers of big data put to undemocratic uses, the thick coat of connections and surveillance that involves individuals in the 'big data society,' the birth of a modern *panopticon*, and how big data infrastructures can discriminate against certain communities. This is a key notion, since it is what triggers the dawn of reactive data activism.

Bauman (2007), Lipovetsky and Serroy (2009), Melucci (1996a; 1996b) and Putnam (2001), among others, help me explore the globalised and individualistic society in which big data and proactive data activism emerge. I draw from Harris and my own explorations on climate change as an example of today's local challenges posed by global phenomena (2014), and how transnational advocacy organisations deal with them. In this section I start exploring

transnational advocacy networks (known as TANs), an issue revisited in the section devoted to the social uses of big data.

Data activism is a form of social organising around information and technology, and social movement, technology and network studies serve to capture some key dynamics of data-based organised collective action as a social construct (della Porta and Diani 2006; Melucci 1996a, 40; Milan 2013; Schradie 2015).

I have looked into the different political perspectives and epistemologies that exist around big data as well, from different perspectives. In this regard, scholars could be roughly distributed between two extremes: the techno utopians and the big data sceptics. Among the ones enthused by the techno utopia, mostly keen capitalists, are the proclaimers of the so called 'Californian ideology' (Rossetto 1998). This enthusiasm has prompted Anderson to proclaim 'the end of theory' (2008). The sceptics, however, include all sorts of authors –from liberal to communitarian— that caution us about the dangers of big data infrastructures (Bernays 1947; Braman 2006; Cruickshank and Chis 2015; Fuchs 2013; Tufekci 2012). I position myself in the middle. In fact, the *positive* and the *negative* of technology are exposed by the two types of data activism. 'Reactive data activists' engage in unconventional and disruptive tactics, such as resistance, and in practices, such as hacking and encrypting, in order to defy massive data gathering and snooping (the negative). They are the pioneers of data activism, and precede 'proactive data activists.' Meanwhile, the latter get involved in data-based awareness campaigns, advocacy, visualisations of critical information, mobilisation and lobbying for social change (Bollier 2010, 40; Milan and Gutiérrez 2015, 127). Both types of data activism work towards norm or practice change as part of the process in which society takes action upon itself (Touraine 1977, 53). The strategic options and actions that data activists plan and carry out to pursue social

change in international power structures can be studied, also, from the perspective of international relations or global studies (Fraser 2014; Tysiachniouk 2012). After all, many of the mentioned actions are not only international but in some cases global in nature, and cannot be tackled only from a local perspective only.

Meanwhile, code –being the technical environment that supports the internet, or 'cyberspace' (Lessig 2006, 5), as well as the language of big data infrastructures— is central to data activism and a worthy object of study *per se*. 'Code is speaking us (pervading and formatting our action)' (Cox, McLean and Berardi 2013, 9). Code can alter the power relations, place muscle in the hands of individuals of the 'low levels' of the power structure (Linstone and Mitroff 1994, 100):[47] the *techies* and *hackers*. Therefore, science and technology studies can help discover the technological environment that enables and sets boundaries to 'proactive data action' and to power associations related to it, as well as to delve into the emancipatory properties of code (Leonard 1999; Lessig 2006; Cox, McLean and Berardi 2013). Useful are too Bijker's studies into the 'technological' and the 'social,' tackled concurrently with the aim of proposing a comprehensive political theory that embraces sociotechnical change and how society deals with the technologies (1995). Melucci's assessment of the 'collective identity' –or the common sense of belonging to the same group— of social movements is key too (1996a; 1996b). And Milan's studies on the 'emancipatory communication practices' and technologies of social movements have also illuminated this work (2013; 2014a).

[47] 'A crisis may instantaneously change the (power) structure (of an organisation) from a hierarchical to a flat organisation in which the power of previously low levels is enhanced and equalised with that of previously high levels' (Linstone and Mitroff 1994, 100). In the case of data-enabled activism, data are the 'fire' that equips normal people with equalising power.

2.2 Qualitative methods

Paradoxically, I have chosen mostly qualitative methods in order to tackle a new phenomenon heavily based on technology and quantitative data. But the qualitative method enjoys 'a tremendous growth' in use and acceptability (Hammersley 1990, 10), and it specially suitable for this study (Hand 2014, 1). This research uses 'unstructured' forms of data collection, including interviewing and empirical observation, and employs descriptions and verbal data rather than quantitative measurements and statistical analysis (ibid.). My main interest is to capture and explore what proactive data activists think, say and do, and whether they are a class apart as I have conjectured. Qualitative data are 'data enhancers' (Vaterlaus and Higginbotham 2011) in that they boost data to make it possible to see aspects of the object of research that might otherwise be neglected. Therefore, they are particularly appropriate to the study of relatively unexplored realms of human action such as data activism. Thick descriptions, depth and detail of social practices, as well as better answers about the how of digital activism, are needed to understand social action (Schradie 2015). The objective in this study is to go beyond a mere description, beneath the surface, and to study a phenomenon that still needs to be described, explained and analysed in an interpretative manner. Nonetheless, I use some quantitative methods to analyse the data gathered as well.

In other to achieve the objectives established in the introduction, and in relation with the research questions (RQs), the methods employed to study this phenomenon comprise:

- A critical account of the epistemologies and discourses around big data infrastructures and their social uses (RQ1 and RQ2)
- An empirical observation, web content and document analysis, and analysis of relevant cases of data activism (RQ1, RQ2, RQ3 and RQ4)
- Participant observation of relevant cases of data activism (RQ1, RQ2, RQ3 and RQ4)

- Semi-structured and in-depth interviews (RQ1, RQ2, RQ3 and RQ4)
- A case study (RQ2, RQ3 and RQ4)

While observation and interviews require an active role of the researcher in gathering and creating the datasets, documents and website content are primary sources produced by activists without intervention. Both are equally considered and inspected in this study. These methods of data collection provide accounts of organisations and initiatives from the point of view of their protagonists and their social practices. Interviewing, empirical and participant observation and the study of a case provide deep descriptions when they are combined.

The methodology includes semi-structured interviews with activists and experts around the world, selected based on their visibility and their experience or role. The sampling criteria employed to decide on the interviewees include any of the following: a) interviewees must depend on and deal with data and data infrastructures in their work; b) they often define themselves as data journalists, even as data activists; and c) they must use data analysis and/or visualisations with a social purpose, either to assists in a humanitarian emergency or to explore the causes of a social wrong. They include journalists using data or devoted data journalists with social interests; data scientists, analysts, *techies* with a social purpose; researchers employing data with social objectives; activists using data or staunch data activists; and developers of data platforms and infrastructures for social change.

The idea is to capture the interviewees' points of view. How accurately is my portrayal of them is a matter of discussion, since social scientists tend to attribute interpretations of the world to the people whose actions they are analysing (Becker 1996, 56). I do that, paraphrasing Becker, by talking to them; by verifying their utterances with information that has been produced without my intervention and is available online; and by giving them questionnaires, which prompt them

to say what they think (ibid., 57). The purpose of the interviews is to capture what the activists and data users believe they do, and their stated reasons and interpretations of their social world, as well as their evolution in time and what changes data usage has introduced in their lives. Based on the interviews, through a systematic and critical analysis, I look for support, or lack of it, for my initial premises; create a typology of action repertoires; and categorise proactive data activism initiatives and projects. Of course, all classification is arbitrary, and any typology provides ideal types that only partially cover all the nuances of reality. I only use them as a heuristic tool to understand action in a complex field in continuous evolution.

Annex I includes the basic questions posed to all interviewees, as well as a list of the interviewees, a brief description of their positions and merits, the date and place of the interview and the manner in which the interview took place. Some of the interviews required several exchanges with the interviewees.

The semi-structured interviews follow the guidelines of qualitative search. That is, the interviewer and interviewees engage in a 'highly formal' interview (Bernard 2006, 210); an 'interview guide' or a list of questions, topics that need to be discussed during the conversation, the particular order they have been developed, and instructions that steer the questioning have been taken into account (Robert Wood Johnson Foundation 2008). Semi-structured interviewing, according to Bernard (2006, 212), is best used when there is no chance to interview someone more than once, such as in many of the cases included here. Most of the interviewees are based in a variety of locations and are highly occupied people. Semi-structured interviewing also allows a degree of flexibility to maintain 'discretion to follow leads' without straying away from the objectives of the exercise (ibid.). The idea is to be in full control of what is needed from each interview, but leaving both interviewer and interviewee space to explore related ideas that

emerge in the interview. The interviewees in this exercise are a group of leaders, directors, innovators and founders who are at the forefront of the data revolution or leading data-based innovation processes in different places of the world. Semi-structured interviewing is especially suitable in, since it 'works very well' with 'people who are accustomed to efficient use of their time' (ibid.).

These cases are not observed from the outside only. As a data activist immerse in some global struggles (i.e. the fight against climate change and biodiversity loss, and their impacts on human development), I also resort to participant observation, which grants me privileged understanding and access to the inner workings of some organisations and campaigns, and insights into activist processes as they take place. Precisely, the purpose of participant observation is to understand events from the perspective of the people that are being observed and understand activism from within (Balsiger and Lambelet 2014, 146). Participant observation allows for first-hand data collection, adapting to concrete realities and direct experiencing (ibid., 145-146). It is especially useful in the context of data activism, because it breaks the illusion of homogeneity within forms of collective behaviour. However, there are limits to the level of diversity and granularity that can be offered in a study such us this. Although I have proposed some categorisations of proactive data activists, I am aware of the fact that, when one looks at groups closely like you do as a participant observer, homogeneity disappears (ibid.); and when homogeneity disappears, the risk is to generate as many classifications as there are cases. I therefore suggest these categorisations –simplifications— as heuristic tools to think about data practices in a systematic way. There is a degree of subjectivity here as well. These categorisations emerge from the interviews and empirical observation, and as such, they depend on interpretation. The texts generated by the interviewees have meaning and interpretation can

get at that meaning. The analysis of the data is about finding patterns in data, and interpreting those patterns is up to the researcher as well (Bernard 2006, 453). Interpretation —expressing what findings mean and connecting the findings to the findings of other research— is a 'pretty qualitative exercise' too (ibid.). As 'interpretations matter hugely, but their status is rarely divorced from action' (Couldry and Hobart 2010), I admit openly that I am both enabled by the experience of being immersed in activism, but constrained by it as well.[48]

The data analysis comprises, too, an empirical observation and classification of cases of data activism. The classification is based on data that are derived from or related to *a posteriori* observation of relevant cases rather than theory. That is, it is based on what proactive data activist organisations actually do, rather than say they do.[49] I examine the actions and outcomes of these organisations, aiming at extrapolating conclusions to apply to a larger set of organisations. In order to come up with categories, I work with concepts, defined and established as attributes –that is, a feature that is characteristic or an inherent part of the examined endeavour—, to which values are assigned to designate intensity of the attributes present in each case. For example, from observation, proactive data activist organisations tend to work in alliances or partnerships (i.e. either big data sets are too large or social problems are too complex to be dealt with by a single individual). 'Works in alliance with other like-minded organisations' is then established as an attribute of proactive data activists. But is it so in all cases? And to what degree? To determine the level of intensity of the attributes, I have also set up a range, from 1 to 3, in which 3 is the maximum degree of intensity of each attribute. The non-assignment of a

[48] I am also biased and enabled by the languages I speak, by my education and experience, the access I have to technology and knowledge, etc.

[49] These data emerging from sources without my intervention is then combined with the data extracted from the interviews, where the interviewees express what they think they do prompted by me. The interviewees' experiences and opinions also populate the dissertation, whenever relevant, since they represent some of the mentioned organisations and cases as founders, directors or experts.

value means there is an absence of an attribute. This is to capture the idea that proactive data activist organisations usually specialise in one endeavour and strategy, but have no qualms in crossing lines separating areas of action such as advocacy and campaigning, funding, research, journalism and media, and humanitarianism. The result is that all proactive data activist organisations examined here are granted values for more than one attribute. Cross-pollination is in itself an attribute of proactive data activism.

I then outline the results in a grid, and employ R Studio in order to determine whether there are correlations among attributes, which there are.[50] To guarantee that the method is adhered to, there is a degree of meticulousness in the assignment of a unit of value to an attribute with regards to a concrete organisation, although I am alone making the decision and there is an unquantifiable amount of arbitrariness in the assignment of a value. The usefulness of this exercise is that, when comparing an organisation that 'works in alliance with others' to an organisation that does not, a clear distinction can be made between the concepts of 'collaboration' and 'non-collaboration' with respect to data activism. This type of categorisation allows the extraction of multifaceted reflections and definitions too. For example, if a proactive data activist organisation is totally dedicated to skills transfer activities (value level: 3), sometimes provides funds or resources to catalyse the creation or the continuation of data projects (value level: 2), produces limited proactive data activist content (value level: 1), occasionally generates data visualisations (value level: 1), sometimes works in alliances or collaborates (value level: 1), and often provides match-making opportunities between organisations to produce proactive data activist projects (value level: 2), it can then be categorised as a 'skill transferer' that also has something of a visualisation-producing

[50] I do not pretend to state that these correlations have any statistical validity, since I do not even know how big the universe of data activists is, and therefore I cannot establish a representative sample.

organisation and a ‘catalyst.’ This is the case of DataKind, which is examined more in detail in this dissertation.

Finally, the qualitative case study targets a proactive data activism initiative and its developers –the Ushahidi platform—, with the aim of tracing the project evolution from its inception to present, and understanding how people use data for social change. The Ushahidi platform is a case of successful proactive data activism. The case study methodology is especially helpful when the boundaries between a phenomenon and its context are not clearly evident (Yin 2002, 13), as is the circumstance of proactive data activism, a phenomenon deeply rooted in the context of the social uses of data infrastructures. In other words, the case study has been chosen as a method because contextual conditions are deeply pertinent to the exploration of the phenomenon known as proactive data activism. The subject has been chosen as an instance – a key or typical case — of a class phenomenon to illuminate proactive data activism. The study is single-focused, diachronic, descriptive and exploratory, as expounded next.

Ushahidi is identified as the subject of the case study because it is a complex example of proactive data activism, which expands over a number of years and offers an insight into the context of the phenomenon besides illustrating it (Davis and Wilcock 2003, 3). The case study is addressed by a single focus on one organisation –Ushahidi— from its inception in 2007 to 2015, as the progress of this organisation illustrates the evolution of proactive data activism as well. Ushahidi is at the intersection of several trends, namely the social uses of big data infrastructures, digital humanitarianism and collective action. The case is an instance of a class of phenomena (proactive data activism) that provides an analytical frame which the case illuminates and within which the empirical inquiry is conducted over a sustained period of time. This is primarily a descriptive study that utilises one case to show what proactive data activism

looks like from proximity, to make the unfamiliar familiar, and to provide a common language and framework for future studies (Woodside 2010, 322).

The case selected for this study is a typical case. I have chosen it conscious that in elucidating lines of causation it is more useful to select subjects that offer a rare set of circumstances. However, the key case was selected because of its in-built relevance (i.e. a pioneering geoactivist platform in a fairly new endeavour, that is, digital humanitarianism), its context and its capability to directly alter reality. The case selection is therefore based on its representativeness, and this is the basis for the development of a viable model of effective proactive data activism at the end of this dissertation. I have considered Ushahidi to be a critical case –that is, as having strategic importance in relation to the general problem (Flyvbjerg 2006, 233). A critical case allows the following type of generalisation: 'If it is valid for this case, it is valid for all (or many) cases' (ibid.).

The case being studied includes an organisation in action, that is, a data-based project (outcome) implemented by an organisation (structure). The case study comprises in-depth interviews[51] with three of its leaders and a deployer, including life histories that focus on different aspects related to the creation of the organisation, and analyses of people, decision-making processes, real-life contexts, policies and the institution, studied holistically. Dozens of documents have been analysed to contrast, verify and contextualise the data extracted from these interviews. The interviews specifically aimed at the case study differ from the semi-structured interviews employed in the data gathering exercise. They resemble the typical of long interview in that they include: a) interviewing the respondent in his or her life space; b) asking open-ended, semi-structured questions with deeper exploration of unexpected topics related to the study as opportunities occur (sometimes several interviews were necessary with the same individual); c)

[51] See questions and names in Annex I.

tape recording of responses during the interview; d) verification of responses by triangulation of research methods (i.e. comparing answers with data from direct observation and documents); and e) developing descriptions of individual cases and people (ibid., 264). However, due to the fact that the interviewees are based in far away locations, they were not conducted in person, but over the phone, and not all exceeded the two hours timeframe. In two of the cases, they involved several exchanges. In short, this case study research relies on multiple sources of evidence, and benefits from the prior development of theoretical notions in the first part of the dissertation.

The general purpose of the case study is basically exploratory, since proactive data activism has not been clearly defined yet (Yin 2002). Although reactive data activism has been explored before, and there are some conceptual distinctions that be made in relation with proactive data activism, exploratory research helps me determine the research design, data collection method and selection of subjects (ibid.). Data activists, whether individual or collective, use data either to fight in favour of a cause, to respond to a crisis or to resist and defy social injustice and civil rights breaches, with the ultimate goal of generating social change in the long run. How do they do it? The primary objectives of the case study are to verify how Ushahidi fits in the theoretical description of a data activism organisation that I have outlined in the first section of the dissertation, and to observe how Ushahidi uses data for social change. More precisely, I look into what their theory of change, tactics and strategy, challenges, and main successes and failures, as well as potentials, are. The questions, concepts and categorisations emerged in (i.e. from the perspective of SMS, what are the unconventional means set in motion by Ushahidi in its pursuit of change?) are then applied to the case study with the intention of contributing to the practice of activism, but also that practice generates new knowledge.

The specific objectives of the case study include: to understand the intersections between Ushahidi's humanitarian efforts and its more long-term aspirations from the perspective of data use; to reinforce the understanding of the rules guiding data-enabled social change and to generate an analytic paradigm that can be used in the future; and to help identify various pressures that surround data-enabled social change in the context of disaster and develop a model for data activism.

2.3 Ethical considerations

Important ethical questions surround the study of big data infrastructures, their social uses, and their relation with activism and collective action. Berendt, Büchler and Rockwell wonder about whether data scientists should cooperate in augmenting data infrastructures' snooping potential (2015). However, there is no list of criteria against which one can measure the 'ethicalness' of every research idea or research process (Bernard 2006, 74).

Milan considers this a specially pertinent challenge, because collective action always entails the production of knowledge and political imaginaries; that is, new ways of seeing the world, which emerge in horizontal, bottom-up processes that should be respected (Milan 2014a, 446). A study can bring undesired attention to activists' doings and put them at risk (ibid., 447). Besides, activists are typically 'highly invested subjects' who 'tend to expect from the researcher some sort of political alignment with the principled ideas they embody,' a factor that should be taken into account and treated with responsibility (ibid.). Finally, the participation of the 'research objects' –that is, the interviewees and others— involves an investment of time and effort on their part, and the researcher should place 'realistic expectations' on them (ibid.).

These considerations have been pondered to conclude that this dissertation involves neither physical nor reputational risks for the participants. It does not face particularly complex

ethical challenges in relation to the treatment of the sources of information and interviewees, as none of them are people at risk or vulnerable (i.e. most of them are experts and leaders). Even if the interviews include personal information about the interviewees, this is not deemed sensitive.

It would have been different if this dissertation were focused on reactive data activists, who often resort to secrecy and insurgent tactics as defence mechanisms against digital intrusion. In that case, I would have been careful and used stricter protocols in order to protect such sources. On the other hand, it would have been important to consider the accuracy of the data obtained from such anonymous sources.

Finally, questionnaires are designed to extract the maximum information with the minimum effort on the interviewees' part. I requested a further review of the interview by the interviewee when necessary.

All interviews were accompanied by an informed consent form.

In general, I faced the ethical challenges of any research effort: to be honest in the presentation of evidence and sources; to be straightforward when including references relevant to the investigation, and to avoid including irrelevant references; to be methodical when referencing only the sources that were actually used; and to be clear in the presentation of intertextuality. Big data are an inherent part of a globalised world where capitalism, consumerism and post-modern individualism triumph. Understanding the conditions that can bring about social change and the role data analysis and visualisations can play to foster it in this context is at the core of this research effort. The social impact of this dissertation derives from its objectives and approach.

3. The Big Data Society

Snowden's revelations about the massive espionage practices over the whole world (with scarce legal protection or simply illegal) have exposed a society in which nobody can escape Big Brother's surveillance, not even Merkel.[52] It has not always been like this, because we were not digitalised and technologies powerful enough to obtain, associate and process such an immense mass of information didn't exist before. The emergence of so called big data and the gigantic databases in communicable and accessible format (such as the huge NSA archives in Bluffdale, Utah) have resulted in a strengthening of the intelligence services after the barbaric attack against New York, as well as of the cooperation between the big technologic corporations and the governments, in particular with the National Security Agency of the US (which is part of the Ministry of Defence, but enjoys of a great autonomy) (Castells 2015).[53]

As cultural materialists attempt to recreate the *zeitgeist* of a particular moment in history to explain social facts, I try to recreate in this section the socioeconomic and technological context and the spirit of the 'big data society' –a society in which big data have flourished and have permeated most human and non-human activities. This chapter deals with the aspects of the post-modern 'big data society' that are relevant to the emergence and development of data activism. The big data society is global, but at the same time deeply individualistic; the big data society is an upshot of the 'network society' (Castells and Cardoso 2005, 3-6); in the big data society there are gaps, as some people are excluded from it (that is, they are not represented by

[52] Edward Snowden is a computer analyst whistle-blower who had worked for the US Central Intelligence Agency (CIA) and in 2013 provided several media outlets with top-secret US National Security Agency (NSA) documents leading to revelations about systematic massive surveillance on phone and internet communications in the US, with the connivance of European governments.

[53] 'Las revelaciones de Snowden sobre las prácticas de espionaje masivo del mundo entero (con escasa protección judicial o simplemente ilegales) han expuesto una sociedad en la que nadie puede escapar a la vigilancia del Gran Hermano ni Merkel. No siempre ha sido así porque no estábamos digitalizados y no existían tecnologías suficientemente potentes para obtener, relacionar y procesar esa inmensa masa de información. La emergencia del llamado big data, gigantescas bases de datos en formatos comunicables y accesibles (como el inmenso archivo de la NSA en Bluffdale, Utah) ha resultado del reforzamiento de los servicios de inteligencia tras el bárbaro ataque a Nueva York así como de la cooperación entre grandes empresas tecnológicas y gobiernos, en particular con la Agencia de Seguridad Nacional de EEUU (que forma parte del Ministerio de Defensa, pero que goza de amplia autonomía)' (ibid.).

big data); and these facets are reflected in the volume of big data that are amassed by governments and corporations. In order to explain big data and their uses, I also define several important phenomena, including mobile technologies –huge generators of data—, the computational politics of governments and the modern *panopticon* (Foucault 1984, 1995), along with the 'space of flows' (Castells 2010b, 407-453).

As sociology appeared in the Enlightenment to explain the transformation of *old regime* societies, big data and big data infrastructures are arising as the new tools and language of sociology to explain post-modern societies. But what is so new about the world today? Castells responds that, among other trends, new are ICTs, the emergence of a network society based on 'a space of flows' and on 'timeless time,' the phenomenal urbanisations process, the global 'inter-linked capitalist economy,' 'the fading away of communism,' 'the widespread challenge to patriarchalism,' 'the universal consciousness on ecological preservation' (2010, 372). These facets are new, but the world of today's big da society has not created itself; the current fixation with data and evidence owes a great debt to the Enlightenment. Paraphrasing Castells: today's computing power, the ubiquitous mobile technology, the global markets, the inter-linked capitalism, the urban labour force, the challenges to patriarchalism, the emergence of the 'network society' can be tracked back to the naissance of the 'Age of Enlightenment' and the enshrinement of evidence-based knowledge. And the big data society is also an evolution of what is known as the 'digital age' (Negroponte 1995)[54] and the 'network society' (Castells 2010b). It is a product of a globalised planet, where most of human and non-human activity generates data or is *datafied*. In the big data society, people and things are connected,

[54] 'In the fall of 1993, when Bell Atlantic agreed to buy cable giant Tele-Communications Inc. for US$21.4 billion, pundits of the *information superhighway* took it as a signal that the digital age had truly begun. The digital ribbon had been cut' (Negroponte 1995, 88).

communicate and produce massive amounts of data and metadata simultaneously and continuously, which are gathered mostly by governments and corporations, creating the 'informational state' (Braman 2006; Tufekci 2012). However, in the 'big data society' there is also space for data-enabled individual and collective action addressed at positive social change.

But what attributes characterise the big data society? The next chapters explore the Enlightenment as the source of key social constructs to understand the emergence of big data. I also define big data, examine the most relevant characteristics of the post-modern societies: mobile technologies, the birth of the post-modern *panopticon*, big data society's socioeconomic conditions and how it excludes some communities, the digital public spheres, individualism in a global planet, and the generation of a space without time or geography –the 'space of flows.'

3.1 The new digital public spheres

Why write about the Enlightenment when we are dealing with possibly one of the most recent innovations in post-modernity? This dissertation moves in two realms narrowly linked to that period: sociology and critical thought are the disciplines that I have employed to explore data activism. And 'sociology arose as the theory of bourgeois society; to it fell the task of explaining the course of the capitalist modernisation of traditional societies and its anomic side effects' (Habermas 1984, 45-46). Meanwhile, big data infrastructures –whose foundations can be traced back to the Enlightenment too— are being employed currently to explain the course of the big data society.

The notions outlined below are key for this dissertation in several ways. The focus of this study is to describe how data are used for social change, democratic progress and emancipation in a world that is the upshot of modernity. Still, a critique of the *status quo* – of ideology, value systems, power structures and dynamics— is embedded in data-enabled activism. And that is

why it is important to understand where predominant ideologies, value systems and power structures that are legatees of the Enlightenment. The Enlightenment opened the door towards a hunger for knowledge and the primacy of reason, and marked the birth of the scientific method and computing, a social contract theory and the notion of public consent, as well as the ideals of equality, feminism, individualism, democracy, freedom of expression and emancipation. It was the starting point for mass media and the public sphere, as well as capitalist liberalism and communitarianism. It saw the creation of civil society as a concept, along with the separation of the state powers in government, judiciary and parliament. This all meant a 'general reformation' of knowledge and 'the way we look on every aspect of life' (Israel 2006, 5). What is today's data activism fighting for? Mainly it is fighting for emancipation, which is the definitive enlightened ideal; that is, economic, social and political rights, and equality for all. As we will see in the following sections, today's post-modern society has created many divides that make people unequal. The paradox is that technology and big data, the very same tools that warrants such divide, can be an instrument of emancipation as well.

Broadly speaking, the Enlightenment period was marked by growing empiricism, rational inquiry, scientific rigor, observation and innovation, and critical thinking, and was not only the origin of the scientific method, but also the source of many modern paradigms. In his key essay *What is Enlightenment?* Kant described the Enlightenment as freedom to use one's own intelligence, as daring to know. 'Have the courage to use your own understanding' he urged the reader (Kant 1996). The centrality of reason, critical thinking and self-reflection, together with the concept of democracy, are some of the main values of the Age of Enlightenment, as opposed to traditional dogmas, religion and the divine right of kings to rule. These ideas converge in two dimensions: the individual (critical thinking) and the collective (the democratic will of the

majority). Another key idea of this period is the contractual basis of rights between citizens (the governed) and the government, which would later bring about the organisation of states into self-governing democracies as well. What would be the Enlightenment in relation to big data? If the Enlightenment is the free use of reason, how do we guarantee it in the digital age? How do we generate knowledge? And how can we go about examining ourselves as subjects? Daring to know remains a challenge for us still today. Foucault thus explains it:

> The critical historical ontology of ourselves has to be considered not, certainly, as a theory, a doctrine, nor even as a permanent body of knowledge that is accumulating; it has to be conceived as an attitude, an ethos, a philosophical life in which the critique of what we are is at one and the same time the historical analysis of the limits that are imposed on us and an experiment with the possibility of going beyond them (1984, 14).

If the big data promise is to be fulfilled, it may make it easier for individuals and communities to use critical thinking and to experiment with 'the possibility of going beyond' our own limits and understand society better. This is a central idea in this dissertation.

Deeply entwined with the Age of Enlightenment is the scientific revolution, pioneered by physicists and mathematicians, such as Bacon, Newton and Descartes. Although at the time it was the source of intellectual disputes across 'the Channel' between the followers of island-dwelling Bacon and the continental Descartes, the modern scientific method combines Bacon's empiricism and Descartes's rationalism, aiming at obtaining knowledge through research. This marks the beginning of a trend that outlives its onset. A belief that dominated the Enlightenment also was the idea of science as an accumulative process that involved the discovery of an initial problem, the formulation of a hypothesis, its critical examination, its review, and the detection of a new problem as a result of the process. This idea engenders the 'doctrine of progress' and the

notion that progress is a law of nature (Mill 1973, 219). But this notion was challenged in the 20th century, after two world wars and the invention of the atomic bomb, when it was understood that devastation could be bred in technical innovation as well (Adorno 2004, 26; Huesemann and Huesemann 2011, 247-251). In the 21st century, when anthropogenic climate change threatens future life in many parts of the planet, the 'doctrine of progress' has been totally discredited (ibid., 230-254). But from the Enlightenment on, the data-based evidence, experience and causation became central issues, and rationality was applied systematically to every problem and all realms of understanding, starting with mathematics, physics and chemistry, permeating other academic disciplines and even creating new ones, like sociology. The scientific method became the paradigm, in the sense that Kuhn uses to describe paradigms (1970). Ironically, the scientific method, which is at the root of the technology that has made big data possible, is being currently challenged by enthusiasts of data infrastructures and artificial intelligence, bringing about another paradigm change, which has prompted Anderson to proclaim 'the end of theory' (2008). This aspect will be examined more in-depth later.

This period gave birth to other key ideas that set the scene for the development, centuries later, of 'big data society.' One of these ideas was the social contract –a tacit agreement between the government and its citizens (the governed), who surrender some of their freedoms, and promise to cooperate and obey the laws in exchange for social benefits and protection (Rousseau 2010, 6-9). The social contract rests on the dilemma that all societies face in getting a balance between individual freedoms and collective security, and on the concept of 'public consent,' referring to the underlying agreement between governed and government. This is an important notion here because, right now, big data and other technologies put to the service of undemocratic causes are also challenging the social contract that has endured since the

Enlightenment. A social precursor is Locke, known for his assertion that individuals have a right to 'life, liberty and property,' the very foundations of individualism and capitalism. He influenced enlightened thinkers, including Voltaire and Rousseau, who argued in favour of a society based upon reason (instead of religion), and a science resulting from experiments and observation. Another social contract theorist, Hobbes of Malmesbury, provided the foundations for modern social contract theory (1651). Though he was a champion of absolutism, Hobbes further developed some of the foundations of European liberal thought: the natural equality of all 'men;'[55] the rights of the individual, and the view that all legitimate political power must be representative and based on the consent of the people. The emergence of modern nation-states and capitalist relations created the conditions under which ideals of citizenship, governmental accountability and civil society established themselves. These conditions gave birth to the public sphere as well, another key concept for this dissertation.

Associated is the notion of the 'commons.' The commons as a concept has come to mean shared –sometimes unregulated— resources, such as fish stocks in international waters today. The internet is another interesting commons. The 'tragedy of the commons' is a notion first explored by Lloyd to convey a situation where individuals behaving rationally in accordance with their own self-interest act contrary to the best interests of the community by exhausting some common resource (1833, 31-32).[56] This dilemma is extremely contemporary today, as it is

[55] Meanwhile, the first feminists started to claim that equality was for everyone, not just 'men.' Wollstonecraft, one of England's earliest feminist philosophers, argued that women, as well as men, should be treated as rational beings (1993, 263-285).

[56] 'If a person puts more cattle into his own field, the amount of the subsistence which they consume is all deducted from that which was at the command, of his original stock; and if, before, there was no more than a sufficiency of pasture, he reaps no benefit from the additional cattle, what is gained in one way being lost in another. But if he puts more cattle on a common, the food which they consume forms a deduction which is shared between all the cattle, as well that of others as his own, in proportion to their number, and only a small part of it is taken from his own cattle. In an inclosed pasture, there is a point of saturation, if I may so call it, (by which, I mean a barrier depending on considerations of interest,) beyond

being defied by how data infrastructures are being used by some governments and corporations. Are data part of the commons as the internet is? The commons is understood today also as a governance doctrine, a way to collectively produce and manage resources that belong to all for the common good (Colaborabora 2016). The concept has gained popularity in the past few years thanks to the open source movement, and after, in 2009, Ostrom was awarded the Nobel Prize in Economics for her defence of the commons, and of people's common sense to manage it (The Economist 2012). The open source movement advocates for open licenses that make software code available for anybody to use or modify, as opposed to proprietary systems, which underwrite 'modern capitalist hegemony' (Milberry 2005). It promotes learning through the unrestricted dissemination of collective creation. Examples of popular open source software products include Mozilla Firefox, Android and OpenOffice.org. The commons and the *open* concept are deeply ingrained in many data activism projects, as it will be seen later. There is currently a strong public pressure for governments to open their data vaults and make them open; some of them practice what is known as 'open government,' that is, a policy that maintains that citizens have the right to access public documents and data for an effective public oversight and accountability. These governments provide access to 'open data,' or data that individuals can access easily, and use, reuse and redistribute (Open Knowledge 2016).

While the theorists of the French Enlightenment were not radicals, their ideas played a key part in weakening the prevailing Old Regime, bringing about the French Revolution with its motto '*liberté, égalité, fraternité.*' Some of the enlightened ideals are still guiding principles of

---

which no prudent man will add to his stock. In a common, also, there is in like manner a point of saturation. But the position of the point in the two cases is obviously different. Were a number of adjoining pastures, already fully stocked, to be at once thrown open, and converted into one vast common, the position of the point of saturation would immediately be changed' (ibid., 31-32).

today's activism,[57] in a sector, civil society, which was generated as a concept during the Enlightenment as well.[58] The concept of liberalism is based on liberty, the initial element of the tripartite *motto* of the French revolution, and linked with the ambition to be liberated from the absolute power of monarchs. But the moment one inspects the concept, '*liberalism* fractures into a variety of types and competing vision' (Gaus, Courtland and Schmidtz 2015). Classical liberals link the concept of liberty with the right to private property, because, they argue, the freedom to obtain private property is simply part of people's liberty (ibid., 5). Here the idea is that the distribution of power that results from a free market economy and dispersed private property safeguards the liberty of individuals against intrusions from the state. What has come to be known as 'new' liberalism challenges this close association between individual liberty and a private property based market order, by which liberalism, losing faith in the fair forces of markets, adopts some of the socialist ideals (Freeden 1986, 25-32). Deeply related to the liberal thought, a free market economy is an economic system in which the value for goods and services is set freely by mutual consent between sellers and buyers, and the prices are regulated by supply and demand, with little state intervention (unrestricted by state intervention, in the case of the *laissez-faire* economy).[59] With this rules, markets are supposed to arrive at some sort of equilibrium.[60] Although they are associated with capitalism, free markets have also been

[57] Kumar recently proposes a new *trinity* –soil, soul, society—, representing the environment, the individual and the community (2015, 92).

[58] Hume, known for his radical philosophical empiricism and scepticism, among other Enlightenment thinkers, developed a 'science of man,' focused on the understanding of all facets of humanity. Hume's philosophical concepts influenced Madison, one of the 'fathers of the US Constitution,' and his writings would be the basis of classical liberalism. In the meantime, Smith published *The Wealth of Nations* in 1776 (2007), which is considered the first book on modern economics; lays the foundations of classical free market economic theory; and has an immediate and profound impact that continues well into the 21st century.

[59] An economic system in which transactions between private parties are totally free from government interference, such as regulations, privileges, tariffs and subsidies (Badhwar and Long 2016, 17).

[60] This notion has been discredited too after the global financial crack initiated in 2007 (Crotty 2008).

advocated by market socialists, anarchists and libertarians and some advocates of profit-sharing. In contrast, a regulated economy is an economic system that is controlled by the government's intervention in the supply and demand. But in reality, there are many hybrid examples of both systems.[61] Meanwhile, capitalism is a social system based on the recognition of individual rights, including property rights, in which much property is privately owned (Badhwar and Long 2016). Liberal capitalism, therefore, interlaces economic and social theories, based on individual rights, liberty, private property and free markets, and will be understood henceforth as a doctrine in which the greatest possible number of economic decisions is made by individuals, and not by collective or state institutions. Liberal capitalism is usually contrasted with *communitarianism*, which will be understood hereafter as philosophy that places the community –understood as a group of people that share a geographical location, an identity, an interest or a combination of them— at the centre. The Enlightenment led to a revival of communitarianism and collectivism, concepts that precede it and will come to be crucial in the 21st century as well. State intervention is here a mechanism to safeguard the interests of the community. Communitarian philosophy is grounded upon the principle that a person's social identity is mainly fashioned by community associations. It is sometimes considered progressive on social issues, and linked with leftism (socialism and communism) on economic issues. After the French revolution, theorists like Morelly, Auguste or Buonarroti emerged and influenced French labour activists. But the first socialist movements, under the umbrella of utopian socialism, did not properly develop until the 1820s and 1830s. Communitarianism would evolve into a critique of liberalism. While the classical liberalism of the Enlightenment can be viewed as a reaction to centuries of authoritarianism and dogma, 'modern communitarianism can be considered a reaction to

[61] An example is the massive bailout of the financial system by capitalist governments all over the world after the financial crisis. The bailout of only two institutions, Fannie Mae and Freddie Mac, for example, was estimated in more than US$300 billion by the Congressional Budget Office (Cover 2011).

excessive individualism, understood by communitarians as an undue emphasis on individual rights, leading people to become selfish or egocentric' (Etzioni 2016, 3). Although the origins of communism are still debated, the industrial revolution, with its surpluses and private property, and its nascent proletariat and dreadful working conditions, generated another fundamental ideal: that of a classless, egalitarian society based on common ownership and shared responsibilities. Industrial manufacture and the production lines brought about something different too: the disconnection of the workers with the product of their labour and their reduction to a cog in the machine. The influx of people into cities to find work contributed to an increasing urbanisation of the population of Europe and the US, but also produced an oversupply of workers which allowed those with capital to exploit the labourers. Marx theorised that the people would always be divided into two classes, those with capital and those without it, and foretold that the industrial revolution would speed the process of history along, and that the people would institute a state run by the workers for the workers. In any case, the deterioration of the enlightened ideals has been observed by Putnam, who noted that, in today's US, all forms of civic involvement and association have experienced drops in membership. An example is that, while today more people were bowling than in the 1950s, there are fewer bowling leagues. This is the result of an erosion of the 'social capital,' defined as the collective value of social networks and their inclination to do things for each other (2001, 5).

Why is it appropriate to define, even so roughly, these ideas here? For two reasons: the first is that our current economic and political system –where big data thrive— is a direct heir of these enlightened notions and the way they have evolved and entwined; the second is that further down this dissertation, I describe the current debates about big data from both liberal and communitarian perspectives, and also label the uses people, organisations and governments make

of big data by looking at their essential goals. Big data and big data infrastructures can be put to either liberal-capitalist or communitarian uses, roughly divided. An action with liberal purposes can have communitarian benefits, or externalities, and *vice versa*. However, for the purpose of this dissertation, I will consider than an initiative is communitarian if its final goal is the common good and has a collective, open nature. Some of the enlightened spirit is still present in a number of digital communities that are ferociously cooperative and egalitarian too. An example is the international network known as Anonymous, which is a decentralised, open-sourced and diverse group of all hands on deck *hacktivists* involved on a number of civic issues.[62]

In many ways, post-modernity is both an evolution and a rejection of modernity. With the Industrial Revolution, 'the development of productive techniques came into prominence as an important dimension of societal evolution' (Habermas 1984, 152). In the post-modern society, offshoring production and the dematerialisation of the economy result in a society where immaterial knowledge and information become the most prominent societal value, a characteristic of the big data society. The big data society is light and ethereal, while modernity is heavy (Lipovetsky 2016). With capitalist growth, 'the economy came to the fore as a functionally autonomous subsystem' too (Habermas 1984, 152). This brought into play the social division of labour, which 'needed no longer be reduced to aggregates of individuals;' now

[62] The group became known for a series of digital stunts and distributed denial-of-service (DDoS) on government, religious and corporate websites. A DDoS is an attack where multiple systems, which are often infected with Trojan *malware*, are used to target a single system causing an interruption. 'We weren't always legal... Still, most people who cover Anonymous get it wrong… We are condemned as criminals without consciences, dismissed as anti-social teens without a cause, or hyped as cyber-terrorists to justify the expanding surveillance state. But *hacktivism* exists within the history of social justice movements' (Hammond 2014, 4). Hacktivism is to be understood hereafter as 'the use of legal and/or illegal digital tools in pursuit of political ends' (Gorenstein-Massa 2013, 63). For Jordan, 'hacktivism is politically motivated hacking. Hacktivism is activism! running free in the electronic veins that enliven our 21st-century, global socio-economies' (2004, 119).

societies 'could be regarded on an analogy with organisms as self-maintaining systems' (ibid., 152). The capitalist economy is another element that completes the big data society landscape.

With the industrialisation, other matters troubled the minds of thinkers and sociologists. Weber's main intellectual concern, for example, was grasping the processes of rationalisation, secularisation and 'disenchantment' that he associated with the upsurge of modernity.[63] Weber proposed that ascetic Protestantism[64] was key for the rise of market-driven capitalism and the nation-state in the western world (2005). Russell too saw Enlightenment as a phase in a progressive development, born out of the Protestant reaction against Catholicism, the seed of democracy itself. In a paper entitled *Was Weber Wrong?* Becker and Woessmann provide an alternative theory: Protestant economies prospered because instruction in reading the Bible and Protestants' higher literacy (and access to knowledge) generated the human capital crucial to economic prosperity (2007). Whatever the reason, the industrial revolution blossomed in Protestant economies and spread out to more countries. Meanwhile other related notions emerged as new sources of attention. In the context of industrialisation and capitalism, for example, what constitutes the state? Weber defined the state as a human community that 'claims the monopoly of the legitimate use of physical force within a given territory' (1946, 77).[65] These are interesting debates to understand the context in which big data emerged later in the 21st century, since they deal with the value of knowledge, the nature of human and social capital, and the concept of 'public consent.' Meanwhile, Comte, applying the principles of the scientific

---

[63] In contrast with Comte, Weber was a key proponent of methodological anti positivism, arguing in favour of the study of social action by way of interpretive means (rather than purely empiricist, as proposed by positivists), based on understanding the purpose and meaning that individuals ascribe to their own actions. This is another very important contribution to sociology.

[64] In comparison with 'the cycle of sin, repentance and forgiveness, renewed throughout the life of the Catholic,' absent in Protestantism, as Anthony Giddens says in the introduction of *The Protestant Ethic and the Spirit of Capitalism* (Weber 2005, xii).

[65] Weber made other major contributions in sociology and economic history, as well as economic theory and scientific methodology, and influenced the critical theory associated with the Frankfurt School.

method, founded the discipline of sociology and stated the doctrine of positivism, for which only information derived from logical estimations and direct experience is the source of authoritative knowledge. Comte was a major inspiration of the 19th century ideas, influencing the work of other social thinkers such as Marx and Mill. His concept of *sociologie* and social evolutionism set the tone for social theorists and anthropologists, evolving into the modern academic sociology represented by Durkheim. Together with other influences, including Kant and neo-Kantian thought, Marx as well as the critical neo-Marxian theory, Weber**,** Freud, the Frankfurt School of social theory[66] and Parsons, Durkheim inspired Habermas's thought. It is Habermas, in fact, who reviews the founding of modern sociology in *The Theory of Communicative Action.* 'Sociology originated as a discipline responsible for the problems that politics and economics pushed to one side on their way to becoming specialised sciences' (1984, 45-46). Sociology – another important outcome of the Enlightenment— is the prism employed in this study to observe the social uses of big data.

Modern democracy is another notion that owes its inception to the Enlightenment thought. Spinoza was one of the thinkers that set the ground for democratic values. The Enlightenment was the birth of democratic ideas, such as the natural rights (those that do not come from a particular form or government, but from being a human being) versus the divine right of kings; the separation of powers, influenced by Montesquieu's *The Spirit of the Laws*; the will of the people and the power of majorities; the separation of church and state; and religious freedom; and equality among individuals. In the 1700s the notion that sovereignty lay with the people (in reality, an ancient idea) became coupled with other notions born in the Enlightenment, especially liberalism, socialism and communism. This is an important conception too, since I am arguing that data infrastructures can help tackle social challenges from a democratic perspective. So,

[66] Especially, Adorno, Marcuse and Horkheimer.

what is a *democracy*? There is not a commonly accepted definition, but one emerges looking at indicators of democratic development. *The Economist* famously publishes a Democracy Index every year, even if it acknowledges that 'there is no consensus on how to measure democracy,' and that 'definitions of democracy are contested' (Kekic 2007, 1). Kekic notes that this is not only a theoretical question, but a practical one. For example, although democracy-promotion is high on the list of US foreign-policy priorities, 'there is no consensus within the US government on what constitutes a democracy' (ibid.). He distinguishes between 'freedom' and 'democracy,' which he defines as 'a set of practices and principles that institutionalise and thus ultimately protect freedom,' and proposes a set of fundamental features of a democracy that, at a minimum, should be included in this concept, 'government based on majority rule and the consent of the governed, the existence of free and fair elections, the protection of minorities and respect for basic human rights' (ibid.). Also, importantly, Kekic says that any democracy 'presupposes equality before the law, due process and political pluralism… (Because simply the) rule by the majority is not necessarily democratic' (ibid.). In order to rank countries, *The Economist*'s Democracy Index takes into account not only political and electoral freedoms, but also the state of civil liberties, including individual rights and the rights of the minorities; the quality of functioning of the government; a democratic political culture; and participation of citizens in political life. 'Democracies flourish when citizens are willing to take part in public debate, elect representatives and join political parties. Without this broad, sustaining participation, democracy begins to wither and become the preserve of small, select groups' (ibid.). Talking about the democratic values and civic virtues, de la Cruz emphasises the collective instead. Civic virtues include 'responsibility, solidarity, social trust, duty, civic sense, altruism, loyal disagreement, patriotism' (2016, 66). In fact, authors as distant as de Tocqueville (2002) and Putnam (2001)

underline vigorous debate, collaboration, participation and sense of community belonging as part of what makes a democracy strong. Based on these concepts, a democracy is to be understood henceforth as a political system that institutionalises and protects both individual and collective freedoms, establishing a balance between them, which is enacted through the separation of powers –legislature, executive and judiciary—, monitored by independent and free journalists, and based on the rule of the majority and the consent of the governed, the existence of fair elections, the protection of minorities and vulnerable groups, and respect for human rights.

How do communities democratise, then? They undergo a democratisation process –that is, a process in which democracy, in all its facets, expands. The different patterns of democratisation often explain and are explained by other phenomena, such as the level of development[67] of a given community, the distribution of wealth among people, social equality of its individuals, the degree of foreign influence and openness, and the level of education and access to information. In the absence of a globally accepted definition, it has been theorised that education promotes stable and democratic societies, and research shows that 'increases in levels of education improve levels of democracy' (Alemán and Kim 2015, 1). Democracy has other properties. For example, Sen states that 'no famine has ever taken place in the history of the

[67] Development does not necessarily mean economic development, as democracy does not imply economic development either (Kekic 2007). A country can develop increasing other facets than the economic, for example, achieving equality or fighting against poverty, without necessarily increasing its Gross Domestic Product (GDP). Examples of this are Rwanda and the Philippines, very poor countries (World Bank 2015) with narrower gender gaps than much richer nations (World Economic Forum 2015). Although development requires economic investments, this type of investment does not automatically translate into GDP growth either. At the same time, economic development can run parallel to social gains and an intensification of civil liberties; however, this link is not always consequential. An example is China, were the remarkable economic growth of the past three decades has not been tied in with an equivalent growth in freedoms at the same rate. Development is understood as the eradication of poverty and hunger, the achievement of healthy lives, equitable education and gender equality, the universal availability of sustainable water, sanitation and energy, the promotion of inclusive economic growth and resilient infrastructure, the reduction of inequality among countries, the transformation of cities into resilient, safe spaces, the fight against climate change, the conservation of oceans and terrestrial ecosystems, and the promotion of peaceful societies (UN 2016).

world in a functioning democracy' (Sen 2000, 16). Why does starvation occur when they are slow moving crises that can be predicted? The answer cannot be divorced from 'the issue of liberties, of newspapers and ultimately of democracy' (Sen 1982). With all its faults, it is the best system that humanity has come across to govern itself. Common features in totally different approaches to democracy are the notions of participation, equality, the collective good, and the balance between individual and collective freedoms. Therefore, tackling problems from a democratic perspective means that, when engaging with the problem and its protagonists, the approach is democratic, that is, it is based on equal participation (in a Habermasian sense) and seeks the common good. And when this process generates a solution, it reinforces democratic relations, that is, the solutions to those problems incorporate the democratic values described above. Concretely, when I talk about data infrastructures being tools that allow tackling social problems from a democratic perspective, I refer to their power to reinforce individual and collective freedoms, to empower and give a voice to the voiceless, and to facilitate the equal participation of citizens in decision-making. Precisely Haque, interviewed for this dissertation, says that data, rendered useful through purposeful analysis, can be democratic in empowering people to make intelligent decisions and take action.

> I am not interested simply in amassing a dataset about air quality… those kinds of datasets serve to show from a very top down perspective a situation without really offering any resolution. But… most people cannot change where they live, they don't have the financial capacity to decide where to live, they cannot do something about their daily activities to change data that is being generated. So the data is this fetishizing activity of amassing data for data's sake. On the other hand, when somebody starts measuring air quality for their own reasons, which might be that one of their children has asthma or there has been some

> kind of health effect in the family, then actually data itself makes almost no difference. What matters is that by doing certain things –for example, by opening a window or by closing the door or by planting a certain type of tree or by planting a certain type of plant—how the measurement gets affect. And that measurement is the demonstration that a peace lily actually filters the air well. That is when the data becomes useful, as a decision-making tool that enables someone to say: "You know what? I am going to plant more peace lilies. And I might even tell my neighbour to do the same thing" (ibid.).

Another important aspect of democracy that deserves examination is its dependence on knowledge, and on a free press and independent media organisations –whether conventional, alternative or citizen—as intermediaries between information and people. On the way towards modernity, democracy progressed hand-in-hand with the mass media, which became central to our society. In spite of the fact that conventional media have seen their influence wither together with their circulation, audiences, rates and sales, and that the old linear communication model sender-receiver is no longer valid in the network society, media organisations still have the power, 'not only convey, but simultaneously create an image of reality, our knowledge and understanding of the world' (Ortíz-Ceberio and Rodríguez 2016, 77). Although the history of mass media could be tracked back to Medieval China and Korea, it was not until the European popular prints of the 1400s that the earliest mass medium was invented. Guttenberg introduced the printing press in Europe in about 1440, and this, together with movable types and increasing popular literacy, gave rise to the mass publication of books and newspapers on a scale never known before. Newspapers expanded from the 1600s and reached a mass audience in the 19th century, when the first high circulation dailies, such as *The Times*, were set up and made possible the invention of the high-speed rotary steam printing presses. Meanwhile, modern transport

systems allowed large-scale circulation of papers. The increase in circulation and volume, however, led to a decline in direct interactivity with the readers, making newspapers more a more a one-way medium. The notion of 'mass media' was restricted to print media until the aftermath of the Second World War, with the popularisation of radio and television, which provided not only information, but also entertainment. Radio and television further increased the tendency of the general public of passively receive content. In post-modernity, a new phenomenon would emerge reverting this: the 'mass self-communication' (Castells 2009): the limits between *senders*/*emitters* and *receivers* would blur and the interactivity between *content generators* and *consumers* would come back, only in a different manner. That is, with the internet, mass communication would evolve into mass self-communication, and the classic one-way communication standard would turn into 'a new form of communication… over the internet and wireless communication networks that would turn the tables in the relations between "power and counter-power," "formal politics" and "insurgent politics," as well as set the scene for 'new manifestations of social movements' (2009, 239).

The notions of the social contract between the people and the government, and the mass media mingle in 'the manufacture of consent,' a concept coined by Lippmann (2009 [1922]), who presents mass media as the main vehicle for governmental propaganda. Later Herman and Chomsky would declare that mass media are 'effective and powerful ideological institutions that carry out a system-supportive propaganda function by reliance on market forces, internalised assumptions and self-censorship, and without significant overt coercion' (1988, 306). This propaganda system became even more efficient with the rise of the national television networks, greater media concentration, right-wing pressures on public radio and television, and the growth in sophistication of public relations and news management (ibid.). However, media could be also

a channel for people's voices in political matters. Exploring the links between media, power structures and 'the governed,' Castells notes how global civil society acted on the public mind 'via the media and communication networks' (2009, 42). Consequently, the quality of people's understanding depended not only on their individual abilities, access and 'daring to know,' but also on the diversity, depth and dependability of the news coverage available. This reliance meant that media played a key role in keeping an eye on the powerful and maintaining people informed. In modernity, media and journalists became *gatekeepers* of what was publicly known and of public debate.[68] The level of freedom of speech and diversity of the media of a given country at a given time, consequently, turned out to be extremely important for both the degree of democratic development. Well into the 20th century, mass media economic power dynamics led to today's concentration of ownership in media organisations, which meant a few corporations controlled the creation of content and its distribution channels, from the designer's table all the way into people's living rooms. And this situation led to a distorted competition within media and a negative impact on the quality of democracy. In some cases, media offered a biased view of reality, ultimately failing the public (Alger 1998; Gutiérrez 2004a[69]; Castells

[68] The process of 'gatekeeping' was described in 1922 by Park in *The Immigrant Press*, where he states that 'out of all the events that happen and are recorded every day by correspondents, reporters and the news agencies, the editor chooses certain items for publication which he regards as more important or more interesting than others. The remainder he condemns to oblivion and the wastebasket. There is an enormous amount of news "killed" every day' (Park 1922, 328). During my time at EFE, the wastebasket was called ironically the *archivo general* (general folder).

[69] 'The marriage of media content (news, films, TV shows) with media distribution (TV or radio networks, internet services and the like) further increases the control of media barons over the audience, as they use their sales power to batter their way into living rooms… One infamous US case of media concentration is that of Clear Channel, which now owns more than 1,200 radio stations in the nation. "A couple years ago, there was a chemical spill in a small town in South Dakota. Authorities wanted to alert the local community about the toxic fumes and related dangers, and contacted the local radio stations," says Lewis. "But they were owned by a national company, Clear Channel, which had no staff locally, and no local news or public affairs programming. So local citizens could not be alerted about a real public health and safety subject via the airwaves in that municipality"' (Gutiérrez 2004a).

2009, 61[70]). In post-modernity, with the internet and mobile technologies, media and journalists lost their monopoly over gatekeeping.

Down the line, democratic development did not evolve in the same manner everywhere. De Tocqueville concluded that the US, in contrast to Europe and its enduring aristocratic codes, was a society where money-making was the overriding ethic, the common person enjoyed an unparalleled level of self-esteem and could elevate him or herself through hard work, and individualism and market capitalism had taken root to a remarkable degree. De Tocqueville writes: 'The public has therefore among a democratic people a singular power, of which aristocratic nations could never so much as conceive an idea; for it does not persuade to certain opinions, but it enforces them, and infuses them into the faculties by a sort of enormous pressure of the minds of all upon the reason of each' (2002, 495-496). De Tocqueville discovered something fascinating: associationism, collaboration and dialogue were the fabric of democratic development. But democracy does not flourish unceasingly. Two centuries later, Putnam would describe 'the collapse (and renewal) of American community' and the spirit that was the substance of the US vigorous democracy (2001, 11). This process is also relevant here, as some data initiatives try to restore these democratic values.

Another essential concept is that of the 'public sphere,' which as a phenomenon appeared during the Enlightenment, although it was not noted or termed then. It was Habermas who, in the 1960s, tracked the origin of the public sphere back to the 'enlightened' bourgeois salons of the 18th and 19th century England, France[71] and Germany, 'until its makeover due to the influence

[70] The case of Miami Herald Publishing Co. v. Tornillo (1974) included a reference to the Twentieth Century Fund Task Force Report for a National News Council saying: 'This concentration of nationwide news organisations – like other large institutions— has grown increasingly remote from and unresponsive to the popular constituencies on which they depend and which depend on them' (Findlaw 1974).

[71] 'While the *bourgeiosie*, for all practical purposes excluded from leadership in state and Church, in time completely took over all the key positions in the economy, and while the aristocracy compensated for its

of the capital-driven mass media' (1991).[72] The public sphere sets the social conditions for a rational debate about public issues conducted by individuals willing to let arguments, and not stations, determine decisions. The question is a crucial one for democratic theory too. In his answer, Habermas reveals a communicative ideal that afterwards would supply the essential normative standard for his theory: the idea of inclusive critical debate, unrestrained and free of constrictions, in which contributors consider each other as equals in a shared attempt to arrive to a common understanding. During the Enlightenment, the public sphere surfaced against the absolute power of royals as a neutral and autonomous arena of critique, as social circles that evaded censorship, and offered freedom of opinion and interaction. 'In a process of enlightenment there can only be participants,' and everybody gets a chance to talk and be listened to (1973, 40). Habermas has described its nature as a communicative ideal: a space of inclusive critical debate, uninhibited and free of restrictions, in which people treat each other as equals in an effort to arrive to a mutual understanding (1991, 33-34). Habermas has been criticised for having created too abstract a model and for having bypassed some dialectic challenges, such as the exclusion of women in general as well as men from lower classes (Webster 2006, 163-168; Fraser 1990, 2014).[73] Plage points out that Habermas 'fails to

---

material inferiority with royal privileges and an ever more rigorous stress upon hierarchy in social intercourse, in the salons the nobility and the *grande bourgeoisie* of finance and administration assimilating itself to that nobility met with the "intellectuals" on an equal footing. The plebeian d'Alembert was no exception; in the salons of the fashionable ladies, noble as well as bourgeois, sons of princes and counts associated with sons of watchmakers and shopkeepers. In the salon the mind was no longer in the service of a patron; "opinion" became emancipated from the bonds of economic dependence' (Habermas 1991, 33-34).

[72] The concept of the public sphere has triggered many discrepancies and has been criticised mainly for being too abstract or elitist. In *Ideal Theory, Real Rationality: Habermas versus Foucault and Nietzsche*, for example, Flyvbjerg compares the Machiavellian *verita effettuale* (effective truth) and the Nietzschean *wirkliche Historie* (real history) with the Habermasian *Diskursetik* (discourse ethics). The former are praised as the best approaches, as they are 'seen as more effective means to understand and limit rationalisation and power' by this author (2000).

[73] Another criticism of Habermas's theory is that he restricts hermeneutic practice to the humanities, while the sciences are also interpretive. In fact, one of the problems of our 'Information Age' is the

acknowledge the heterogeneity of the public, with its many competing interests' as well (2009, 8). That may well be. But even as an idealisation, the enlightened public sphere includes some key features, the most interesting of which is that, in the public sphere, everybody gets a chance to talk and be listened to, that there is an opportunity for the best argument to win, because nobody has the upper hand (Habermas 1973, 40). Habermas's description of the public sphere is aspirational, not an account of something real; but this concept is of tremendous importance today for several reasons[74]. One is the possibility for data infrastructures to generate alternative digital public spheres and empower people so they can really participate as equals. Habermas's requirement of an equal and unobstructed participation in the public debate as a condition must be qualified, since the bourgeois public sphere, in its early days, was an elitist bubble for educated, reading, propertied men – most women and all other members of society excluded. Its character was 'equal' only for its privileged few, but 'exclusionary' for others (Calhoun 1992, 223). The inclusion of dissimilar social groups happened progressively in the course of history,

---

'information overload' (also known as *infoxication*) that exists, which requires constant interpretation of everyday life situations. The term was disseminated by Toffler in *Future Shock* (1984), but is mentioned in other works. He defines it as 'the distress, both physical and psychological, that arise(s) from an overload of the human organism's physical adaptive systems and its decision-making processes. Put more simply, future shock is the human response to overstimulation' (ibid., 168). But in the 21th century, we are in a situation where interpretation has never been more difficult in spite of the technological artefacts that we have developed. Castells says: 'We live in confusing times, as is often the case in periods of historical transition between different forms of society. This is because the intellectual categories that we use to understand what happens around us have been coined in different circumstances, and can hardly grasp what is new by referring to the past' (2009, 27).

[74] Dean argues that the notion of the public sphere is not only 'inapplicable to the Net,' but also 'damaging to the practice of democracy under the conditions of contemporary technoculture' (2003, 95). He argues that whereas the Habermasian public sphere is rooted in the nation, has as telos to generate consensus by applying legal and rational procedures, is ruled by inclusivity, equality, transparency and rationality, and is enacted by actors, net-based neodemocracies are rooted in the web as a 'zero institution,' aim at contestation, are based on networked conflict and credibility, among others, and are driven by issues (ibid., 108). Although Dean is right in noting that Habermas theorised the public sphere in those terms, it is also valid to think that, in retrospective, we do not need to. For example, one could discard the 'nation' as a necessary context to theorise about the public sphere, since the essential element here is the 'community.' In fact, the alternative digital public spheres I theorise about capture only some of the features of the original Habermasian public sphere.

contributing to the openness of the public sphere but, in Calhoun's interpretation of Habermas, that very process of gradual inclusion introduced 'degeneration in the quality of discourse' (ibid., 6). However, the real deterioration of the public sphere and its 'structural transformation' came with the mass consumption and commodification of culture and private affairs reinforced by the mass media (Habermas 1991). The media turn active citizens into passive spectators creating a pseudo-public sphere. The world fashioned by the mass media is a public sphere in appearance only. By the same token the integrity of the private sphere is also an illusion (ibid., 1991). As these behind-closed-doors circles matured into mass publics in the 19th century, ideas were transformed into commodities and became engulfed by the economic machinery of mass media consumption. People became clapping consumers. 'In the manipulated public sphere an acclamation-prone mood comes to predominate' (ibid., 217). Indeed, in a society dominated by mass media, acclamation, and not critical discourse, is what becomes important, says Calhoun. 'The public sphere becomes a setting for states and corporate actors to develop legitimacy not by responding appropriately to an independent and critical public, but by seeking to instil in social actors motivations that conform to the needs of the overall system dominated by those states and corporate actors' (1992, 26). The effect of the mass media on the public sphere is to invert its initial significance.[75] The public sphere becomes the sphere 'for the publicising of private biographies, so that the accidental fate of the so called man in the street or that of systematically managed stars attain publicity, while publicly relevant developments and decisions are garbed in private dress and through personalisation distorted to the point of unrecognizability' (Habermas 1991, 172). The spectator develops 'a sentimentality toward persons and corresponding cynicism

[75] Habermas perceived the media as adding to the decline of the rational-critical discourse and causing the weakening of the public sphere. However, the concept has been reviewed by some authors that believe that, quite contrarily, the media have the capacity to instigate public argument and give it a constructive turn. The internet regarded as a public service and big data analysis made available to the people offer the possibility to contribute to the same purpose and multiply its effects.

toward institutions which with social psychological inevitability result naturally curtail the subjective capacity for rational criticism of public authority' (ibid.) .[76] Television's allure and the willingness of the media to acclimatise to the commercial rules of the market caused a deterioration of the critical discourse. A personalised politics provokes a representative publicity by making political candidates into media stars.[77] The result is that 'non-public opinions are at work in great numbers, and "the" public opinion is indeed a fiction,' concludes Habermas (1991, 244). But rather than give up on his ideal, Habermas calls for a viable concept of public opinion. A public opinion 'that is historically meaningful, that normatively meets the requirements of the constitution of a social-welfare state, and that is theoretically clear and empirically identifiable can be grounded only in the structural transformation of the public sphere itself and in the dimension of its development' (1991, 244). I argue that data infrastructures, in the hands of people, can offer the possibility of a viable concept of the public sphere, only different. As we will see later, innovation and big data infrastructures have generated two opposing forces: they facilitate governmental and corporate intrusion, manipulation and control; but on the other hand, they can be a powerful tool for equal interaction, activism, participation, social change and democracy. 'For every negative, there will be a counter response that has the potential to be a substantial positive,' say Schmidt and Cohen (2013, 56). Two covers of the *Time* magazine broach these overlapping realities. In 2006, "the person of the year" was "You," or the content-generator and sharer, web-user, social media contributor, blogging "You" (Grossman 2006). "The You whose miasma of bits, bytes and the powerful images they express are becoming

[76] Bauman will later explore this detachment between citizen and public administration, contributing to the growing uncertainty that surround us in his books, especially in *Liquid Times: Living in an Age of Uncertainty* (2007).

[77] Calhoun also criticises Habermas for judging 'the 18th century by Locke and Kant, the 19th century by Marx and Mill, and the 20th century by the typical suburban television viewer' (1992, 33). The consequence, for Calhoun, is that 'Habermas's account of the 20th century does not include the sort of his approach to 17th, 18th, and 19th centuries' (ibid.).

increasingly problematic" (Holmes 2016), craved and hoarded. Four years later, the cover of *Time* is "The Protester" (Andersen 2011). "You" becomes the adversary. "As governments, police forces and other power players… crack down on voices of dissent, it is only You, The Protester, armed not with a press pass, but with a smartphone and a Twitter account, who brings the rest of the world its news" (Holmes 2016). Further down the line, protesters and activists started to utilise data and data infrastructure too, which alone cannot rebuild the public sphere, but can generate alternative digital public spheres by empowering people, generating truly democratic spaces for equal participation and action.

Another key Habermasian concept here is the communicative action. The teleological action presumes links between an actor and facts in an objective world (Habermas 1984, 85-88). In this type of action, the actor is bestowed with cognitive abilities (realisations that an actor has about the facts) and volitional faculties (an actor's intentions that are meant to attain a goal). Consequently, the chief question for an actor is dual: he or she must try to bring together his/her opinions with real life and facts, but has to integrate the real-life situations and facts in his/her wishes as well. By always communicating with one another to coordinate actions, actors work at certain goals, and that is why the teleological structure is fundamental for collective action. For the teleological action, the objective world alone is of significance, whereas a norm-regulated action occurs in the social world. Actors belonging to the same social world share normative and value systems. Values regarding the handling of particular situations turn normative because of the general recognition by the members of a social world, who are then allowed to expect another actor's action to be consistent with the accepted social norms. Systems of norms are composed of prior experiences and lessons learnt that were useful and are built-in as a code of behaviour. The norms that are well integrated are typically not questioned, but interiorised and

automatized. They are found useful; they help navigate a society. Norms are *conditionalities* (systems of social or cultural control) that rule a society too. They are made of the social values received in the socialisation process. But this concept is defied when an actor purposely challenges the social world and takes actions that deviate from the social norms (ibid., 89-93). There is a whole line of social action –including data activism in many cases— that is focused on precisely changing normative systems. Meanwhile, looked at from the point of view of a dramaturgical action, social interaction is thought to be an exchange between two or more actors, in that the actors form an audience for one another and present something to one another. This type of action needs a subjective world that the actor refers to. By showing something of his/her subjectivity, an actor wishes to be accepted by the audience. That means that the expression of feelings in a genuine form is not possible (i.e. as found in his/her inner self), and every actor routinely plays a particular role (that is, becomes a *persona*, wears a mask). A sincere actor tries to communicate feelings as faithfully as possible, persuaded that the idea of reality that he performs is the reality. At the other extreme is the actor who is not deceived by his/her own performance (ibid., 91). Finally, the communicative action focuses on the prerequisites for an inter-subjective understanding, which requires an integration of the three concepts of objective, social and subjective world. That is to say that an actor is keen to employ the power of rationality of all three worlds to obtain the goal of mutual and clear understanding in interaction with others. Understanding is therefore only possible if (1) the statement is true, (2) the actors act in harmony with accepted social norms, (3) and if the actions undertaken are sincere. That requires that every actor who is truly interested in understanding another actor is forced to reconcile another actor's opposing description of a situation with his/her own description. However, unambiguousness is an exception in real-life situations; what is more realistic is the picture of a fragile and constantly

revised communication (ibid., 99). Since only communicative action makes possible a mutual understanding of each other, it is the type of action required for successful campaigning and activism –a type of communicative action that includes a clear purpose within a normative system and displays the subjectivity of the dramaturgical action to connect with different actors (audiences) and seeking inter-subjective understanding. Evidence-based campaigning seeks an undistorted understanding among stakeholders and target audiences, based on shared opinions of existing, real facts and equality, with a normative structure as a context, and conditioned by sincerity. This type of campaign shares Habermas's resolute conviction that, through rationality and critical thinking, we can change our minds if we hear a better argument. Habermas argues that implicit in all speech is the possibility of an ideal speech situation where all interlocutors in which undistorted communication aims at reasonable consensus and communicative rationality. And he lays out a series of conditions for such communication. One is that undistorted communication would have a symmetry condition: everyone has an equal opportunity to talk and to listen, and an equal right to question and to answer (1984, 41-42). This concept is crucial.

Box 1: A Habermasian look at proactive data activism

Applying Habermas's definition of the communicative action, proactive data activism could be firstly considered a social meaning-building process realised through 'communicative action,' defined as a collaborative action undertaken by people (and in this case, by machines too) based on their mutual understanding that seeks to influence through knowledge extracted from data analysis. This process includes elements of the 'teleological action' (since it entails a kind of instrumental knowledge),[78] the 'norm-regulated action' (because it refers to and conforms to a set of norms and rules, from the structure of mass media to political power), and the 'dramaturgical action' (since campaigning, even if based on data analysis, also normally relies on the emotional language of subjectivity in order to communicate

[78] A data-based campaign, as any other campaign, has a teleological purpose; it generally seeks to influence to change policy or practice, or both.

efficiently and draw adherents). This communicate action demands an integration of the three concepts of teleological, social and subjective actions.

In proactive data activism, the ‘agent’ anticipates decisions on the part of at least one additional goal-directed actor. Thus, Habermas would consider it also a ‘strategic’ action, since this model of action could conform to the following description:

> This model is often interpreted in utilitarian terms; the actor is supposed to choose and calculate means and ends from the standpoint of maximising utility or expectations of utility. It is this model of action that lies behind decision-theoretic and game-theoretic approaches in economics, sociology, and social psychology (1984, 85).

This is very important in campaigning, since no research will be undertaken if it does not serve a purpose in a defined strategy.

The symmetry condition has incredible political power: it is an egalitarian principle at the bottom of what Habermas understands as real participation. In an undistorted communication act, only a free human being can recognise that peculiar, natural force of the better argument. A better argument, presented in a dialogue where everyone is equal, can change people’s minds (Habermas 1989, 47). The symmetry condition has other implications. Big data infrastructure can be applied to negative goals (i.e. manipulation), as well as positive goals (i.e. digital humanitarianism). In a sphere were everybody –campaigners, decision-makers, interested audiences, citizens, voters, relevant communities, vulnerable people, victims and involved corporations– have parallel access to information and participate as equals in the decision-making processes, a real undistorted communicative action, and change, is possible. In these circumstances, proactive data activism presents immense opportunities as data analysis can empower people so they can sit at the decision-making table.

The public sphere is a space where communicative action can happen. In communicative action, actors seek to reach common understanding and to coordinate actions by reasoned argument, consensus and cooperation, rather than simply strategic action in pursuit of their own goals (Habermas 1984, 17-18). This precisely what activists strive for.[79] Alternative digital public spheres have some of the same attributes of the original public sphere; the main difference is that they have the potential to be more compliant with the symmetry condition and more inclusive, as many more people can now access information and participate,[80] and that alternative digital public spheres are aimed at action. Several factors enhance the capabilities of the digital public spheres: a) The huge connectivity of the 'network society' (Castells and Cardoso 2005, 3-6) resulting from how individuals and machines are associated simultaneously in a thick network of digital relations. b) The abundance of big data and the fact that they are a reflection of society itself, since practically everything we are and do can be, and is, *datafied*. c) The fact that big data infrastructures can be adequate tools to study post-modern complexity, and generate diagnosis and solutions. With those enhancing capabilities, digital public spheres can unleash processes of real communicative action. It is not the only property found in big data infrastructures. Both campaigning and journalism are being revamped by a revival of facts and statistics. Big data infrastructures have a lot to do with this departure from the purely subjective into reason. From

[79] This is not necessarily the case according for example to Fraser, for whom TANs's global campaigning is strategic and teleological, and that their impacts in enhancing public spheres are mere side effects. In *Transnationalising the Public Sphere*, Fraser notes that the TANs 'are concerned above all to be useful; they are not directly concerned with democratic legitimacy' (2014, 72). Fraser says that is only secondarily that TANs are focused on 'making institutions of global governance more accountable to stakeholders… Their concern is very rarely directed at improving the quality or the inclusivity of public debate as such, and nor does it necessarily do so. It is much more likely to be focused on a particular issue, place or group of people and on what can practically be done to remedy injustice and ameliorate conditions that are causing suffering' (ibid.).

[80] This happens despite of the different divides that exist, which are examined later.

an empirical observation and literature review,[81] I have detected a slight, but distinct, shift from *opinionism* into data journalism. Social networks[82] and blogs made it possible for anybody to *be* a journalist, write about any subject and disseminate commentary, which is a phenomenal swift in power relations within communication. With digital media and an internet connection or mobile phone you do not have to be in Afghanistan to write about its long war; you can form an opinion and spread it from the comfort of your coach. Do not need to stick to the rules of journalism as Kovach and Rosenstiel envisaged them either (2007). If you have you own blog, there is no fact-checker or editor on top of you demanding that all assertions are confirmed and that your story only includes all relevant information.[83] However, data journalism has brought back *The Guardian's* 1921 motto: 'Comment is free, but facts are sacred' (Scott 2002). And data-based reasoning has returned to journalism and restored it to its roots. Some of the interviews carried out for this dissertation point to this direction as well. Cabra, an investigative data journalist who started in television, moved away from 'human stories' into data because data's potential to tell the whole story, while faces tell a fragment of it. Clark tells how data has transformed not only his work, but also his points of view about issues like energy and climate change.

Apart from humans' fundamental interests in communicating, Habermas's third hypothesis is that humans have a common interest in emancipation, or the liberation from unnecessary constraints in labour and distortions in communication. Habermas says that we could labour

[81] See more details in the chapter dedicated to the uses of big data infrastructures.
[82] Online social networking services include Twitter and Facebook. The analysis of data produced by social networking provides the means, to study social entities to identify local and global patterns, to locate influential social actors in the network, and to examine network dynamics and social interaction between actors.
[83] I would not consider that journalism. The issue is not whether who does it is a salaried full-timer in a media organisation or not. The question is whether the result adheres to the mehod of journalism.

more humanly and communicate more clearly through reason (1984, 357). Froomkin explains how Habermas's theory can apply to activism and civil action: the forces needed to push public decision-making 'are likely to come from a re-energised, activist, engaged citizenry working together to create new small-scale communicative associative institutions that over time either merge into larger ones or at least join forces' (2003, 753). Froomkin proposes too that 'new technology may increase the likelihood of achieving the Habermasian scenario of diverse citizens' groups engaging in practical discourses of their own' (ibid.). He admits that technology, by itself, 'may not compel outcomes, but it certainly can make difficult things easier' (ibid.). Taking Froomkin's discourse a bit further, indeed technology does not induce outcomes, whether positive or negative, but it certainly can help reconstruct new digitalised public spheres where citizens engage in practical discourse and decision-making as equals, as we will see.

How does the Enlightenment tradition and the digital world come together? Habermas and Castells have the answer. They 'have provided two of the probably most significant metaphors' to explain the role of media, information and the wider landscape of communications: the 'public sphere' and the 'network' (Plage 2009, 1). Plage explores the question of whether Habermas's theory and Castells's description of empirical communicative phenomena might be complementary. Castells defines the 'network society' like this:

> A network society is a society whose social structure is made around networks activated by microelectronics-based, digitally processed information and communication technologies. I understand social structures to be the organisational arrangements of humans in relationships of production, consumption, reproduction, experience, and power expressed in meaningful communication coded by culture (2009, 24).

'Network society' is an expression that relates to the social, political, economic and cultural changes caused by the spread of networked, digital information and communication technologies. The intellectual origins of the idea could be traced back to other authors,[84] but for Castells networks constitute the new social morphology of our societies, and 'digital networking technologies, characteristic of the Information Age' drive networks in ways that allow 'their endless expansion and reconfiguration, overcoming the traditional limitations of networking forms of organisation to manage complexity beyond a certain size of the network' (2010, 18). A network is a set of interconnected nodes, which are the points in a network where the links intersect; it has no centre, just nodes, which can be increase their relevance for the network by absorbing more relevant information and processing it more efficiently (2004, 2). The accent in Castells's account of the network society is not on social networks *per se*, as they have existed since the beginning of human life, but on social networks that process and manage information, and use digital-based technologies to do so. In this work, Castells describes how networks have become the basic units of modern society. Castells argues that it is not purely the technology that defines modern societies, but also cultural, economic and political factors that make up the network society. While Habermas is suggesting an approach to explain 'the conduct of communication,' Castells seems to accept a pre-existing structure –the network –, which is shaping 'the measures necessary to promote equality' (Plage 2009, 1). Habermas's concept paved the way to include theories of democracy and governance within media studies, whereas

[84] Van Dijk defines the network society as a society in which a combination of social and media networks shapes its prime mode of organisation and most important structures at the individual, organisational and societal levels. He compares this type of society to a mass society that is shaped by groups, organisations and communities organised in physical co-presence (2005). In an information society 'the information intensity of all activities becomes so high that this leads to an organisation of society based on science, rationality and reflexivity; an economy with all values and sectors… increasingly characterised by information production; a labour market with a majority of functions largely or completely based on tasks of information processing requiring knowledge and higher education…; a culture dominated by media and information products' (ibid., 19).

Castells offers the analytical potential to theorise social transformation in a globalised world (ibid.). The Habermasian idealisation of the public sphere and Castells's network society model are not only compatible, but complementary. In commenting on both sets of theories, Plage observes that neither Habermas nor Castells are 'overly enthusiastic' about the impact of media (and the internet in the case of Castells) on democratisation processes. Habermas highlights the negative effect of mass media in the public sphere, and Castells points out: 'neither utopia nor dystopia, the internet is the expression of ourselves –through a specific code of communication, which we must understand if we want to change our reality' (2001, 6). Castells maintains a cautious view of technology, in fact, without denying internet's transforming power. The very same could be said about big data: however flawed or incomplete, big data are a representation of ourselves, and we need to understand and dominate them if we want to change our reality. Apart from the differences between these perspectives and the timeframes of their studies, there are also some important similarities. Plage points out that 'both formulate a way to define the role of the media as an essential intermediary between the individual and society' (2009, 8). Equally these concepts provide a point of departure towards theorising the media as a means of inspiration and political participation as well. In this way, Castells stresses that 'the structure and dynamics of socialised communication is essential in the formation of consciousness and opinion, at the source of political decision making' (2005, 12). Webster concludes that the notion of the public sphere 'may be utopian, but it is as real as is our conception of… the idea of democracy… It does not deserve being dismissed because nowhere has it been fully achieved' (2006, 202). Webster referred to the way the notion of democracy is valued in spite of the fact that is nowhere to be found 'in perfect working order' (ibid., 201). Another similarity between Habermas and Castells is that both thinkers put both individual freedoms and the collective good

at the centre of their work, and in both the rebellion of collectives, made of reasoning and equal individuals, is fundamental. Talking about his personal experiences in the Paris of the 60s, where he was exile because of his previous political activism against the Franco regime in Spain, Castells says that the 1968 movement influenced his theory in fundamental ways, because it showed that 'things could change… by protest articulated with the interests and values of society at large' (2001). Summarising, Castells and Habermas have created two of the most potent metaphors to explain communication processes, and the role of media and information – the supreme value of the Information Age. These notions are very useful also to understand big data and their uses. The public sphere may seem idealistic, but, as Webster suggests, I prefer to hang on to this ideal as something to aim for. After all, most mission-driven organisations strive for visions of a better world that will never materialise, and that does not make them less useful. As for the notion of the network society, it provides a framework for the emergence of big data, their infrastructures and their uses. Habermas produces a monumental approach to explain the conduct of communication, while Castells describes the pre-existing structure of the network society. These two authors careful optimism is also a source of inspiration. Paraphrasing Castells, things can change, not just by protest, but by protest fuelled by the interests and values of the community. And the spaces where this protest can be articulated, with the help of technology, are the new digital public spheres.

3.2 A definition of big data

The notion of big data is essential for this dissertation: data activism relies precisely on data, metadata and data infrastructures as enabling know-how, vehicles and tools. That is why it is important to define what big data really are and clarify many of the popular perceptions that circulate about them: are they actually valuable? Are they the new *oil*? (Singh 2013). If they are

the basis for data activism in the hands of individuals and organisations, who generates and controls them? In the previous sections, we have examined their distant origins in the Enlightenment. Here I review what space they occupy in the knowledge hierarchy, examine the different definitions of big data and provide my own.

A lot of big data was previously considered garbage, and the sheer volume of it –which made it unanalysable— was considered to be a private protection against prying eyes. Big data were similar to the introns[85] of genetics. Until very recently, nobody knew what they were for, and therefore they were considered 'waste AND.' As science and technology advance, we are now able to find explanations for introns and to analyse and render data and metadata useful. But in spite of it, there is no universally accepted technical definition of 'big data,'[86] as the International Telecommunication Union (ITU) admits. However, different organisations offer several useful definitions based on their qualities and 'behaviour.' It was META Group analyst Doug Laney who, in a 2001 research report, described data as being 'three-dimensional:' they are characterised by a mounting volume, the velocity of data coming in and out, and the variety of data formats and sources (2001, 1). Today the industry carries on using this '3vs' model in order to explain big data, although more recently it has been updated and expanded to include additional facets.

Let us start with trying to differentiate between business intelligence and big data analysis. As big data grow bigger and the concept more mature, firmer difference between big

[85] Introns are found in the genes of most organisms. An intron is any nucleotide sequence within a gene (Todo Sobre la Ciencia 2016).
[86] Hence the relative newness of the bibliographic references of this section.

data and business intelligence is taking root too. Business intelligence[87] normally refers to performing statistical analysis of data with high information density to take measurements, detect trends and correlations in order to make companies more efficient. Big data analysis uses inferential statistics from nonlinear system identification[88] to infer laws (relationships, causal effects, dependencies and nonlinear relationships) from large datasets with low information density to perform predictions of outcomes in many fields. With this perspective, Brown offers this definition: 'Business intelligence helps find answers to questions you know. Big data helps you find the questions you don't know you want to ask' (2015, 1). That is because when you deal with enormous sets of data, of which many are not structured, it is difficult to really know what you are looking for. And this is a very relevant issue, since some scholars are announcing 'the end of theory' (Anderson 2008). With big data analysis, instead of producing hypotheses and research questions and testing a theory, we scientific insights 'born from the data' can be obtained (ibid.). That is, the machine has the questions and the answers. Cukier explains it like this: 'One of the most impressive areas where this concept is taking place is in the area of machine learning. Machine learning is a branch of artificial intelligence, which itself is a branch of computer science. The general idea is that instead of instructing a computer what to do, we are going to simply throw data at the problem and tell the computer to figure it out for itself' (2014).[89] I disagree with Anderson when he says that big data are the end of theory, although this

[87] In business analysis, business intelligence (BI) is the set of tools to transform data into functional information, usually in companies or institutions, with the main objective of improving efficiency. BI technologies can manage large amounts of unstructured data to recognise, develop and generate new strategic business opportunities.

[88] System identification is a technique to measure the mathematical model of a system from measurements of the system inputs and outputs. The applications of system identification include economic data, biology, medicine, industrial processes and social systems. A nonlinear system is any system that is not linear; that is, any system that does not satisfy the superposition principle. There are very many diverse sorts of nonlinear systems (Golnaraghi 2012, 52-54).

[89] This matter is examined when looking at the different epistemological approaches to big data.

matter is much more complex than what I have described so far and it can be considered a false dilemma. In fact, Miller, and I agree, suggests that both approaches are compatible and complementary. He says that, although statistical analysis is viewed as an inductive process (i.e. from observation of a pattern to a formulating a hypothesis and a theory), it is 'often embedded within a broader deductive process' (i.e. from a theory and hypotheses to observation and confirmation) (2009, 18). It is probably through a relevant combination of methods that more and better insights can be originated. Besides, big datasets are always originated by an existing scientific, social and political framework, and can contain biases, errors and gaps. Big data analytics give the idea that they generate insights spontaneously, without the need for research questions, but the algorithms used to extricate those insights do result from an aprioristic scientific framework as well (Gitelman 2013, 147-167).

After offering this simple definition, based on the one produced originally by META Group. That is, big data are understood 'as being datasets whose volume, velocity or variety' is extremely high in comparison with normal or previous datasets (International Telecommunication Union 2014, 173),[90] ITU provides one of the most compete characterisation of big data, since it explores two facets that are missing in previous analyses: it adds 'veracity' and 'value' to the 'three vs. that are typically associated with big data's main dimensions. According to ITU, big data's velocity is the speed at which data 'are generated and analysed,' either real time or near-real time; variety indicates the range of data types and sources encompassed by big data, including 'unstructured data;' value involves big data's potential for

[90] ITU included a big data chapter for the first time in its annual report in 2014.

socio-economic growth;[91] veracity shows the 'level of quality, accuracy and uncertainty of data and data sources;' and volume suggests 'the vast amount of data generated through large-scale datafication and digitalisation of information,' measured in quasi astronomical units: terabytes and petabytes.[92] Let us explore some of the most interesting and relevant features in-depth: namely, volume, veracity and value.

Governments around the world are opening up their data vaults, allowing anybody access and (in the best cases) reuse to them as a result of the open data movement.[93] In the academia, big data infrastructures have facilitated the availability of previously inaccessible or isolated repositories of data. When data are not open,[94] individuals and organisations are generating their own datasets thanks to the increasing access to technology and doses of ingenuity. In many realms of life, there is a huge datafication trend, 'the ubiquitous quantification of social life' (Kennedy, Poell and van Dijck 2015, 4), which turns many aspects of life into computerised data. An example is how Twitter *datafies* wandering thoughts and utterings.[95] The Internet of Things (IoT) is another source of big data, to a point that Dull wonders whether IoT and big data are two sides of the same coin (2015). The result of this data exuberance is that we are living in a

---

[91] As I have suggested in the introduction of this dissertation, the value of data is a debated issue. I do not agree totally with this definition. Potential is not value *per se*, it depends on the processes and uses that data are put to. Only data rendered useful are valuable, what we do with them is valuable.

[92] A bit is a single numeric value, either '1' or '0', that encodes a single unit of digital information. A byte is a sequence of bits; usually eight bits equal one byte. One terabyte is one trillion (short scale) bytes and one petabyte is one quadrillion (short scale) bytes.

[93] The aspirations of the open data movement are similar to those of other *open* movements, such as open source and open access, based on the idea that data gathered by governments about people should be available to them, unless sensitive or confidential.

[94] 'Open data are data that can be freely used, reused and redistributed by anyone— subject only, at most, to the requirement to attribute and share alike' (Open Knowledge 2016). That is, the data –in a 'convenient and modifiable form' – must be available as a whole and downloadable over the internet. The data must be provided under terms that 'permit re-use and redistribution including the intermixing with other datasets;' and everyone must be able to use, reuse and redistribute' them (ibid.). For example, an image of a chart with some numbers on it is not open data.

[95] Twitter permits access to some of its data through a Public API or a publicly available application programming interface (API) that provides programmatic access to a proprietary software application.

time in which we are surrounded by data. Vast datasets are repeatedly produced and routinely stowed employing different technologies such as aerial sensors, ubiquitous mobile devices, and radio-frequency identification readers. As we move in cyberspace, our activities leave behind digital footmarks of our doings, in a myriad of software logs and communications metadata collected by service providers. The ability of generating and taking in of ever larger amounts of data has driven observers to speak of a new development phase in human history, which Hellerstein termed 'the Industrial revolution of data' (2008). Thus the most obvious of big data's properties, its volume (which is ingrained in their very name), does not generate much polemic: big data are essentially big. Volume is one of the first features that authors mention, but big data are not characterised only by it. Undeniably, governments, companies and academia have accumulated huge datasets for a long time (i.e. national censuses). However, given the costs and complications of producing, processing, analysing and keeping such datasets, until recently these data have been created in strongly controlled ways employing sampling techniques that limit their scope, temporality and size. The challenge of analysing big data is coping with copiousness and diversity, timeliness and dynamism, messiness and uncertainty, high relationality and the fact that much of what is being analysed was produced with no explicit query in mind or is a spin-off of another operation. That is, big data's complexity is also big. The magnitude of big data has led some observers to produce analogous definitions by negation: Big data are large unstructured datasets that cannot be managed and analysed by standard traditional data processing applications. Big data implies extracting value (i.e. recommendations based on consumers' behaviour, grouping of similar text, classification and prediction) from an immense quantity of data. However, these are outcomes that people have been obtaining before from 'not-so-big data.' That is why some experts differentiate between both groups by the tools you can

use to tackle them. For example, distinguishing between (small) data which can be handled by one computer or by one human brain, and big data, which need to be tackled in clusters of servers.[96] Here what is considered big data ends up depending on the capabilities of the organisation confronting the dataset. Big data turn out being data that are 'one bit too much for you to be comfortable' (Rutter 2014). Big data have also been described by some as a 'moving target:' what is considered to be big today may not be so big tomorrow. The world's technological per capita capacity to store information has almost doubled every forty months since the 1980s (Hilbert 2011); as of 2012, every day 2.5 exabytes[97] of data were created (IBM 2012). It is to be expected to grow even bigger in the near future. Innovative technology is being used to face these challenges, and each dimension of the big data poses a different challenge. The European Big Data Public Private Forum (BIG) project[98] –which examines the three traditional vs (volume, velocity and variety)— determines that volume 'places scalability at the centre of all

[96] There are many tools you can use to handle big data. Apache Hadoop –a set of algorithms or an open source software framework—, for example, allows for a stable and distributed processing of very large datasets on computer clusters built from commodity hardware. It works in clusters of servers with JavaScript –the programming language of the web— as a basis, and a Hadoop-based storage system, Hadoop HDFS. On top of that, there is a layer of distributed data processing, Hadoop MapReduce. Many projects generate applications and tools that work with Hadoop, but working on one sole computer with Hadoop does not make it big data, as it is not a distributed solution. In machine learning and data mining, big data processing is distributed and scalable. Other options include a NoSQL (often interpreted as Not Only SQL) database, which provides a mechanism for storage and retrieval of data that is modelled in means other than the tabular relations used in relational databases. A NoSQL is distributed, open source and scalable horizontally (you can add more servers and nodes), i.e. Redis (Value key), Cassandra (columns), MongoDB (documents), Neo4j (*graphos*). Another option is cloud computing, a metaphor based on utility and consumption of computing resources that involved deploying groups of remote servers and software networks that allow centralised data storage and online access to computer services and resources.

[97] One exabyte is one quintillion bytes (short scale).

[98] 'The BIG Project (http://www.bigproject.eu/) is a European Union coordination and support action to provide a roadmap for Big Data within Europe. The work of the project is split into groups focusing on industrial sectors and technical areas. The BIG project is comprised of: Sectorial Forum that gathered Big Data requirements from vertical industrial sectors, including Health, Public Sector, Finance, Insurance, Telecoms, Media, Entertainment, Manufacturing, Retails, Energy, and Transport. And the Technical Working Groups that focused on Big Data infrastructures for each activity in the data value chain to examine the capabilities and current maturity of technologies in these areas' (European Union 2014, 14).

processing' and that, in order to undertake it, 'large-scale reasoning, semantic processing, data mining, machine learning and information extraction are required.'

The veracity promised by big data, though, is possibly one of their most interesting, relevant and controversial features. As we have seen before, the veracity of big data has been questioned by authors looking at the political prejudice and their scientific framing (Gitelman 2013). Even if we would agree that the analytic method could be unbiased, the generation and extraction of those data are never free of prejudices. Because the thing is: not even the harmless most innocent, factual, empirical data, let us say data on stars, are free of biases. In order to obtain a number or a value, one needs 'other numbers,' which only exist once they pass through 'textual, historical and psychological filters' (Stanley 2013, 90). Each of these filters can be used 'positively or negatively, to either exclude a record from reliability or to detect reliable records' (ibid.). To be precise, if you dig deep enough in how any dataset has been originated, there is always a degree of subjectivity and prejudice. However, big data analysis can be extremely useful and produce novel insights. For Miller, science has functioned until recently in 'a data-poor environment,' and therefore 'measurements of reality were difficult, expensive and cumbersome to obtain, store and manipulate' (2009, 2). The study of the 13rd century Korean Buddhist canon, which contained 52 million characters distributed across 166,000 pages, illustrates this. The Korean Buddhist canon is a perfect example because of its huge size and complexity. Before big data infrastructures, it was only possible to examine it in fragments; and therefore science has focussed on getting relevant information from 'scarce observations' (ibid., 2). The limits of the scientific method have been noted by Huesemann, who believes that the scientific method 'is based on mechanistic reductionism, which means we are trying to understand the whole by looking at isolated parts; by looking at isolated cost and effect

relationships. That's what we do all the time. When we do research in the laboratory we just look at one parameter at a time' (2014). With big data analytics, for example, the whole Buddhist cannon can now be examined in its totality: 'all of the texts, all of the words and all of the metadata in every search' (ibid.). It seems more truthful than just analysing parts of it, but paraphrasing Stanley, to obtain a word, one needs more words, which have been necessarily gone through filters (Stanley 2013, 90). But can biased data be 'truthful'? Can biased data represent the reality? Or better, can the analysis of biased data generate scientific truths? It is a question of how biased a data set is and how transparent the study is about the biases and errors contained by them, and the processes involved in extracting and analysing them. Traditional social scientists have been happy before interrogating just selections of large databases, deemed relevant by previous research, and they have managed to extract relevant insights from these exercises. Gabrys, Pritchard and Barratt defend 'just god enough data' in activism, and I agree with them. Schmidt, for example, says that conventional sampling is not even representative in many cases. 'Humanists and scientists alike, trained in the language of survey research, tend to ask of datasets: "Is it a representative sample?" I doubt there is a single dataset of interest to historians that is' (2012). The field of digital history is a good perspective to examine the issue of the supposed veracity of big data and how it can be combined with other methods. Schmidt argues that the digitalisation of historical data makes the traditional skills of historians more relevant than ever. The difference is that historians 'need to *reinvent*, not *reaffirm*, the way that historians do history' (2012). And this rejuvenation should take place in relation with three practices: source criticism that 'explains what's in the data;' 'a hermeneutics that lets us read data into meaningful form;' and 'situated argumentation that ties the data in to live questions in their field' (ibid.). Relevant to the issue of whether data are 'truthful' or not is the first one,

source criticism, that is, examining how close data are to reality and how they were originated. Given the fact that access to data is today granted to almost anyone with the means, Schmidt sees a great opportunity for historians in creating a new 'source criticism' that really explains what is behind the datasets (ibid.). The same way journalists are no longer the 'gatekeepers' of what is widely known and the public debate, historians are no longer the only 'guardians' of historical archives, and their role should change to embrace transparency about datasets' origins. At any rate, veracity is merely a word that starts with 'v' and fits nicely in the 3vs paradigm. But veracity is not the same as truth. If a stricter sense of the word were applied, 'truthfulness' should be applied to big data instead, since, as Habermas would say, big data belong in the 'objective world,' where a 'valid claim' is a 'propositional truth' (1984, 71). In contrast, in the social and subjective worlds, valid claims represent either 'normative rightness,' (former) or 'subjective truthfulness' (latter) (ibid.) I favour the viewpoint expounded by Kitchin, who thinks that data are neither neutral nor complete (2014). To start with, big data do not portray the whole reality. There are millions of people who are not even considered by big data, who are for now 'invisible' to the big data watchful eye because they are not considered by algorithms (Letouzé and Sangokoya 2015, 11).[99] For example, there are big data gaps when it comes to information on developing nations, and poor and forgotten communities. In other words, there are many holes in the thick coat of connections that big data infrastructures have knitted over this planet. Millions of people are not represented by big data and escape their net because they do not have access to mobile technology or the internet, they do not communicate via WhatsApp,[100] emails,

[99] Letouzé and Sangokoya, though, refer to minority users; while I refer to a person who does not have access to technology because he or she is not an *homo consumericus* (Lipovetsky 2008).

[100] WhatsApp is an instant messaging and telephone app for smartphones that operates under a subscription business model. The proprietary *app* uses the internet to send text messages, images, video, user location and audio media messages, call other phones and create very popular chat groups. WhatsApp Inc. was acquired by Facebook on February 19, 2014, for approx. US$19 billion.

Twitter or Facebook, their movements are not captured by cameras and sensors, they do not drive vehicles connected to GPS technologies, they do not perform searches or commercial operations online, they do not interact with the IoT, and they neither own credit cards nor hold bank accounts. In other words, data-based research and analysis 'is not a neutral, objective activity that produces a view from nowhere.' And how that research is used 'is not ideologically-neutral, but is framed in subtle and explicit ways by the aspirations and intentions of the researchers and funders/sponsors, and those that translate such research into various forms of policy, instruments and action' (Kitchin 2014, 9). Besides, the mechanisms employed to transform input data into a desired output, that is, algorithms,[101] do not emerge from vacuum either. However, they are unfathomable (Letouzé and Sangokoya 2015, 15). Letouzé and Sangokoya advocate for 'a sociological inquiry into algorithms' that aspires to expose 'the complex workings of this knowledge machine, both the process by which it chooses information for users and the social process by which it is made into a legitimate system' (ibid.). According to these authors, algorithms are 'deliberately obfuscated,' and because they are designed to work without human intervention, they may hide a secret (human) desire to be relieved from the task of 'being sceptical about information we cannot ever assure for certain.' There are other related forms of asymmetry in big data that make them not entirely *true*. Andrejevic describes as the 'big data divide' the differences between those 'who collect, store and mine large quantities of data' and those who become 'data collection targets' (2014, 1673). But who collects the data of the collectors? Later in this dissertation I quote Telefonica's boss, Cesar Alierta, bragging about his old-fashioned cell phone; nobody is going to collect *his* data. This key distinction exacerbates 'power imbalances in the digital era,' says Andrejevic (ibid., 1673). While the Aliertas and the

[101] They are defined as 'a set of steps that are followed in order to solve a mathematical problem or to complete a computer process' (Merriam Webster 2016).

very poor of the world remain in the shadows, the middle classes' personal thoughts, moves and relations are up for grabs. In short, the veracity of big data has been questioned, but it is not such a great discredit. Admitting that they can include biases, gaps and errors (Gitelman 2013), it does not mean that big data analysis is insignificant, useless or false. We just have to be aware and open about the limits and promise of big data analysis.

The applications of big data infrastructures and their impacts have started to be felt everywhere, from businesses to science, from governments to the arts, from journalism to campaigning organisations. However, tackling big data –and the inherent challenges that materialise when working with huge amounts of datasets— is not an easy task. Innovative technology offers different options to face these challenges, providing new ways to extract, gather, clean, use, reuse, visualise and redistribute data, and eventually obtain value from them. The *value* of big data in economic terms and how businesses explore data to obtain efficiencies (which translate into benefits and savings) are other interesting facets of big data. Most importantly, the perceived value of big data is driving the accessibility and user-friendliness of data infrastructures as well. However, my starting position is that data are not valuable *per se*; data *rendered useful* are valuable instead. Benefits obtained by companies using big data infrastructures include intelligence and efficiency.

> The ability to effectively manage information and extract knowledge is now seen as a key competitive advantage, and many companies are building their core business on their ability to collect and analyse their information to extract business knowledge and insight. As a result, big data technology adoption within industrial domains is not a luxury but an imperative need (European Union 2014, 13).

This ability can be translated into money, to a point that Lawrence, among others, proclaims that data are 'the new currency' (2015). But as said before, data are not valuable in themselves; their value is in how they are turned into useful information and eventually into practical knowledge, in how organisations turn their organisation into an information-centric places that rely on 'insights derived from data analyses for their decision-making' (Boulanger 2015). And that is why big data infrastructures are considered an emerging field full of promises for socio-economic development, both for companies and other sectors. Specific value can be found in the processing of the different elements of the so called big data value chain.[102] A value chain is structured in a series of subsystems each with inputs and outputs. In its analysis, the BIG project applies the value chain model to the high-level activities of information systems in big data processes –from data discovery to data integration and exploitation— to examine the value-creation of data infrastructures. The data value chain also shows the complexity involved in tackling big data through the steps in the process of data usage. As said before, the business community has long acknowledged the economic opportunities for big data uses, and it is already segmenting this sector and generating revenues from each of the value chain segments. 'Big data represents a fast-growing multibillion-dollar worldwide opportunity. It is the topic on executive agendas and a driver of technology and services investment' (Vesset et al. 2012, 1). But not everything is good news. There is another economic impact that is not so welcome, according to Cukier: Big data are going to 'steal our jobs' (2014). There is no doubt that big data are going to challenge white collar, professional knowledge work, in a similar way that the assembly line challenged blue collar labour in the 20th century. Cukier thinks that, although big data technology will create new jobs (and new businesses), some of the old jobs will simply be

[102] The value chain is an analytical model used as a decision making tool that segments the activities that support the production of a service or product to manage them efficient and exploit their added value.

eliminated (ibid.). In any case, both as a generator of internal intelligence and efficiency, and as an originator of businesses linked to the activities of their chain value, big data have already created one of the fastest sectors in developed economies. The interesting consequence is that actual and expected gains have propelled the big data infrastructures development, have formed economies of scale, and ultimately are making these infrastructures available to 'normal' citizens and non-for-profit organisations, moving from the elitist periphery of the skilled *techies*, civic hackers and technology corporations to the centre of society, and from large datasets to small ones.

Learning Analytics offers yet a more 'expanded model' of the 3vs prototype, which includes volume, variety, velocity, variability, veracity, volatility, complexity and value (Boulanger 2015). Let us consider now the extra facets. Variability 'refers to the inconsistency which can be shown by the data at times,' thus hindering the process of handling the data effectively (GSR 2015). 'Additionally, data flows can be highly inconsistent with periodic peaks. Daily, seasonal and event-triggered peak data loads can be challenging to manage' (ibid.). So, big data infrastructures must be able to handle high levels of variability. Volatility 'refers to how long is data valid and how long should it be stored. In this world of real time data you need to determine at what point is data no longer relevant to the current analysis' (ibid.). Complexity is sometimes present in large datasets, but not necessarily. It refers to the intricacy of linking, connecting and correlating data coming from multiple sources (ibid.). Other sectors talk about the vs of big data too. In *The Five Vs of Big Data Political Science,* Monroe, exploring 'the challenge and opportunity that big data presents to political science,' identifies five vs.: volume, velocity, variety, *vinculation* and validity (2011, 1).

But what do big data vaults (governments' and otherwise) contain, exactly? We saw in the previous section that data can be facts, signals or symbols (*digital objects*).[103] But what are really these objects that can be scrapped, gathered, standardised, analysed, edited and communicated? Looking at how they are created could provide some answers. Big data can comprise:[104] a) a wealth of digital objects and user-generated online content (video and audio files, texts, tweets, messages, links and tags) resulting from online distribution and archiving, and commercial online transactions; b) other types of transactional data generated by human interactions in social networking platforms; c) private and public records that companies and governments amass and stock about their clients and citizens; d) signals captured and emitted by sensors and portable devices; e) digital tracks captured by the Internet of Things; f) digital traces left behind by web clickstreams and indexing processes; g) mass-interception data and metadata gained by 'snooping' machinery; h) data from far away extra-terrestrial objects captured by potent tools, such as the Large Synoptic Survey Telescope.[105] In addition, they can result from the process of datafication, that is to say the 'ability to render into data many aspects of the world that have never been quantified before' (Cukier and Mayer-Schoenberger 2013), such as friendships in the form of *likes*. Governments (mainly gathering demographic information or through surveillance activities) and corporations (collaborating in surveillance operations or

[103] Digital objects or artefacts 'differ from physical objects and other cultural records' in that they are 'editable,' that is, they can be modified 'continuously and systematically' (Kallinikos, Aaltonen and Marton 2010, 3). Digital objects are also interactive 'in the sense of offering alternative pathways along which human agents can activate functions embedded in the object or explore the arrangements of information items underlying it and the services it mediates' (ibid., 3). Digital objects are possible to access and to modify by means of other digital objects, 'as when picture–editing software is used to bring changes to digital images' (ibid., 4). Finally, digital objects are distributed and are thus 'seldom contained within a single source or institution' (ibid.).

[104] Elaboration by the author.

[105] 'The LSST is a new kind of telescope. Currently under construction in Chile, the LSST is designed to conduct a ten-year survey of the dynamic universe' (Large Synoptic Survey Telescope 2016).

collecting data about their clients' behaviour) control the biggest data repositories. But people can, and do, 'liberate' data or create their own datasets, as Sampedro suggests:

> Databases do not longer belong to the bureaucracies. There is no need to wait for them. If bureaucracies do not publish them, the public will liberate them. If they are censored, internet users will complete the missing data. If they cannot be analysed, others will change their format so they can be searchable and visualisable. If they do not exist, they will be generated collaboratively. The public debate does not depend on who registers, stores and (de)classifies data anymore. That is, of who holds the power (2014, 211).[106]

In other words, Sampedro thinks that the asymmetries described by Andrejevic can and will be straighten by the ability of people and organisations to generate their own datasets and analysis about both the Aliertas of this world (i.e. the 'Luxemburg leaks'),[107] and the very poor (i.e. the ODI's Illegal, Unreported, and Unregulated Fishing project).[108] Actually, Letouzé and Sangokoya prefer to talk about the 3 cs of big data (2015, 4). The c of *crumbs*, i.e. those 'digital bread crumbs' or those 'digital translations of human actions and interactions passively emitted

---

[106] 'Las bases de datos ya no las proporcionan las burocracias. No hace falta esperarlas. Si no las publican, el público las liberará. Si se censuran, los internautas completarán los datos que falten. Si no pueden analizarse, otros cambiarán el formato de los documentos para realizar búsquedas y visualizaciones. Y, si no existen las bases de datos, se crearán colaborativamente. El debate público ya no depende de quién registra, almacena y (des)clasifica. Es decir, de quien detenta el poder' (ibid.).

[107] Luxembourg Leaks is a financial scandal exposed in 2014 by the *International Consortium of Investigative Journalists*, which had many ramifications in Spain. It is based on confidential information about Luxembourg's tax rulings established by PricewaterhouseCoopers from 2002 to 2010 to the benefits of its clients, among them many Spanish corporations. Eighty journalists from different media organisations were drawn in collaboratively assessing 28,000 pages of document. It generated a scandal about tax avoidance schemes, and contributed to the implementation of measures designed to mitigate tax dumping and regulate tax avoidance schemes.

[108] As I work on this dissertation, I am also involved in the research and publication of a big data-based report on IUU fishing in Western African countries. IUU fishing has huge development impacts, both local on fishing communities and national and regional on poor countries' economies along the Western coast of Africa. The data for this project has been independently collected and analysed in spite the veil of secrecy that covers most fishing operations, agreements and joint ventures.

and captured by digital devices;' the c of *capacities*, i.e. tools and methods to collect, aggregate and analyse data; and the c of communities, i.e. 'all those involved in generating, governing and using data, including data producers, end users, policymakers, experts, privacy advocates and civic hacker communities' (ibid.).

Based on the above mentioned literature and my own observations, big data can be considered as a wealth of *digital objects* and user-generated online content resulting from users' digital activity, from mass-interception and metadata, as well as from the datafication of human activity, which are so large in volume, can be generated, analysed and computed with such velocity, are so varied in form, have so much economic and development potential when analysed and rendered useful, and show so much accuracy and complexity that can be considered really *big* and, therefore, can only be extracted, managed, analysed, stored and exploited using new infrastructures and methods. Big data, through machine learning, can produce new insights, new knowledge and 'new forms of value' (Mayer-Schönberger and Cukier 2013, 5). The analysis and visualisation of big data often produces 'small data,' which can be easily comprehended and handled by humans.

3.3 Data, information and knowledge

To understand how knowledge is produced from data, researchers from different disciplines have resorted to 'the knowledge pyramid' –the DIK(WE) pyramid of Data, Information, Knowledge (Wisdom and Enlightenment)—, also known as the 'information hierarchy,' which sometimes includes 'wisdom' and 'enlightenment' as fourth and fifth tiers, and sometimes it does not. I will focus, above all, on the first three elements of the pyramid, as the

higher you go, the vaguer concepts become. The DIK(WE) model[109] is often used in literature about information and knowledge management, but there is no scholar consensus about the definitions used in the model or the processes that transform the elements of a lower tier into the elements of a tier above them. Zins says in an extensive analysis of the conceptualisations of data, information and knowledge that the pyramid's components refer to no less than five different models, depending on whether they are conceived of as internal (subjective) or external (universal-collective) phenomena, or both. In Zins's study, subjective and objective 'are not related to arbitrariness and truthfulness, which are usually attached to the concepts of subjective knowledge and objective knowledge' (2007, 486). For him, the distinction between subjective knowledge and universal knowledge 'differs from the distinction between private knowledge and public knowledge' as well (ibid., 486-487). Private knowledge is the individual's intimate knowledge, while public knowledge refers to thoughts known to other people as well. The most common model, according to Zins, sees data and information as external phenomena, while knowledge is seen as internal phenomena. 'It underlies the rationale of the name *Information Science*; that is, Information Science is focused on exploring data and information, which are seen external phenomena. It does not explore knowledge, which are seen as internal phenomena' (ibid., 489). So, how do these components relate to each other? Let us start with the first step. Data can be conceived as signs or symbols, which in some cases represent stimuli or signals (ibid.). What is their value and how can they be used? Data are 'of no use until… in a usable form,' according to a review of textbooks published by Rowley and Farrow (2008, 5). Zeleny uses the term 'know-nothing' as a metaphor to describe data's purposes (1987, 72). Data are, therefore, of no use *per se* and have no meaning or value until put in a usable format and

[109] The pyramid is a model for representing purported structural or functional relationships between data, information, and knowledge.

analysed in the context of other data. For Rowley and Farrow, data are 'discrete, objective facts or observations, which are unorganised and unprocessed and therefore have no meaning or value because of lack of context and interpretation' (2008, 6). However, since facts relate to objective realities and can be verified, such definitions would disregard false, biased, inconsequential data –which are frequently included in large datasets— from the DIK(WE) model. In fact, the process known as Garbage In, Garbage Out (GIGO)[110] would not be taken into account. Besides, some data and metadata[111] which were considered *garbage* and nonsensical in the past can be processed and analysed now with big data infrastructures. Therefore, I do not consider data in big data sets as *only* facts. Data can be deemed as signals too. In the subjective domain, data are conceived as 'sensory *stimuli*, which we perceive through our senses' (Zins 2007, 480), or as 'signal readings,' including 'sensor and/or sensory readings of light, sound, smell, taste and touch' (Liew 2007, 4). Harmon defines data as 'one or more kinds of energy waves or particles (light, heat, sound, force, electromagnetic) selected by a conscious organism or intelligent agent on the basis of a pre-existing frame or inferential mechanism in the organism or agent' (ibid., 483). The meaning of sensory stimuli may be deemed, then, as subjective data (as opposed to 'objective data'). According to Harmon, information is 'the meaning of these sensory stimuli,' or data, through empirical perception (ibid., 487). But data can be thought of as symbols too. Bellinger, Castro and Mills say data are 'symbols' (2004, 1),[112] which are understood by Zins as 'sets of signs, which represent empirical stimuli or perceptions' (2007, 9). Data, in this sense, are 'recorded (captured or stored) symbols,' including 'words (text and/or verbal), numbers,

[110] Garbage in, garbage out (GIGO) in the field of computer science refers to how computers, since they operate by logical processes, will automatically process accidental, even senseless, input data ('garbage in') and produce superfluous, often nonsensical, output ('garbage out').
[111] 'Metadata are data that describe other data. Meta is a prefix that in most information technology usages means "an underlying definition or description"' (Rouse 2014).
[112] Based on Ackoff (1999).

diagrams and images (still &/or video), which are the building blocks of communication.' Their purpose 'is to record activities or situations, to attempt to capture the true picture or real event,' such that 'all data are historical, unless used for illustrative purposes, such as forecasting' (Liew 2007, 4). In conclusion, data (either structured or unstructured) is the basic material –facts, but also signals and symbols— with which information is produced through their analysis. Big data infrastructures can render useful more and more data and metadata that were previously considered *garbage*. But again, data cannot be 'raw' –Gitelman clearly alludes to that in this manner in the title of her book *"Raw data" is an oxymoron* (2013) —, and that they are not valuable or meaningful in themselves (Trottier 2014, 52). Hogan, interviewed for this dissertation, says that 'I think it's important to think critically about data and big data in particular. I don't think numbers speak for themselves. I think it's important to understand the context, source, method of acquisition, and underlying politics of any data set as well as its representations and visualisations.' Haque speaks out 'quite strongly against the idea of data as a fetish,' since data are'(not) a particularly useful goal,' but the by-product of many other processes. They are 'not inevitable' either. Talking about the DIK(WE) model, he finds 'so ridiculous to think that somehow these things lead to the next. Because the point is data is [sic] crafted, it is not plucked from the air, nor is a neutral substance... Someone has decided what to measure, where to measure, how to measure, why to measure, why not to measure something else. There have been so many value judgements that go in any data capturing exercise' (2016).

Information is a step higher in the ladder towards wisdom and enlightenment. In an article entitled 'From big data to big wisdom,' McKeever notes, 'stories may get aggregated into mineable datasets, but the work of turning this data into "evidence" — that is information which is helpful in forming a conclusion or hypothesis — involves something more' (2013, 1). Data-

based information paints a bigger picture than simple data; it is organised data with relevance, purpose and meaning. It may indicate a trend in the environmental patterns over time or perhaps show a correlation between the sales of a given product and a social development. Information is found 'in answers to questions that begin with such words as who, what, where, when, and how many' (Ackoff 1999, 160). For data to become information, it must be contextualised, categorised, calculated, corrected and condensed (Davenport and Prusak 2000, 3). Technology is vital in turning data into information, particularly with large amounts of data across multiple sources and formats. In the context of DIK(WE), information is differentiated from data in that it is meaningful and useful, and in that information is inferred from data in the process of answering questions, thereby making the data useful for 'decisions and/or action' (Liew 2007, 1-4). Cleveland noted that information is different than material things in that it can be shared, it is expandable, compressible, substitutable and transportable, and drips profusely and naturally (1982, 34). The difference is one of order of complexity. Cleveland explains the hierarchy like this: 'Information is horizontal, knowledge is structured and hierarchical, and wisdom is organismic and flexible' (1982, 34-35). A step further up the ladder, knowledge is related to understanding. The knowledge possessed by each individual is a product of his or her experience, as well as lessons learnt, and embraces the norms by which he or she takes in and assesses new inputs. Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or repositories but also in organizational routines, processes, practices, and norms (Liew 2007, 2). Knowledge is action; it is appropriating; it is impactful information. In fact, the level of knowledge has a massive impact

in the network society, which is an incredibly interactional, knowledge worshiping society. Knowledge has been understood in different ways by various authors. In *Knowledge and human interests*, Habermas identified three types of knowledge: empirical-analytical, hermeneutic and critical,[113] which respond to three areas of human interest. An empirical interest can best be attended through science or technology, which fall under the sphere of analytical knowledge; interests that have to do with putting ourselves within an historical or cultural context are best tackled by hermeneutically-derived knowledge (interpretative); and our interest in freedom (the emancipatory interest) links with the critical element of our knowledge-base (1971). 'Examples of these three areas of knowledge would thus be the natural sciences or mathematics in the analytical-empirical sphere, the social sciences or humanities in terms of hermeneutic-historical aspects, and political theory or psycho-analysis as a means of conceptualising our critical-emancipatory aspirations' (Terry 1997, 3). For Terry, the first two of these types of knowledge correspond to Ryle's distinction between 'knowing that' and 'knowing how,' Ryle relates to Habermas's analytical and hermeneutic knowledge respectively (2009, 14-47). Terry extends this comparison to include a third category of 'knowing why,' which would correspond to the critical sphere. Computerised data and information can be found easily through the internet, but 'for the data to be useful — to come to life, as it were — they have to be linked to another rung or category of data. The result is knowledge' (ibid.), which is more difficult to transmit. For Toffler, 'the recognition that no knowledge can be complete, no metaphor entire, is itself humanising. It counteracts fanaticism' (1984, 11). This way, human beings can err, especially when doing large-scale syntheses. In fact, to ask ever larger questions is 'to risk getting it wrong, but not to ask them at all is to constrain the life of understanding (ibid., 11). Zeleny defines knowledge as 'know-how' (i.e., procedural knowledge) (1987, 72). Further, he declares that knowledge is not

[113] A model which was reminiscent of the three Kantian spheres of science, aesthetics and morality.

abstract, that it is action instead, 'so it can only be manifested in a specific, individual context of space and time' (ibid., 64). In that sense, knowledge can be a matter interpretation and justification.

Post-material elements, particularly knowledge, are valuable in a society (western society) where all other material needs have been satisfied. And today's economy has never been more dematerialised; big businesses go public because there is more money to be made in the stock exchange than producing services and products; they delocalise their factories where salaries are cheaper, taxes, weaker and social rights, frailer. This logic has transformed everything. For example, the simple transfer of information at university is not valued as it was before: what is important now is the conditions in which that knowledge can be transferred and the competences and skills one gains at university. And the universities that have not caught up yet are outdated, generating a sort of institutional divide. 'Our students have changed radically. Today's students are no longer the people our educational system was designed to teach' (Prensky 2001, 1). According to Prensky, 'the single biggest problem facing education today is that our Digital Immigrant instructors, who speak an outdated language (that of the pre-digital age), are struggling to teach a population that speaks an entirely new language' (ibid., 2). Precisely language (one of the most powerful and important institutionalisations in life and, together with social interaction, the centre of the construction of reality, as Berger and Luckmann would say)[114] is transformed though this interaction. In the information society, the key value is knowledge,

[114] *The Social Construction of Reality* is one of the most important and influential works in sociology. In it, Berger and Luckmann propose a theoretical basis for the sociology of knowledge, inspired in part by Schütz's phenomenology (1991). Their fundamental thesis is that reality is constructed socially and that sociology must analyse the processes in which that process occurs. Reality is understood by these authors as a series of phenomena, external to subjects, and knowledge is information about the characteristics of these phenomena. Reality and knowledge are found intimately linked in a process in which a body of knowledge about a given phenomenon is established socially as a reality. This work introduced the notion of social construct. Like what happens with Habermas, these authors have been criticised for being too idealistic and valuing meaning and representations before reality.

and in the knowledge economy, the capitalisation of information for teleological purposes generates wealth. Whereas the much paraded *value* of data does not deserve much consideration, the value of knowledge has always been a central topic within philosophy. An important question to address, which can be traced right back to Plato's *Meno*,[115] is: what is it about knowledge that makes it more valuable than mere belief? According to the *Stanford Encyclopaedia of Philosophy*, 'interest in this topic has re-emerged in recent years, in response to a rediscovery of the *Meno* problem regarding the value of knowledge' (2007). Recent discussions of the value of knowledge have begun to explore the possibility that it is not knowledge which is the distinctively valuable epistemic standing, but rather a different epistemic standing altogether, such as understanding (ibid.).

As said before, some DIK(WE) models include wisdom and enlightenment as well. For Zeleny, wisdom is 'know-why' and enlightenment 'know-yourself.' Enlightenment in this context largely means enabling clarity of perception, a sort of individual awakening to ultimate truths of life. In English, the term can have spiritual connotations; in fact, it can mean spiritual enlightenment (*Erleuchtung*) and intellectual enlightenment (*Aufklärung*). But as said, further up the pyramid, the more difficult it is to find consensus or to keep away from subjectivity. All conceptualisations that include these two last stages in the pyramid connect knowledge with enlightenment, via wisdom.

3.4 Mobile ubiquity: Data 'on the go'

Mobile technology is one of the most important ways in which data are being generated and used for surveillance purposes; it is also a technology that has transformed the way people

[115] The question why knowledge is idiosyncratically valuable has a significant precedent in Plato's *Meno* (Plato 2008), in which Socrates, one of its protagonists, raises the question of why knowledge is more valuable than mere true belief. Initially, we might note that knowledge seems to be more practical than true belief, but, as Socrates notes, this claim is far from obvious on closer inspection. The issue of why knowledge is more valuable than mere true belief is known as the Meno problem.

mobilise and protest, reasons why I am devoting a section to this matter. Walking or driving a hundred metres down any street or road in any city is almost inevitable that these movements are being registered by cameras, recognition systems, satellites, sensors under the asphalt or radiofrequency identification gadgets. We live in cities where urban objects, buildings, distribution systems, even pieces of urban furniture are receptors and gatherers of data, and are connected with networks. 'The 21st century city is a city under surveillance' (de Vicente et al. 2014, 7). We are *human antennae*, we carry a devise that continuously releases signals –the *smartphone*—, which connect wireless in three forms: mobile broadband, Wi-Fi and Bluetooth. The internet of things, combined with mobile technology and conventional internet, knits a compact coat of connections over everybody. Because of their key role in digital mobilisation and activism, in the generation of big data and eventually in the birth of reactive and proactive data activism, they are worthy of an in-depth examination.

By 2020 there will be 4.3 connected devises per person on the planet, when 33 billion of these gadgets are expected to be in use, with the Internet of Things (IoT), smartphones and wearables[116] driving the next growth wave (Strategy Analytics 2014). The smartphones, with more capabilities than any device before,[117] have become the quickest-selling devise in history, outshining the simple mobile phones that preceded them (The Economist 2015a). That of the smartphone is not a success story of a consumer product: it is a story of business far-sightedness

[116] Wearable technology, tech togs or fashion electronics are clothing and accessories incorporating computer and advanced electronic technologies capable of connecting to other devises.

[117] As *The Economist* says, the smartphone is the real personal computer (The Economist 2015b). They have the features of a miniature computer, and can also place and receive calls and connect to the internet. Though there an industry-standard definition of smartphone does not exist, the simplest way to tell a simple cell phone apart from a smartphone is to determine whether or not the device has a mobile operating system, which is much like what is powering a personal computer. Smartphones can be powered by Google's Android, Windows Mobile, iPhone OS, Symbian OS, Palm's WebOS and Linux, mainly (RIM disappeared with the BlackBerry).

as well. From 2009 to 2013 the mobile industry invested US$1.8 trillion on improving its infrastructure around the world (Bezerra et al. 2015). Download speeds have also increased by a factor of 12,000. Along with Wi-Fi[118] in homes and offices, this has made it feasible to add the power of external data-centres to the phones' own computing capabilities. Amazon Web Services, the world's biggest provider of cloud computing,[119] says it is now adding as much server capacity every day as its e-commerce parent required to run its entire global infrastructure ten years ago (The Economist 2015b). A key driver in the growth of mobile technology has been the prices of mobile services. Since 2005 the cost of delivering one wireless megabyte has plunged from US$8 to a few cents. Another driver has been the dropping average price of a smartphone as a hardware product, which has also been collapsing since 2008. The falling prices of smartphones as appliances are a huge economic engine, but they have a more sinister face: the global trade of cobalt that put thousands of children –as young as seven— to work and die in countries like Congo (Amnesty International 2016a, 29). Cobalt is used in lithium-ion batteries common in consumer electronics, especially in portable devices such as smartphones, since they have a high energy density. 'Millions of people enjoy the benefits of new technologies but rarely ask how they are made. It is high time the big brands took some responsibility for the mining of the raw materials that make their lucrative products' (Amnesty International 2016b, 1). Hogan, among many others, wonders about the impacts of technology in the environment as well. 'I think Big Tech often gets looked to solve environmental problems, with solar and wind technologies for example, or for its capacity to crunch large data sets, to then visualise impacts

[118] A local wireless networking technology that allows digital devices to connect to the network and among themselves. It is also defined as a wireless local area network or WLAN. The commercial name 'Wi-Fi' comes from 'Wireless Fidelity.'

[119] Cloud computing is internet-based computing system that provides shared, ubiquitous resources, data and information to a pool of computers and other devices on demand, reason why it is also called 'on-demand computing'. Cloud computing and storage offers users increased storage and process capabilities based on economies of scale.

and predict disasters, etc., but it's less often looked at for its ecological impacts' (2015, 1). It is a proven fact that the deterioration of the environment can impact directly and negatively on people's food security, access to water and capacity of escape poverty, among other development areas (Gutiérrez, McFarland and Fonua 2014). This generates a paradox. These are the issues proactive activists care and campaign about, many times using technology that harms people and the environment.

Powered by their access to technology and their dynamism, and because of their population density, cities are called places of extreme connectivity: they have become 'informational cities' (Castells 1989). And the human race increasingly lives in cities;[120] an unparalleled process of urbanisation accompanies the development of mobile technologies. In Western Europe, we exhibit some of the highest rates of mobile usage, with 129% of penetration (Kemp 2014). This means that some people carry more than one of these devises in her or his pocket. But mobile technology is not only an urban phenomenon. Precisely what makes it so popular is that it can be used from remote locations where no conventional facilities are in place. In fact, global penetration of mobile technology is 93 (ibid.). Kemp also says that the mobile subscription penetration is, more surprisingly, 67% in Africa – in comparison with 13% of internet penetration and 5% of social penetration in the same region (ibid.). Even in places with little access to technology, millions of people own a cell phone. According to another UN study, 'more people worldwide have mobile phones than toilets' (UN News Centre 2013). Some rural, remote areas of the world have leap-frogged into post-modernity by accessing mobile technology. In spite of the grim picture painted by Amnesty International's report on child labour in Congo (2016a), Africa is swiftly turn out to be a 'mobile-connected continent,' where 'mobile

[120] Today, 54% of the world's population lives in urban areas, a proportion that is expected to increase to almost 66% by 2050, according to a United Nations report (United Nations 2014, 1; UNICEF 2012).

phones are often the only technology available' (de Simone 2015). Places where no commercial interests were strong enough to provide fixed-line telecom infrastructure are now experiencing a 'technological explosion' thanks to mobile technology 'changing people's lives in several ways to an even greater extent than in developed countries' (ibid.). This is important because mobile technologies are facilitating data activism in Africa as well. *The Economist* has called this explosion of mobile technology 'the dawn of the planet of the smartphones' (2015a). The fact is that practically everyone on the surface of this Earth carries a mobile phone, with the probable exception of the children that work in conditions of virtual slavery in the cobalt mines. This has positive and negative implications: mobile technologies have enhanced the capacities for activism, but they are also a great producer of big data that can be used in massive surveillance.

Who would have thought only a few years ago that we would be hooked to the phone, as if it were a McLuhanian extension of ourselves, continuously checking new entries in several apps? By their nature, seminal technologies challenge society, while society changes them as well by subjecting them to different uses than the ones they were originally conceived for, as Fischer notes of the land-line telephone (1994). Looking at these transformations could present some insights about what to expect of big data infrastructures. There are different approaches to study of technology and society. Looking at the coupling 'technology and society,' some put 'technology' first and others go the other way around. Others, like Williams, reject both technological determinism and the notion of a 'determined technology' as an equally one-sided version of human development (1990, 123). I am going to review some technological deterministic approaches as well as a critique of determinism, because both views, though antagonistic, offer interesting explanations that could illuminate the way big data infrastructures are affecting us, and how we use and influence them too. Let us start with technological

determinism, which focusses on the social transformations brought about by technology. In the sixties, McLuhan pointed out the great social transformations derived from technological innovation in media (1994). In his much quoted book *Understanding Media: The Extensions of Man*, McLuhan coined the phrase 'the medium is the message,' and described 'the personal and social consequences of any medium' as 'an extension of ourselves' (ibid., 6, 9). The result of new medium is a 'new scale that is introduced into our affairs by each extension of ourselves or by any new technology' (ibid.). That is to say, the form of any medium embeds itself in the message, generating a symbiotic relationship by which the medium influences how the message is received and understood. McLuhan proposed too that media themselves, not the content they carry, should be the focus of study. Kostelanetz explains explain this notion:

> Electric modes of communication —telegraph, radio, television, movies, telephones, computers— are… reshaping civilisation in the 20th century. Whereas print-age man saw one thing at a time in consecutive sequence –like a line of type— contemporary man experiences numerous forces of communication simultaneously, often through more than one of his senses (1967, 2).

Has any device previously been more an extension of ourselves than the smartphone? The answer is probably no. Wearable technology has that potential as well, but it has not been realised so far. *Understanding Media: The Extensions of Man* seems written to explain smartphones and their new codes. Abbreviations such as ';S' (meaning gentle warning) are a clear example of the medium influencing how the message is formulated, received and understood. Technological artefacts can influence the course of history as well. According to McLuhan, television helped bring about the end of the Vietnam War, which is considered to be the first 'television war' by scholars and journalists alike. His interview with the *Montreal*

Gazette in is often quoted saying: 'Television brought the brutality of war into the comfort of the living room. Vietnam was lost in the living rooms of America –not on the battlefields of Vietnam' (Misiroglu 2002, 165). Without being a determinist, Habermas blames television and mass media for the deterioration of the public sphere too. The technological determinism approach to the question of human development is not unprecedented. McLuhan acknowledges his debt to such works as Innis's *The Bias of Communication* (2008 [1951]) among others. In this book, Innis describes the social changes derived from competing media in Ancient Egypt, where the use of hieroglyphics and stone was replaced by the papyrus and the brush as 'a new and more efficient medium' (ibid., 2). This change disturbed the 'monopoly of knowledge' and brought about the emergence of the profession of scribes. 'The social revolution involved a shift from the use of stone to the use of papyrus and the increased importance of the priestly class imposed enormous strains on Egyptian civilisation and left it exposed to the inroads of invaders equipped with effective weapons of attack' (ibid., 2). Reinterpreting McLuhan's vision of the end of the Vietnam war, mobile technologies may have been, not only instrumental, but decisive in overthrowing governments and breeding revolutions.[121] They also have the potential to support data activism. Nevertheless, I think that in all cases it is a conjunction of facts that make it possible for people to revolt against authority. In any case, in the McLuhanian universe, cell phones 'reversed' into smartphones when pushed to their technological limit, and smartphones are now part of our *retribalisation* process (1964, 380): they have taken us out of our individualist bubbles and returned us to a tribal village on a global scale by making public interaction possible (instead of the private conservations offered previously through cell- or traditional phones). First Innis, followed by McLuhan, and more recently Rheingold, among others, have stablished an association between different media and certain mental and social

[121] See Annex II for examples of cyber and mobile-based mobilisations.

settings. Innis believed that the social, cultural, political and economic developments of each historical epoch can be associated directly to the technology of the means of communication of that period. In this sense, ICTs acquire dynamic life of their own and are seen be as a key driver of social phenomena. Media and technology determinism has been defined as a perspective that places technological innovation as the main causal factor determining social change. In my view, technology does not explain all social change alone: context, social movements, social differences conditioning access to it and other social factors are crucial too. While not in agreement with that school of thought, however, examining how technology has moulded society at some points in time, and generated new power structures as well as behaviours, is interesting. After all, big data are both a shaping force and expressions of our society. Innis' priestly class is to papyrus what today's techno-bureaucracies to big data; only the hold that bureaucracies have on data is not as steady and unyielding as priests' hold on papyrus in ancient Egypt. As Sampedro suggests, people will *occupy* big data if techno-bureaucracies do not relinquish them. Interestingly Ogburn says the impact of a technological invention produces a chain reaction in a way similar to a billiard ball, where the effect is at the same time cause of another effect (1949). The same happens with how different parts of society interact with others; change in one part produces a chain reaction and reaches other parts of society (that is, in part, how data activist tactics are copied and adapted from one place to another). For some determinists –for whom technology is an external force introduced into a social situation—, as a technology stabilises, its design tends to dictate users' behaviours, and progress is a cumulative process. The doctrine of progress –born in the Enlightenment—[122] says, basically, that scientific advancement is a governing force in society, that social problems can be solved by technological evolution (or

[122] Adam Smith was the first to formulate a systematic doctrine of progress in the 18th century.

revolution), and that 'you can't stop progress,' which implies that 'we are unable to control technology' (Green 2002, 8). But not everybody agrees, I included. In *Why Technology Can't Save Us*, Huesemann doubts that advanced technology will rescue us from social, environmental and economic ills. Negative unintended consequences of science and technology are inherently unpredictable; *techno-fixes* do not offer lasting solutions; and technology, in the presence of continued economic growth, does not promote sustainability but instead hastens collapse (2014). Adorno goes even further believing that total destruction could be triggered from technical innovation as well (2004, 26). The relevant question here is: technology to do what?

It is also interesting to observe technology from a critique of technological determinism that puts society in the centre. Fischer studies the influence of the telephone in society from this perspective. As with many technologies, initially the industry that developed telephones did not imagine how it would be used when it first became integrated into the private lives of average people. Women, says Fischer, were not supposed to have much use for it.; but the industry soon learnt to promote its use among women when it realised the business potential (2014, 25). 'People typically turn new technologies into devices for doing what they have always wanted to do. And people like to stay in touch… (And that is how people), especially women, a century ago, turned two new technologies marketed for other purposes, the telephone and automobile, into "technologies of sociability"' (ibid., 25). Williams ends up describing technological development as a complex process in which not one force explains totally an outcome.

> We have to think of determination not as a single force, or a single abstraction of forces, but as a process in which real determining factors —the distribution of power or of capital, social and physical inheritance, relations of scale and size between groups— set limits and

> exert pressures, but neither wholly control nor wholly predict the outcome of complex activity within or at these limits, and under or against these pressures (1990, 123).

In a nutshell, he proposes that social movements and people influence technological processes (2005). His contribution to the Marxist critique laid the fundamentals for the 'cultural materialist' approach, in the tradition of the Frankfurt School's critical theory. From a critical theory of technology perspective, Feenberg says that the degradation of labour –as well as of many environmental, educational and political systems— is rooted in 'the antidemocratic values that govern technological development' (2002, 18). I feel closer to this approach, but it is somehow partial too. Our access to user-friendly, cheap mobile technology depends on thousands of people, many of them children, working in slavery conditions in countries like Congo. If we paid decent salaries to professional adults, with social benefits, for the same job, mobile phones would be accessible to the elite, at least for a while. However mobile technology has empowered others, like the demonstrators summoned in 2011 in Zuccotti Park, New York, via mobile phones and other technologies, to protest precisely against social and economic inequality. Reactive data activists, although far from being the majority of people, do seek alternatives to the values embedded in the dominant technologies, even creating their own, closed-doors technologies and communication networks. A new technology does not necessarily transform people into passive users or determine their behaviour. Castells talks about 'mass self-communication,' or the use of new media for messages that are able to reach masses, who can, in turn, respond to those messages in a way where receivers are also emitters. There are many examples, too, of new technologies that have been transformed by their users. The computer is one. Thomas Watson, president of IBM, is supposed to have said in 1943: 'I think there is a world market for maybe five computers' (Strohmeyer 2008, 1). If it were for Watson, there

would still be a handful of computers based in some military compounds and universities around the world. Drones were created as military applications, but, as we will see later, communities are appropriating and adapting them for data activism. Users definitely influence technology and not always react to new inventions as their creators first imagined. A similar process was observed by Rheingold, who describes how people turned cell phones into instruments of massive social mobilisations (2002, 157; 2005, 226). However, in this dissertation, I dwell on how people use big data infrastructures for social change. It is not clear to me whether cell phones have evolved into smartphones because of the new ways in which users employed them (that is, a bottom-up process in which producers adapt their product to what users want of them),[123] or whether smartphones have been imagined and improved by an elite of expert *technorati* –led by Apple's Steve Jobs– so as to transform how people live (in a top-down process). There is probably a bit of both: user experience, foresight and ingenuity, and market forces. Talking about the influence of technology in international relations, Ogburn says that 'when it is found that technology affects international relations, it is not to be implied that no other factor is of any influence. Several causes often exist, of which only one is a new invention' (1949, 17). The problem is that what is new is not one factor, but many. Paraphrasing Castells again: 'Chips and computers are new; ubiquitous, mobile telecommunications are new; genetic engineering is new; electronically integrated, global financial markets working in real time are new' (2010, 372). However, the matter of whether current exponential technological change is to be considered an evolution or a leap is not so relevant. What is interesting here is the use of mobile technologies for social change. The mobile phone has been instrumental in facilitating

[123] Nowadays, UX, or user experience, is everything when it comes to commercialising new devices. User Experience (UX) refers to a person's feelings and attitudes about using a product or service –which includes the experiential, affective, meaningful and valuable aspects of human-computer interaction. It also includes a person's perceptions of utility, user-friendliness and efficiency. UX is dynamic as it is modified over time due to changing usage circumstances and changes as well as the wider usage context.

new forms of social mobilisation and, in some cases, making social insurgency possible. That is certainly something that was not anticipated by either market forces or mobile phones makers. Castells praised the personal computer as a tool for individual independence and freedom (2001, 3). But size and connectivity make the smartphone the truly first personal computer, as it is easy to carry, and has even more capabilities than the original PC. 'The phone takes the processing power of yesterday's supercomputers —even the most basic model has access to more number-crunching capacity than NASA had when it put men on the Moon in 1969— and applies it to ordinary human interactions' (The Economist 2015a). This does not pertain only to ordinary interactions, but to extraordinary interactions –including social mobilisation— as well. The so called 'smart mobs' behave efficiently because of its increasing network links, enabling people to connect to information, to perform a task allowing a form of social coordination (Rheingold 2002, 1). Mobilising has become much cheaper, centralised organisation less necessary, because mobilised crowds emerge, behave and adapt as networks: they are flexible and scalable, and they survive better than hierarchical structures. Networks adapt to new environments, retaining their goals, 'while changing their components' (Castells 2009, 23). They can lose parts, and continue moving. Mobile technologies are changing the ways in which people create awareness, share information, organise and mobilise, showing, as Schmidt and Cohen sustain that technology and data offer 'the potential to be a substantial positive… More people will fight for privacy and security than look to restrict it, even in the most repressive parts of the world (2013, 48-49). *The Economist* explains the power of mobile phones in the face of authoritarianism like this: 'The same phones that allow governments to spy on their citizens also record the brutality of officials and spread information and dissenting opinions… A device that hands so much power to the individual has the potential to challenge authoritarianism' (2015a).

Mobile technology has engendered a greater fear over privacy as well. Reactive data activism is a response to it. The smartphone turns you into a human antenna and the person next to you into a potential publisher of your most private moments.

> (For many, smartphones are) miniature versions of the tele screens in George Orwell's 1984, omnipresent tools, which allow the thought police to recognise enemies of the state… Incredibly enough, around the world, people are rushing to buy machines and download apps through which they can be monitored at previously impossible levels of closeness (The Economist 2015b).

But data collected this way does not stop there. Many app vendors know 'a great deal about you,' and sell your data without proper disclosure, since mobile-privacy policies 'routinely rival *Hamlet* for length' (The Economist 2015a, 2). As said, mobile technology is not only a revolution tool but a big generator of data. This takes us to the next section on how big data and their infrastructures are amplifying surveillance capabilities of governments and corporations to create a new *panopticon* and computational politics, but also the possibility of a new digital public sphere and data activism.

3.5 Space of flows: The culture of real virtuality

Another consequence of technological revolution is a change of time and space. To explain this phenomenon, Castells creates the concept of 'space of flows,' which plays a central role in his vision of the network society and has been mentioned previously when referring to the post-modern panopticon. The space of flows allows for simultaneous social practices without 'territorial contiguity' (contrary to what happens with the pre-network society 'space of places') (1992, 126-172). This is a space made up of a technological infrastructure of information systems, telecommunications and transmission lines, which is the material support of

simultaneous, ubiquitous social practices. The 'space of flows,' the material and immaterial elements of global information which networks use for the instantaneous, long-distance coordination of the economy, and their dynamic interactions with the digital society (ibid.). The space of flows has weighty implications for Castells: post-modernity and the end of history. 'Because the spatial manifestation of the dominant interests takes place around the world, and across cultures, the uprooting of experience, history and specific culture as the background of meaning is leading to the generalisation of ahistorical, a cultural architecture' (2010, 449). Post-modernism expresses the new dominant ideology: the end of history and the supersession of places in the space of flows (ibid.). In post-modernity, time and space have shrunk in physical terms. Big data infrastructures defy even Newton's laws, since the *machine* can anticipate our thoughts and desires even before a human being formulates them. This is the world of 'predictive analytics,' which forecasts what might happen in the future with an acceptable level of reliability and can allow you look for 'patterns that are indicative of current or future behaviour' (Gibbs 2014). An example is Amazon's 'anticipatory shipping' algorithm-based scheme. This system allows Amazon to ship products before anyone even place an order to the destination geographical area without completely specifying the delivery address at time of shipment. The final destination is defined *en route*, when the actual order confirms the prediction made by the machine (Ulanoff 2014). What took hours, and even days, today is immediate, simultaneous or can even happen before it does.[124] Ubiquitous, instant, real-time communication produces a tightening and shortening. As never seen before, contacts and friendships can be built and destroyed in seconds, creating a 'culture of the ephemeral' (Castells 2010a 386), what Bauman

[124] Ryft, a big data outfit, launched a new analytics platform that claims to be '200X faster than conventional hardware' in analysing historical and streaming data simultaneously. It promises to 'provide actionable business insights by simultaneously analysing up to 48 terabytes of historical and streaming data at a 10 gigabytes/second or faster' (Dataconomy, 2015).

calls liquid and fluid (2007), and Lipovetsky light (2016). ‘The network society disembodies social relationships, introducing the culture of real virtuality’ (ibid.). The haste in the generation of knowledge originates new spaces of social interaction that change with extreme swiftness too. Caused by rapid technological development, changes in relationships of production, power and experience converge toward the transformation of material foundations of social life, space and time (ibid.). And new challenges emerge. Businesses have gone from producing things to producing intangible services and generating innovation, not products. The speed and complexity of economic, social and political change place individuals in a process of constant adaptation in advanced post-industrial societies. Capital escapes from time and space. ‘Capital circulates, power rules and electronic communication swirls through flows of exchanges between selected, distant locales, while fragmented experience remains confined to places. Technology compresses time to a few, random instants, thus de-sequencing society, and de-historicising history’ (ibid.). Innovation becomes a central element of survival in unforeseeable scenarios, where planning is difficult because of the uncertainly attached to these ‘liquid times’ (Bauman, 2007), and the only possible thing we can do is managing complex short-term realities. Some people cannot endure or face these challenges, and social exclusion happens.

3.6 The post-modern panopticon

As seen, big data and communications technologies can make digital public spheres possible. But as I have been pointing out, they have dark side: The rise of big data infrastructures has also resulted in the mass collection of data by governments and corporations, massive surveillance and ‘computational politics’ (Tufekci 2012), what Braman calls the *panspectron* of the ‘informational state’ (2007, 314). In contrast with old strategic surveillance, ‘the new surveillance involves scrutiny of individuals, groups and contexts… This means the ability to go

beyond what is offered to the unaided senses and minds or what is voluntarily reported' (Ball, Haggerty and Lyon 2012, 26). In the face of massive vigilance, society has experienced a 'counter response,' which is embodied in 'reactive data activism:' the practices of resistance to the threats to civil and human rights, enabled by socio-technical artefacts to work with data (Milan and Gutiérrez 2015, 127). The informational state has replaced the bureaucratic welfare state, as information has changed the nature of the nation-state and its relation with citizens. The informational state is characterised by 'shifts in the nature of power and its exercise via information policy' (Braman 2006, 1). Instead of protecting and looking after, the state now shadows and pre-empts. For Braman, 'we have passed the tipping point: While information policy is among the most ancient forms of governance, there has been a phase change —a change of state— in the extent to which governments deliberately, explicitly, and consistently control information creation, processing, flows and use to exercise power' (ibid., 1). As we have seen in previous sections, data rendered useful can be translated into power and economic gain, thus data have become increasingly important in contemporary politics and economics. An example is among Snowden's revelations: the NSA requested metadata about millions of US phone calls from Verizon in 2013, without informing its clients, under a top secret court order. (Greenwald 2013, 2). Thus collecting data is not just about counting people, it serves a purpose. According to Tufekci, the combination of big data and computational practices 'allow for massive, latent data collection and sophisticated computational modelling, increasing the capacity of those with resources and access to use these tools to carry out highly effective, opaque and unaccountable campaigns of persuasion and social engineering in political, civic and commercial spheres' (2012). And with 'the rise of big data,' governments can now focus not only on groups, but also on individuals 'as individuals.' Being able to pull apart voters on an

individual basis, and not as members of an aggregate, was the 'holy grail of political campaigns' (ibid., 9). Computational politics emerges with the 'shift away from demographics to individualised targeting, the opacity and power of computational modelling, the use of persuasive behavioural science, digital media enabling dynamic real-time experimentation, and the growth of new power brokers who own the data or social media environments' (ibid., 1). That is, with big data infrastructures, governments now have the means to purposely and systematically watch everyone on a personal basis, as well as control information generation, processing, flows and use to exercise power. This development has a profound impact on the social contract between the governed and the government that has been valid until very recently. Tufekci also examines what is left of the social contract and the public sphere. With big data and computational techniques, he notes, manufacturing consent is easier: it has moved from the 'magnifying glasses and baseball bats' of the 20th century into 'telescopes, microscopes and scalpels' of the 21st century (2012, 6). For Tufekci, the public sphere has been completely reshaped, making it less *public* (ibid.). Political communication has become 'increasingly personalised,' as it is directed to interact individually with voters. The use of big data infrastructures enables the creation of new tools, like the microscope or the telescope, 'turning unseen objects into objects of scientific inquiry and manipulation' (ibid.). This influence is exercised in different ways, including 'an exponential increase in the amount and type of data available;' the possibility of targeting and profiling particular individuals; the prospect of 'enhanced, network–based social engineering;' the immediacy and speed in which these methods can be tested and implemented granted by digital networks; and the opacity of most of these techniques and algorithms, which, in hands of corporations and governments, are 'proprietary and undisclosed' (ibid., 3). Big data analytic tools promote an even more successful — and less

transparent — 'engineering of consent' (Bernays 1947, 113) than society has ever seen before. Computational politics is 'informed by behavioural sciences and refined using experimental approaches' and testing, and is often used to profile people, sometimes in the aggregate 'but especially at the individual level,' and to develop methods of persuasion and enrolment, which can be extremely individualized (Tufekci 2012). This creates an *information asymmetry* dividing those holding the data from those whose data are being harvested. Extensive consumer data were collected 'on a regular basis' during Bill Clinton's presidential campaigns of 1992 and 1996 in the US, for example, 'to frame' the campaign's communications' (Newman 2001, 14). More recently, Ted Cruz's presidential campaign has been accused of using 'psychological data' based on research bridging tens of millions of Facebook users, gathered without their permission (Davies 2016). However, 'computational politics' introduces unprecedented qualitative enhancements to earlier efforts to manufacture consent. Previous 'data collection efforts (for example, collating magazine subscriptions or car type purchases)' entailed 'complicated, roundabout inferences about their meaning (does a magazine subscription truly signal a voter preference?),' and resulted in broad profiling. Big data provide 'significantly more individualised profiling and modelling, much greater data depth, and can be collected in an invisible, latent manner and delivered individually' (Tufekci 2012). The expansion of big data capabilities of governments and corporations dramatically amplifies these trends, and the possibility to target individuals serves to discriminate against 'members of chronically underserved communities' (Gangadharan 2012, 1). Gangadharan argues that while online surveillance activity affects all, members of excluded communities become potentially more vulnerable to the harmful effects of snooping technologies. In *Digital inclusion and data profiling*, Gangadharan examines concrete examples of commercial data profiling against members of these communities.[125] Big data,

[125] It would not be the first time. In *IBM and the Holocaust*, Black details IBM's strategic alliance with

therefore, should be scrutinised as a political process connecting questions of transparency, power, privacy and surveillance (Tufekci 2012).[126]

As with the inmates of the *panopticon*,[127] we are being gazed at, but we do not know when or where. And there is no more powerful and sophisticated internalised coercion than ignoring when and where we are being watched. There are no observable bars, chains or wardens, but the post-modern panopticon of the governmental and corporate surveillance machinery creates a consciousness of permanent visibility as a form of power (de Vicente et al. 2014, 7). The standing of this design today stems from Foucault's famous analysis of it and from the capabilities of big data-based surveillance that make it possible across the world. 'It is no exaggeration to say that nowadays there are very few moments when we are truly anonymous' (ibid.). The post-modern panopticon does not operate in the same Cartesian coordinates of time and space than the original panopticon, it operates in a space of flows. Braman calls it *panspectron* (2007, 314). In comparison with the panopticon, the panspectron gathers information about everything, all the time, simultaneously, and can manage many more subjects

---

the Nazi government and how, among other things, IBM provided the technology to identify and catalogue people in the 1930s (2008).

[126] In *The Net Delusion: The Dark Side of Internet Freedom*, Morozov describes how the internet weakens, rather than encourages, global democracy. The book's main thesis is that the internet will eventually lead to trouble-free surveillance, censorship and manipulation by authoritarian regimes, and that the initiatives of Western policy elites to stimulate democratisation by digital development are ill-advised (2011, 9). Morozov's views fall short of reality. His book came out in 2011, and he does not delve much on the possibilities of big data for manipulative purposes. In fact, he refers to 'security by obscurity' as the way in which citizens are protected by digital intrusion, since 'most of our information is simply lost in the endless ocean of digital ephemera produced by others' (ibid., 163). However, big data infrastructures permit the analysis of previously meaningless data and metadata, and the profiling and targeting of individuals.

[127] Michel Foucault used Jeremy Bentham's design, the *panopticon*, is a metaphor of modern power in *Discipline and punish: The birth of the prison*. The panopticon is a type of institutional building designed by Jeremy Bentham in the late 18th century. The concept of the design is to allow a single watchman to observe all the inmates of an institution (prisoners, patients in an asylum). The fact that the inmates cannot know whether they are being watched means that they must behave as though they are observed at all times. Foucault concludes proposing that, not only prisons, but all hierarchical structures have evolved to resemble Bentham's panopticon (Foucault 1995, 170-177).

at once, as well as target particular subjects of surveillance, who never know when, how or why they might become visible on the *panspectral* screen. In any society, the dilemma 'freedom versus security' has always been a burning issue, and the communicative constructs necessary to facilitate civic life occur in that context. Each community in each point in time has dealt with this quandary, and has created a certain space –a normative system— where freedom goes as far as security allows it, or the other way around, where security goes as far as freedom allows it (in the best cases). Bauman captures it in *Liquid Times*: 'The violent destruction of life and property incident to war, the continual effort and alarm attendant on a state of continual danger, will compel nations the most attached to liberty to resort for repose and security to institutions which have a tendency to destroy their civil and political rights. To be safer they at length become willing to run the risk of being less free?' (2007, 8-9). We need a new social contract; the problem is to find the right doses of security and freedom, in a context where surveillance is not just about networks of CCTV cameras or signals captured by the IoT, but about a systematic, planetary data gathering blanket. *Citizen Four* describes it like an infrastructure built by the NSA in cooperation with other governments 'that intercepts basically every digital communication, every radio communication, every analogue communication that it has sensors in place to detect,' which allows those in power o retroactively search any communications (Greenberg 2014).[128] Surveillance today can even be *liquid*: a data flow increasingly free and fluid,

[128] Snowden revealed in 2013 that, between February 8 and March 8, 2013, the NSA had collected about 124.8 billion telephone data items and 97.1 billion computer data items through the world, as was displayed in charts from an internal NSA tool codenamed Boundless Informant. The European data had been collected by European military intelligence agencies during operations abroad, and subsequently shared with NSA. Boundless Informant employs big data databases and cloud computing technology to analyse data assembled. The first item revealed by Snowden on June 6, 2013, published by *The Guardian*, was that the NSA, using an order from the Foreign Intelligence Surveillance Court (FISC), had required the telecommunications giant Verizon to hand over metadata from millions of US phone calls to the Federal Bureau of Investigation (FBI) and the NSA. Verizon itself was forbidden to disclose to the public either the order or the request for customer records, (Greenwald 2013, 2).

reproducing a symbiotic relationship with the liquidity visible in contemporary social, political, personal and economic arrangements, that are often short-term, frail, *fissiparous* (Bauman and Lyon 2013, 123). One theme of *Liquid Surveillance* is the need for properly ethical practices in the management of big data, since according to Lyon, there is a 'strong affinity' between governments and corporations 'particularly in relation to surveillance' (2014, 9). Namely, big data represent a convergence of business and political interests. National security is a business goal as much as a political one, and there is a revolving door between the two in the world of surveillance practices.[129] 'There is an inherent political economy around surveillance, which is used in response to commercial priorities and to reinforce a business's market position. Such surveillance bolsters processes of capitalist value creation and the extraction of material, labour from workers and immaterial labour from consumers' (Ball, Haggerty and Lyon 2012, 9).

Several post-modern phenomena intersect here, which have to do with the state-individual, security-freedom dichotomies. The layer of data discussed previously rests on the globalisation, which transforms a financial disaster in one place into a global catastrophe everywhere, and on the 'information superhighway,'[130] which makes it possible for the human suffering of others to enter our homes via electronic means. We live in an 'open society,' exposed to the 'blows of fate' (Ball, Haggerty and Lyon 2012, 7). Justice is not guaranteed, as the state is progressively abandoning individuals to their own fate. With the gradual dismantlement of the state-supported protection against existential fears and the loss of credibility of the state, pressured by the competitive market that 'erode the solidarity of the weak,' individuals have to fend for themselves (ibid., 14). Bauman mentions the documentary *The Power of Nightmares: The Rise of*

[129] See the section dedicated to the uses of big data by governments and businesses.
[130] The information superhighway was a term used through the 1990s to refer to digital communication systems and the internet.

the Politics of Fear, a 2004 *BBC* documentary series, written and produced by Adam Curtis. The films compare the rise of the Neo-Conservative movement in the US and the radical Islamist movement, claiming likenesses between the two. More contentiously, it argues that the menace of radical Islamism as a pervasive, ominous organised force is a myth manufactured by political propaganda in an attempt to manufacture consent by uniting their people under the fear of an attack. In a moment when politicians have lost their credibility, a fake enemy keeps them in power (ibid., 15-17). Since 9/11 (as the 2001 Al-Qaeda terrorist attacks against the US are known), there has been a reduction of freedoms across the world, but especially in the US. 'One of the major consequences of 9/11 has been its dreadful impact on the nation's conception and experience of freedom, its core secular value' (Patterson 2011). According to this author, after experiencing 'the greatest threats to civil liberties since the McCarthy era,' the contradictions between the reality of the enacted measures to *protect* the US against a terrorist attack and the positive perception that citizens still have of their freedoms were 'severely exacerbated' (ibid.). In the balance of today's society, the weight of security offsets the weight of freedom, in a reversal to Unamuno's motto: '(Give me) the truth before peace.'[131] In a post-truth world,[132] is this surrender of the truth, voluntary or not, conscious or not worth it? Are we all safer in spite of the loss of freedom?[133] The coat of surveillance that wraps us all, curtailing our freedoms, does

[131] 'Mi divisa es: primero la verdad que la paz.' This motto is emblazoned on the façade of the Unamuno Museum, in Salamanca.

[132] The Oxford Dictionaries Word of the Year 2016 is *post-truth*, which it is defined as 'relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief' (Oxford Dictionaries 2016).

[133] Manny think that the attacks against the Twin Towers and the Pentagon would not have happened in today's world of big data. Although there has been no attack of that magnitude, dozens of terrorist attacks and suicide bombings have happened in the world from 2001until the end of 2015. To name a few of the most tragic, only in western countries, from January to December, 2015: At least 129 people were killed by assassins and bombers in restaurants, a concert hall and a sports stadium across Paris (France) in November. Two armed gunmen moved in the Paris offices of satirical news magazine *Charlie Hebdo* and killed twelve people, including two police officers, in January. The same month, a man attacked a kosher

not seem to protect us with the same intensity. Admitting that intelligence and undercover work is necessary, I do not think the post-modern panopticon is justified. The panopticon pre-empts, assumes that everyone can be guilty beforehand, and gathers information on everyone just in case one goes astray, doing away with the presumption of innocence. Contrary to what Morozov says about being protected from digital intrusion by information overload, we are being protected by being ignorant of the panopticon. Not everyone is aware of it. But there will be a moment when everyone is, and then it will deploy its full potential for fostering massive self-censorship.

Massive surveillance generates massive *secret* big data, and big data intensify and enhance surveillance, in a perverse circle. This occurs in three main ways: One, by producing more intelligence in the interrelation of complex datasets and use of big data analytical tools. 'Existing dynamics of influence, risk-management and control increase their speed and scope through new techniques, especially predictive analytics' (Lyon 2014, 1). Two, while big data it is also about size, a qualitative transformation in surveillance practices is also 'perceptible,' due to *big* big data. 'Important trends persist – the control motif, faith in technology, public-private synergies, and user-involvement–, but the future-orientation increasingly severs surveillance from history and memory and the quest for pattern-discovery is used to justify unprecedented access to data' (ibid.). Lyon suggests, thirdly, that 'an ethical turn becomes more urgent as a mode of critique. This is so at several levels, but particularly in the kinds of ways that Snowden himself indicates through his repeated questions about "what kind of society do we want?"' (ibid., 9). According

---

market in Porte de Vincennes (France) killing four people. In February, another man opened gunfire at the 'Art, blasphemy and the freedom of expression' exhibition, in Copenhagen (Denmark). The gunfire left one spectator dead and three police wounded. Later in the night, a Jewish community member was shot outside the Great Synagogue. In March, a neo-Nazi killed one person and wounded five others in a Mesa, Arizona (US) in a shooting rampage. However, outside the western world, thousands of people have died as a result of what are considered terrorist attacks in Mali, Nigeria, Cameroon, Iraq, Yemen, Philippines, Pakistan, Tunisia, Libya and Somalia only during the three first months of 2015, while big brother was watching us.

to Lyon, today's inclination to define and interpret the notion of *privacy* differently deceives the 'subjects of surveillance' (us), who, 'far from conforming to the abstract, disembodied image of both computing and legal practices, 'are engaged and embodied users-in-relation whose activities both fuel and exclude surveillance' (ibid., 1). These are the kinds of capabilities denounced by Snowden that are at the origin of 'reactive' data activism.

The reverse of the coin is the use of big data to form new digital public spheres, where everyone has the change to participate in, or at least influence, the decision making, empowered by new tools and data-based knowledge, as equals. Moreover, big data infrastructures allows for a global exercise of these rights.

3.7 A myriad networked individuals

So what is the place of individuals in a globalised world? Is it *retribalizing* us? According to Putnam and Bauman, some of the effects of globalisation have been quite the opposite. The associationism and lust for a good discussion that de Tocqueville identified at the heart of US's vigorous young democracy has languished, as Putnam describes in *Bowling alone*. And another phenomenon emerges: a new sort of individualism. Not the one discovered by the enlightened thinkers, who dared readers to think and speak critically and independently as individuals. Not the Habermasian recognition that everyone had dignity as a human being and the right to participate in a communicative action as an equal, but a new sort of individualism that is a symptom of a growing weakness of human bonds and languishing of solidarity. What is the connection of this with my object of study? Individualism, which is closely linked with consumerism, is a great source of big data as well. The *homo consumericus* described by Lipovetsky (2008) produces vast quantities of data and is the main target of marketing and big data analysis. Looking at trade, Cruickshank and Chis conclude that 'the Digital Age of big data

is one that re-inscribes neoliberal corporate capitalism' (2015, 53). Meanwhile, activism and social campaigning fight generally for the exact opposite values: solidarity, collectivism, co-creation and communitarianism. Consumerism goes hand-in-hand with this new type of individualism. Publicity feasts on it. Advertising identifies –replaces— participation with purchase, says Sampedro (2014, 84). The same media organisations that are vehicles for news and information place the financial interests of the conglomerates that support them before that of their publics' in the form of ads and publicity (ibid., 11). People are bombarded by ads that claim: 'You deserve it!' 'It's all about you!' Welcome to the 'me-culture.'[134] This is the other side of the coin of 'negative globalisation' (Bauman 2007, 24). People today recklessly seek individual freedom, but in consumer's terms and within what commercial brands have to offer. Today's basis for solidarity is even negotiated in the individual realm; we join in solidarity groups for selfish reasons, says Melucci (1996a, 8). Melucci states that the conflicts have moved from the industrial-economic system towards the cultural space: they focus on personal identity, motivation and the codes of quotidian behaviour. 'Increasing systemic differentiation simultaneously threatens social life with fragmentation, lack of communication, atomised individualism, and calls for deeper integration of individual and collective practices' (ibid.). We continue needing a sense of belonging, a community; however, the sum of many individual reasons does not make for a deontological code of a community. This new evolved individual, the *homo consumericus*, could be defined as 'a figure of hyper individualism, a consumer disconnected from class cultures, mostly seeking the pleasure, experience and emotion of social standing; but it is an anxious consumer' (2008).[135] The post-modern individualism is a weird one:

---

[134] I could not find an origin for this expression, but the media are full of it. A Common Dreams article, for example, explores the possibilities of going from a 'me culture' to a 'we culture' (Mattis 2011).

[135] 'Il n'y a pas de définition simple de l'homo consumericus, néoconsommateur ou hyperconsommateur. C'est un modèle théorique, modèle idéal au sens de Max Weber qui n'existe nulle part mais donne sens à

in their individual choices, hyper consumers look all like clones, because their choices are restricted to what big brands, using big data analysis, have on offer. Another paradox arises: the *homo consumericus* is a highly connected individual, it is part of the network, but feels alone and disengaged. In *Bowling alone*, Putnam examines the 'complex of factors that lies behind the erosion of (US) America's social connectedness and community involvement over the last several decades' (2001, 14). He blames the 'downturn in civic engagement' on the 'breakdown of the traditional family unit,' on the disconnection from friends and neighbours, and ultimately people's detachment from their democratic structures. In summary, Putnam warns that US's stock of 'social capital' (the makeup of strong networks) is degenerating. The *homo consumericus* signs fewer petitions, belongs to fewer organisations, knows fewer neighbours, meets with friends less frequently, and even socialises with families less often. They are bowling alone. However, the US has reinvented itself before, and it can reinvent itself again, says Putnam. The key is social capital –the space where creativity, productivity, networks, relationships, solidarity, favourable contexts are generated.[136] Every human being needs to feel that he or she belongs, but today's relationships do not imply a strong commitment. With the ICTs, in this society of consumers, relationships are articles to be consumed too. Bowling alone, one only thinks about new cars, clothes, accessories, music and smartphones that must be

---

ce qui change dans le monde contemporain. Ce néoconsommateur peut se définir comme une figure de l'hyperindividualisme, c'est un consommateur décroché des cultures de classe, plus en quête de plaisir, d'expérience et d'émotion que de standing social, mais c'est un consommateur anxieux' (ibid.).

[136] Called also 'relational capital,' the more extended and stronger the network of relationships, the bigger is the social capital. Social capital refers to the generation of processes of social interaction that activate responses in order to development a certain environment. It is 'capital' because it has value (i.e. knowledge, abilities and relationships). It is 'social' because it generates, and thrives on, social relationships (Baradiaran 2015). And it can be communitarian when ethical dimensions are shared and there is trust among its members that allows them to face new challenges in a collective and effective way. The more social capital that person has, the thicker the net (ibid.).

replaced every certain time (due also to 'planned obsolescence').[137] The level of social integration of an individual is appraised by his or her level of purchase power, as the level of development of a country is measured by the purchase parity power of its inhabitants (known as PPP). The accumulation of many purchasing actions generates economic growth. You become part of the pack if you wear the right training shoes and have the right king of bag on your shoulder, while someone is watching and taking note of your personal preferences within the allowed choices. Acquiring new things is not about usefulness or even possession anymore; it is about individual identity. Adverts make it clear: it is not about having, it is about a lifestyle, a state of mind, it is about simply being. We are focused on the self at the expense of the community. Individualism creates the structure and consumerism provides its dominant ideology and the navigation chart for individuals to find the way in a world where there is an appearance of choice.[138] Bauman also points out that, in the Information Age, everything is fluctuating and shifting constantly, like liquid: relationships, family, workplace, entertainment, fashion, spiritual and material life… Facebook *friendships* replace real human bonds. Aesthetic commitment communities substitute real commitment communities (Bauman 2001, 65-66). And 'the need for aesthetic community generated by identity concerns' –the 'favourite grazing ground of the entertainment industry'— is never to be fulfilled. The 'vastness of the need' explains this industry's continuing success: it is 'self-perpetuating as it is self-defeating,' never to be gratified, (ibid., 66). Mobile behaviour is one example. Observe those people chatting away on their

[137] Planned obsolescence in industrial design is designing a product with an artificially limited useful life, so it will become obsolete. Firms that pursue this strategy believe that the additional sales revenue it creates more than offsets the additional costs (The Economist 2009).

[138] Individualism and consumerism take us to abuse natural resources without taking into account future generations. For example, millions of tonnes of fish that is not commercially viable are 'discarded' every year. 'Discards' in industrial fishing is the portion of a catch of fish that is thrown, often dying, back to the sea (Food and Agriculture Organisation 2016). The practice is driven by economic factors; discarded fish are frequently unmarketable species, below minimum landing sizes, and catches of species are not allowed to land for instance due to quota restrictions designed to combat overfishing (ibid.).

smartphones while traveling on a train. Are they chatting to each other? As they navigate, buy, sell, search for company on their smartphones (while corporations and governments collect their data), they seem alone. In 2013 security footage on a San Francisco Muni train showed a number of passengers that were 'too distracted by phones' to notice a man playing with a gun until he shot someone (O'Connor 2013). The commuters in the San Francisco Muni train were mesmerised by the screens on their phones, while being recorded by a camera, as a man with a real gun was shooting someone. Lipovetsky and Serroy talk about an era of a multiplication of screens too (2009).[139]

The result is that the parameters of evaluation of the social action begin with the *I*, while all the elements that construct the *we* are in crisis. At the same time, there is an extraordinary structural diversification of individual identities. We live in plurality, which can be a fertile ground if it happens in a common civic situation. But this complexity takes us to an unstable equilibrium, to institutionalise a permanent crisis. There are social problems that do not have a unique or clear solution. What to do with our elders when they cease to contribute to society in monetary terms? How do we integrate migrants that come from different cultures? Reality shows that concrete cases are so complex that cannot be tackled from the normative system or traditional categorisations any more. Public structures are challenged by these new realities, and some of them loose credibility because they cannot cope with these changes. Non-governmental organisations and charities fill the gap left by the state, and the public administrations become irrelevant; they become 'communication agencies' that cannot tackle real social problems or conjure real participation, and are reduced to transmitting messages instead (Barandiaran 2015).

[139] However, they did not refer to omniscient surveillance or to smartphones.

But history has shown that the survival of institutions is linked to their producing solutions to social problems; when they cease to produce solutions, they decline and die, or are toppled.

In spite of all this, things are not as they seem. There is a reaction against consumerism and individualism. Melucci puts is like this: 'While we can agree with much in this description, in our everyday lives we nevertheless encounter others in important ways that are not immediately visible; consequently, the texture of our relationships is far richer than what the accounts of social atomisation would have us believe' (1996b, 99). Activism and journalism, together with other forms of public service and social change endeavours, are experiencing a revitalisation thanks to big data and communications infrastructures, as we will see.

3.8 Exclusion in a globalised planet

Contrary to what Rossetto and other like-minded writers say, technology does exclude people; post-modernity, globalisation, individualism and consumerism exclude people as well. However, not all the types of exclusion are the same. As I see it, technology-related exclusion today could be classified in three main types: purely technological, economic and structural, in descending order, from a subtler to stronger level of exclusion. This final notion is important as well, since exclusion –a negation of the Habermasian ideals of equal participation— is one of the wrongs that data activists are trying to fix. Additionally, exclusion in the big data society can be more than just a question of lack of access to technology (Deibert et al. 2010). Innerarity talks about a widespread 'technological ecstasy' that relegates the digital gap to merely 'inequality in access.' But there is inequality 'in the use that people make of the possibilities opened up by the digital technology' as well (2016a, 2). This ecstasy is linked with a 'deterministic and reductionist vision of technology, which is then not considered as a social and cultural phenomenon' (ibid.). Making a defence of culture, Innerarity argues, for example, that the hopes

placed in big data could be misdirected. 'I am starting to think that big data are all about the illusion that the analysis of data correlations will allow us to renounce to theories, in a way that could lead to assert: "big data, small theories"' (ibid.).[140] In the big data universe, 'what we could call *data rich* and *data poor* exist as well. This difference has its causes, on the one hand, in the inequality that is present in the production, interpretation and utilisation of data, and on the other hand, in their relation with issues such as reputation, valorisation and visibility that they provide' (Innerarity 2016b, 1).[141] Innerarity highlights the fact that, in spite of the growing access to both data and data infrastructures, only an elite achieves competency when it comes to controlling the whole data value chain: from the access to data to the capacity to use data, including the different abilities to extract, clean, store, manage, analyse and visualise them.

> The current big data ecosystem generates a great inequality, in spite of the fact that it is a different type of poverty or wealth that relates to the possession of material things. There are three types of people when it comes to databases: those who produce them, those who can store them and those who know how to extract value out of them. The latter is the smallest and most privileged (ibid.).[142]

---

[140] 'Y empiezo a pensar que eso de los big data se corresponde con la ilusión de que el examen de las correlaciones de datos nos permitiría renunciar a las teorías, de manera que cabría asegurar: grandes datos, teorías pequeñas… El éxtasis tecnológico suele ir unido a una visión determinista y reduccionista de la tecnología, a la que no considera un fenómeno social y cultural… Otro ejemplo que ilustra dicho reduccionismo es que hayamos concebido la brecha digital como desigualdad en el acceso y no por el uso que hacen las personas de estas posibilidades abiertas por la tecnología digital' (ibid.).

[141] 'En el universo de los big data hay también lo que podríamos llamar ricos y pobres de datos. Esta diferencia tiene sus causas, por un lado, en la desigualdad que se refiere a la producción de datos, a su utilización e interpretación y, por otro, en relación con la reputación, valorización y visibilidad que estos medios realizan' (ibid.).

[142] 'El actual ecosistema de los big data provoca una gran desigualdad... Hay tres clases de personas en relación con los bancos de datos: quienes los producen, quienes tienen capacidad de almacenarlos y quienes saben cómo valorarlos. Este último grupo es el más pequeño y el más privilegiado' (ibid.).

I agree. In big data infrastructures, the technological exclusion can be found not only in the lack of access to these technologies that most people experience. It could be understood as well as a) the inequality in access to the very data; b) the unbalances that exist within datasets, especially when it comes to data from developing countries and poor people; c) and the power disproportion that occurs between the surveilled, whose data are harvested, and the surveillants, who do the harvesting. These are taken into account in the following section and across this dissertation. One example is the asymmetry that Andrejevic calls the 'big data divide,' which differentiates between those 'who collect, store and mine large quantities of data' and those who become 'data collection targets' (2014, 1673). To this kind of asymmetry, the gaps that emerge from the fact that big data do not capture everyone can be added (i.e. a poor person living in a poor country, who does not own a smartphone, does not live in a city and is not an *homo consumericus*). But the results of big data analysis depend entirely on what data you are looking at (Leetaru 2015a). If datasets are biased, incomplete or wrong, the analysis will be faulty. For this reason, relaying on big data and social network analysis[143] to observe social movements is problematic as well. Talking about why online big data are 'bad data' for researching social movements, Schradie lists several big data flaws, for example that 'hashtag data are often cherry-picked' (2015, 1). A problem in some social networks studies has been choosing case studies based on high levels of internet and social networks use, for example. Another problem is that big data generated by social networks are 'too small,' because 'tweets and Facebook posts exclude those who are on the other side of the digital divide' (ibid., 2). Schradie says that 'a gaping hole in the digital cloud is that online big data can leave out the poor and working class'

[143] A social network will be understood hereafter, not simply as a network of social interactions or relationships, but more specifically as a dedicated digital application, which enables users to communicate with each other by posting text, pictures, videos and audios (i.e. Twitter).

(ibid.). Taking into account two hiatuses in large sets of data –the non-existent data on the very rich and on the very poor—, it can be concluded that big data tend to represent better the middle class hordes of *homines consumerici.* Feenberg points to another form of technological exclusion, which is generated when the majority of people are excluded from the decision-making processes of producing such technologies (except maybe as consumers and users who influence decisions in the manner proposed by cultural materialists). Since today's 'public life' is mediated by technical decisions, and what human beings are decided 'in the shape of our tools,' the design of technology is 'an ontological decision fraught with political consequences,' and 'the exclusion of the vast majority from participation in this decision is profoundly undemocratic' (2002, 3).

Nonetheless there are other sorts of technological exclusion even among consumers and users. Indeed, the extreme velocity of the development of ICTs –together with the issue of access— excludes people who cannot keep up with them, and generates the 'digital divide.'[144] People who have the mechanisms to participate in this society are called the 'digital citizens;' whereas people born in this age are called 'digital natives.' Meanwhile people who are born before the Information Age are called 'digital immigrants' (Prensky 2001, 1-2). At an individual scale, 'digital immigrants' who cannot keep abreast of new technologies are left behind, also in the labour market and in other spheres of society, because of their lack of skills. Fraser notes that some marginalised groups are excluded from a universal public sphere. However, she claims that marginalised groups can form their own public spheres, and terms this concept a *subaltern counterpublic* or *counterpublics* (1990, 61). Fraser uses Habermas's notion, but observes that

[144] For example, more than 80% of the content on the internet is in one of just ten languages out of more than 6,000 spoken today (Leetaru 2015a). In a paper about translation in news, I argued that the last frontier in communications in the network society is a 'language divide' (Gutiérrez 2006).

‘Habermas stops short of developing a new, post-bourgeois model of the public sphere’ (ibid., 58). In her evaluation of the bourgeois public sphere, Fraser argues that rather than opening up the political realm to everyone, the bourgeois public sphere shifted political power from ‘a repressive mode of domination to a hegemonic one’ (ibid., 62). To deal with this hegemonic domination, Fraser argues that repressed groups form ‘subaltern counterpublics’ that are ‘parallel discursive arenas where members of subordinated social groups invent and circulate counter-discourses to formulate oppositional interpretations of their identities, interests and needs’ (ibid., 67). In today’s network society, precisely the easier access to technology facilitates the generation of other digital public spheres or *counterpublics*.

There are other marginalised groups, such many elderly people, who will never be ‘have-laters’ (Rossetto 1998).[145] But even so, an old person in a rich country receiving a decent pension regularly, with property to call his or her own, discounts to grant him or her access to trips, treatments and cultural treats, and admittance to the best public health systems that post-modernity can offer maybe well be a digital migrant, but can hardly be called an economic migrant. Economic exclusion can be observed at different levels. The economic context in which this technological development occurs dictates that, in any given developed economy, as industrial and manufacturing activity declines and industrial plants are delocalised, the tertiary sector (also known as the service sector) –and above all the quaternary sector (the knowledge-based part of the economy)— flourish. The quaternary sector –which did not exist before the post-industrial society (Touraine 1971, 1) — typically includes innovation-based services like information technology, information generation and sharing, the media industry, blogging, designing, research and other knowledge-based services like consultation, education or financial

[145] My father writes code and is an enthusiast of technological innovations, although he is an exception.

planning. Technology, and more concretely ICTs, also separates countries that have initiated the process of developing a strong quaternary sector from those which have not.

This world is extremely unequal. An Oxfam report shows that, by 2016, the share of the global wealth of the top 1% population, the richest of the rich, will surpass that of the rest –the bottom 99% — (Oxfam 2015). This is also a key issue for data activism. Occupy Wall Street, with its motto 'We are the 99 percent,'[146] is possibly the first movement to have data as its political slogan. The Oxfam report confirms this economic divide: The wealth of 80 rich individuals is now the same as that owned by the bottom 50% of the global population, such that 3.5 billion people share between them the same amount of wealth as that of these extremely wealthy 80 people (ibid., 3). Stiglitz (2011) and Krugman (2011) maintain that similar gaps exist between the top richest –whose incomes have been growing at exponential rates in the past few years—, and the rest. The inequality between poor people in developing countries, and middle and high class citizens in rich countries, what Rosling would call playfully the people living above the 'air-line' (2011), is stark. But the divides within cities are arguably starker. Bauman points out that cities used to be raised to lock strangers out, but today they have turn out to be dangerous places, where some of the deeper inequalities on earth exist. As a result, the urban poor become ghettoised in the most unattractive zones and are permanent exiles from much of their own city (2007, 46). The economic divide can propagate indignation. Stiglitz blames global inequity for the outrage that ignited the global protests at the World Trade Organisation meeting in Seattle in 1999, and every subsequent summits ever since (2002, 245). And when people lack the key information that bears on the decisions they have to make, or when institutions are absent

[146] 'We are the 99%' became the mantra of the 'Occupy movement' in August 2011 after a Tumblr blog 'wearethe99percent.tumblr.com' was launched in late August 2011 by a New York activist going by the name of 'Chris' together with Priscilla Grim. Tumblr is a microblogging platform and social networking website (WeArethe99% 2013).

or flawed, inequality thrives. Another subtler form of economic exclusion occurs when salary and personal resources simply do not allow you to be a *homo consumericus*, and therefore part of the big data society. The crisis that started in 2007 created another new phenomenon: 'the working poor,' or people who work and earn a salary, but live below the relative poverty line.[147] Spain is a good example. Based on data from the 2009 compendium of EU indicators and Spain's own living conditions surveys for 2004, 2006 and 2008, Cabrero Rodríguez concludes that the Spanish working poor is a serious problem (2010, 19). The most vulnerable people in rich countries of Southern Europe, including the working poor, have suffered terribly due to the financial crisis that started in 2007, threatening the collapse of large financial institutions. It was only prevented by the massive bailout of banks by national governments. The blasting of the housing market bubble resulted in mass evictions, foreclosures and prolonged unemployment.[148] The crisis played a significant role in the failure of key businesses, the decline in consumer wealth, and the downturn in economic activity leading to the global recession and contributing to the European sovereign-debt crisis. While the US recovered quickly with expansionary economic policies, Europe, especially Spain, Italy and Greece, still lags behind, trapped in its austerity policies imposed by Germany and the so called *troika*: IMF, the European Union and the

[147] Relative poverty refers to a standard which is defined in terms of the society in which an individual lives and which therefore differs between countries and over time' (The Poverty Site 2012). In this sense, poverty can be absolute too. As of October 2015, the global poverty threshold was updated to US$1.90 by the World Bank; that is, a person, in any country, that lives with less than US$1.90 a day is considered poor (World Bank 2015).

[148] In Spain, for example, as a consequence of the crisis, 'the number of court-ordered home evictions for non-payment of mortgages, rent or other legal reasons reached 67,189 in 2013, according to judicial statistics. The number of open cases –pending eviction requests– stood at 82,860, which is 9.8% fewer than the previous year, said the General Council of Judiciary legal watchdog in its annual report' (Hernández 2014). In Spain, the repossession of homes by banks does not necessarily mean the cancelation of the owner's mortgage debt; mortgage lenders have a claim on all the borrower's assets, not just their homes, while homeowners have to take on all the risk of their house losing value. Meanwhile, in 2012, Spain received around 40 billion euros (US$52 billion) in European aid for its nationalised banks – some of them holding large house portfolios— as agreed in a bank rescue package requested by the government.

European Central Bank. In these circumstances, the Habermasian interests of human beings in labour, communication and emancipation seem to dissipate. However, instead of admitting defeat and surrendering to extreme fatalism (or doing just the opposite, falling into an unrealistic optimism and techno-faith), there is the middle way of alliance-building, collaboration, research-based campaigning and mobilisation, and co-creation. New unchartered paths become visible, where big data analysis can help illuminate the way. An example is DataKind,[149] an initiative that tries to harness data for development. What they do is bring together data scientists with social change organisations to collaborate on 'analytics and advanced algorithms to maximise social impact' (DataKind 2015). The trick is to put them together to work for a common goal.

> The same algorithms and techniques that companies use to boost profits can be leveraged by mission-driven organisations to improve the world, from battling hunger to advocating for child well-being and more. However, most social change organisations don't have the budget or staff to take full advantage of this data revolution and most data scientists don't realize just how valuable their skills can be (ibid.).

Finally, there is the structural exclusion that runs deeper and goes beyond the lack of access to technology or economic means. It is a total exclusion of the network. The inequalities of the network society precede the 2007 financial crisis and are possibly inherent to capitalist societies. The network society 'diffuses selectively throughout the planet… Societies are deeply fragmented by the double logic of inclusion and exclusion in the global networks that structure production, consumption, communication and power' (Castells 2009, 23). And this fragmentation is more than the expression of the time-lag required by the gradual incorporation

[149] As said, DataKind accepts individual donations, as well as corporate and private sponsors, and in-kind donations. Its supporters include Microsoft, AT&T, MasterCard and IBM. DataKind also actively recruits volunteers.

into the new dominant logic, says Castells. ‘It is, in fact, a structural feature of the global network society.’ From a societal point of view, people are influenced by networks and ‘the network society is a global society’ (ibid., 25), but it does not mean that all people are included in them; in fact, many are not. That is, people are indeed influenced by these global trends and processes, even if they are not part of the network. The network excludes people who are neither connected nor generators of data. This issue runs parallel to other trends, such as the level of data openness of a country. Twenty-six of the top thirty countries in the Organisation for Economic Co-operation and Development’s Open Data Barometer are high-income countries. ‘Half of open datasets in our study are found in just the top ten OECD countries, while almost none are in African countries. As the UN pointed out last year, such gaps could create “a whole new inequality frontier” if allowed to persist’ (Organisation for Economic Co-operation and Development 2016). It is no coincidence that many of the people who are excluded structurally by the network are probably located in Sub-Saharan Africa. What nobody can avoid is to be affected ‘by processes that take place in the global networks that constitute the social structure,’ since the ‘core activities’ that control human life are organised in global networks, including financial markets; transnational production, management, and the distribution of goods and services; science and technology, including higher education; the mass media; the internet networks; culture; art; entertainment; sports; international institutions managing the global economy and intergovernmental relations; religion; the criminal economy; and the transnational NGOs and social movements that assert the rights and values of a new global civil society (ibid.). Globalisation ‘is better understood as the networking of these socially decisive global networks’ (Castells 2009, 25). Therefore, exclusion from these networks, often is a cumulative process of exclusion, is tantamount to structural marginalisation in the global network society.

So far, I have described the big data society and its social, economic and cultural context, however imperfectly and selectively. I have described new phenomena such as the rise of the smartphone, the 'computational politics' and the emergence of a new post-modern *panopticon*, the birth of a digital public spheres, the space of flows, a new type of individualism never seen before and the kind of exclusions that occurs in a complex, highly technological, networked world. This paints an extremely multifaceted landscape where these phenomena are intertwined, sometimes working in unison fostering each other, sometimes working as opposing forces. The more powerful and encompassing the *panopticon*, the fiercer the reaction from reactive data activists; the more unequal the economy, the louder the Indignad@s. As I have been repeating: for every negative, there will be a positive (Schmidt and Cohen 2013, 51).This notion brings me to the end of this theoretical section, and takes me to the next section, which gathers the data mobilised for this dissertation. I observe empirical cases of uses of big (and small) data by different sectors, including governments and intergovernmental agencies, private sector and journalism, with special emphasis on the uses by the third sector, where data activism belongs. I also draw on the disciplines of critical thought, and journalism, alternative media, international relations and social movement studies to generate data that allows me to group and classify the different types of proactive data activism. The classifications also frame the case study. Finally, the data is analysed at the end of the next section.

4. The subjects of big data generation

> "There is a movement of quantification rumbling across fields in academia and science, industry and government and non-profits," says King, who directs Harvard's Institute for Quantitative Social Science, a hub of expertise for interdisciplinary projects aimed at solving problems in human society. Among faculty colleagues, he reports, "half the members of the government department are doing some type of data analysis, along with much of the sociology department and a good fraction of economics, more than half of the School of Public Health, and a lot in the Medical School." Even law has been seized by the movement to empirical research —"which is social science" he says. "It is hard to find an area that hasn't been affected" (Shaw 2014).
>
> (Big data have) the potential to transform development and accelerate social progress around the world, but there are issues surrounding understanding, ownership, privacy, capacity, measurement and more that need further dialogue and discussion (Skoll World Forum 2015).
>
> Data can be generated, collected and analysed by alternative actors (to governments and corporations) to enhance rather than undermine the agency of the public. Indeed, it is precisely because the massive flows of data circulating between devices, institutions, industries and users usher in new and troubling practices of dataveillance that it becomes vital to reflect on whether there are alternative forms of big data, forms which enable the less powerful to act with agency in the face of the rise of data power (Poell, Kennedy and van Dijck 2015).

As we have seen, big data are not perfect or complete: inequalities and omissions in society translate into inequalities and omissions in big data as well. Datasets are full of imperfections and gulfs, they embody 'a rather crude definition of society,' they are 'marked by strong asymmetries of power,' and they represent 'only a passing moment in the traceability of the social connections' (Latour et al. 2012, 612). Both their appropriation and impact cannot be completely understood by newly gathered 'game-changing' discourses and analyses, or by a unidimensional approach (ibid.). Big data may be stored in databases in remote locations or in the *cloud*, but they do not occur in a social void. They do not arise out of nowhere, free from the

‘the regulating force of philosophy’ (Berry 2011, 8): they are framed by previous existing policy structures, value systems and science. As a result, they have a ‘limited generalisability: internet users, or users of specific platforms, are not a random sample of the population’ (Rojas 2015). So, contrary to what statisticians do, one cannot infer an encompassing and real public opinion from a sample of online users. At best, one can ‘show how one dataset might correlate with “real world data” through extra analysis’ (ibid.).

So far, I have explored big data with some degree of detail, particularly what they are, how they are generated –namely, how they emerge from society. In this section, I examine big data’s social uses – specifically, what can be done politically with big data to understand and transform our society. Big data are been praised for its power to enhance our knowledge in new ways and to enlarge our understanding of society (Lazer et al. 2009, 721; Lohr 2012). Data are like water, yeast, bacteria and starch molecules in the Zeleny’s model (which uses bread making as a metaphor) in the stepladder towards wisdom (1987, 72). Big data are not valuable *per se*. It is what we do with them that *is* valuable; and their worth is not only in the economic gains that can be extracted of their *value chain*. Henceforward I look in-depth at how big data can be meaningful in use (Trottier 2014, 52), and I review the social functions of big data and data analysis from the point of view of different theories and sectors, including government and intergovernmental agencies, the private sector, journalism and civil society, also known as the third sector. Although brief, this review of the uses of data in different sectors is relevant to this dissertation for two basic reasons. First, there is an incessant and unapologetic transference of skills, contents and perspectives when it comes to employing data (i.e. ONGs producing data journalism). The exploration of data uses exclusively in the third sector would be incomplete without a look at how technologies and approaches have evolved in journalism or in the private

sector, so example. Second, the birth of data activism is inherently linked to the birth of governmental and corporate data-based surveillance technologies, as seen in the previous section. All sectors are trying to find uses for big data, and although in some cases the nature of the endeavour is not clear, I have tried to distinguish between them according to what their main objective is, distinguishing between liberal-capitalist or for-profit, on the one hand, and communitarian and non-for-profit, on the other. Some of the examples quoted in this section do not belong neatly to any of the four identified sectors (public, private, journalism and third sector), but sit on the intersection of two of them. For example, a news organisation can have an NGO's 'business model' and legal status; that is, it can have the income and fundraising structure and legal status of a non-profit, while producing purely journalistic content, and adhering to journalistic rules and techniques. Journalistic organisations can be, and often are, private companies nevertheless. I have classified and group all cases in four groups just on behalf of simplicity. I start with examining the uses of data in the governmental and intergovernmental sector, with special emphasis on development issues, followed by a glimpse of how some private companies are using data in the private sector. Next I dwell more exhaustively on the uses of data in journalism, because of the close association that exists between data journalism and data activism, to end with the main part of this section, which is dedicated to the uses of big data infrastructures in advocacy, campaigning, assistance and development within the third sector.

4.1 Uses by governments and intergovernmental agencies

Rulers have always been interested in demographic data about the ruled. Today in countries where the coat of connections is thick, 'governments, researchers and development agencies' are reluctant to implement costly surveys in order to classify, observe, watch over, tax and police their populations; instead, increasingly they resort to big data infrastructures to

produce ‘new socio-economic value’ and demographic insights (International Telecommunication Union 2014, 177). Big data are being employed also in providing better services to citizens and in guiding better public interventions (Letouzé and Sangokoya 2015, 22). ‘Systems that rely solely on human intuition or predispositions are far from perfect and human interventions often mean that prejudice rather that the law rules. A famous study showed that sentences handed out by judges differed widely for similar offenses depending on whether they were pronounced before or after lunch. Algorithms don’t get hungry, nor angry’ (ibid.). The possibilities offered by big data analysis in all fields of the public sphere seem endless: they can be used to increase tax compliance and revenues (Campbell 2014, 2); detect and predict criminal behaviour (Gibbs 2014; Bogomolov et al. 2014, 427); improve public service delivery (Spotfire 2013; Rutter 2014); in an area where the public and the private intersect, help airlines avert bad weather (Andronie 2005, 153); and generate extremely accurate rainfall information (Andrade et al. 2014, 81). Due to their capacity to predict human behaviour, further areas of public service that stand to benefit from big data infrastructures include health, education, transport, personalised public services, surveillance, policing, and early warning systems. In fact, public services were amongst the first to realise the potential of data analysis. An early example of data visualisation used for public service is John Snow’s map showing the cases of cholera in London in 1854. This is a paradigmatic case of a Baconian ‘instance of the fingerpost’[150] facilitated by a data visualisation. When positioning the cases of cholera on the London map, this precursor of epidemiology realised that they did not spread uniformly, but clustered around fountains, which were contaminated by faeces, demolishing other explanations such as the miasma theory (which

[150] ‘Sometimes these instances of the fingerpost meet us accidentally among those already noticed, but for the most part they are new, and are expressly and designedly sought for and applied, and discovered only by earnest and active diligence’ (Robertson 2013).

put the blame on 'bad air').[151] Administrations started collecting and analysing demographic data systematically in the Enlightenment,[152] and with time moved into the more controversial pursuit of digital data about their citizens' activities and movements. Data produced by mobile technology is the source vast amounts of data on people's behaviour, and they can be real 'digital census-takers' (The Economist 2015a), therefore especially apt for demographic studies. That is because smartphones are creating 'a more detailed view of society than has ever existed before' (ibid.), and in real time. Reigned 'by suitable regulations, anonymised personal data can be used, among many other things, to optimise traffic flows, prevent crime and fight, epidemics' (ibid., 3). Governments are applying big data analysis to areas such as tax compliance, crime detection and public services.[153] The OECD and the G20 have long acknowledged the need to upgrade the use of data to increase indirect tax revenues.[154] In response, many national tax authorities have begun to explore how they can leverage their data to advance their ability to catch irregularities and loopholes. One of the most progressive, the UK tax authority, has started to use big data approaches to lessen potential fraud, investing in their capabilities to spot 'missing trader fraud,'[155] a source of tax loss. This approach allows authorities to profile 'high risk' individuals

---

[151] The Snow example can be interpreted as an early case of advocacy as well. For Tactical Tech, this map is an example of 'information design applied to analysis and advocacy… used to tell a powerful, persuasive story on behalf of a cause' (Emerson 2008, 9). 'Snow gathered data by talking to local residents... Despite scepticism, he collected enough evidence to prompt officials to shut down the pump, after which the epidemic quickly ended.' Snow's work promoting the idea that the disease was spread through contaminated water' became a 'turning point' not only in the history of public health (ibid.), but also in campaigning, by establishing a connection between a data-based research and policy-makers influencing for a social cause.

[152] Romans, as well other civilisations, had registered censuses. Whether Roman censuses included only all adult free males or all citizens *sui iuris* regardless of age and sex is still debated (Hin 2007, 1). However, I am referring to the Enlightenment as the main reference for this dissertation.

[153] These areas have been selected here because, from an empirical observation, they are three of the most common at the time of writing. I am excluding digital surveillance because it has been examined earlier.

[154] Called also goods and services tax (GST), an indirect tax (such as the value added tax or VAT) is a tax collected by an intermediary from the customer who acquires the goods or service.

[155] Missing trader fraud is the misappropriation of Value Added Tax (VAT) from a government using the loopholes generated by the way VAT is treated within multi-jurisdictional trading. Normally it consists of

and companies for deeper investigation (Campbell 2014, 2). In a not so distant future, criminals could be stopped even before they attempt to commit any crime by profiling them and anticipating their criminal plans. 'While individual behaviour can be hard to predict, determining the average behaviour of a population and then matching individuals to that template to determine "fit" can be surprisingly accurate' (Gibbs 2014). Big data can help forecast 'crime hotspots' based on data originated from mobile network activity in combination with demographic information (Bogomolov et al. 2014, 433). This is done by using 'predictive policing' or the employment of 'mathematical, process-oriented data in order to anticipate a crime, with the goal of preventing it from happening or mitigating its effects' (Letouzé and Sangokoya 2015, 8). Plagued with budget cuts, overwhelmed governmental agencies have increasing difficulties to meet the needs of citizens and businesses. However, big data present 'a view into new ways to improve services and processes,' including opportunities 'to make better decisions faster, forecast outcomes by evaluating different scenarios under controlled situations, and improve productivity by finding information and/or deliver services more efficiently' (Spotfire 2013). Examples of how governments are using big data infrastructures to respond to citizens' needs include the analysis of the adoption of new guidelines for primary care doctors working within the UK health system (Stokes, Baker and Pigott 2014, 4). City governments are also stimulating data-driven projects to solve some of their challenges. Funded by Microsoft, the city of New York organises a competition, BigApps[156] to promote initiatives that, using technology and data, address the metropolis's challenges. Seven winners got US$100,000 each

---

a manoeuvre to charge VAT on the sale of a given party of goods, and then instead of paying this over to the government's collection authority, disappear with it as a trader. The term 'missing trader' refers to the fact that the trader goes missing with the VAT (Out-Law.com 2016).

[156] See nycbigapps.com/ [accessed on February 23, 2016].

in cash prizes in 2014, including data-based applications that offered employment and housing services, energy efficiency tools, maps showing crime hotspots, and the like.

Along with governments, intergovernmental agencies are exploring big data too in relation with some development areas and human rights. In fact, there is also a new demand for data analysis in development, and most UN agencies have changed their strategies accordingly. Discussing its strategic framework, Food and Agriculture Organisation says 'the emphasis on evidence-based decision-making in governments and organisations at all levels puts a greater focus on statistics, and their role in measuring and monitoring progress towards national and international development goals and targets' (2016). The UN Statistical Commission (UNSC) and national statistical organisations (NSOs) 'are looking into ways of using big data sources to complement official statistics and better meet their objectives for providing timely and accurate evidence for policy-making' (International Telecommunication Union 2014, 173). In spite of their biases and gaps, big data are perceived as presenting great opportunities for development in the international arena. In 2013, the UN' Department of Economics and Social Affairs Under-Secretary-General Wu Hongbo called for more data in development. 'Statistics is shaping our understanding of the world,' he said addressing the UNSC (UNDESA 2014). Although it is difficult to say who is influencing who, this interest may have a trickled down effect within the UN spheres of work, since it is likely to alter the advocacy strategies of a myriad of civil society organisations working in the same fields, which will move away from moral arguments into statistical evidence in advocating for their causes (Milan and Gutiérrez 2015, 5). ITU sees great potential to produce development-related evidence of relevance to public policy in big data from the ICT services industry as well. A much quoted example, in an area of confluence between the

public and the private, is Google Flu Trends (GFT).[157] In 2008 Google[158] discovered that there was a connection between how many people search for flu-related matters and how many people have flu symptoms. Obviously, not every individual who searches for the words 'flu' or 'cold' is really ill, but a pattern would appear when massive flu-related search queries are combined together. GFT was supposed to track the spread of influenza in near real-time 'more rapidly than the Centres for Disease Control and Prevention' (International Telecommunication Union 2014, 178). Although it was deemed a failure by some (Lazer and Kennedy 2015),[159] this exercise set off similar efforts focusing on the use of search-engine data to, for example, predict unemployment claims in the US (Choi and Varian 2009, 1).

On the other hand, the deficiency and scarcity of data on development issues (i.e. millions of poor people in developing countries are off the radar) is currently an obstacle for the application of data analysis to development issues. 'We know the least about those who are most lacking, and people often lack the information and capacities they need to bring about change'(Overseas Development Institute 2015). There is an acute need for better and more data to come up with diagnosis and solutions that tackle developing countries' challenges. This is a pending issue that has to do with the sort of technological, economic and structural exclusions that have been examined earlier. However, some of the most innovative approaches to social problems come precisely from developing countries. An example is Ushahidi, originated in

[157] See google.org/flutrends/about/how.html [accessed on February 23, 2016].

[158] Google Inc. is a US technology company specialised in internet-related services and products. These include a popular internet search engine, online advertising technologies, cloud computing and software (Google 2016).

[159] 'And then, GFT failed—and failed spectacularly—missing at the peak of the 2013 flu season by 140%. When Google quietly euthanized the program, called Google Flu Trends (GFT), it turned the poster child of big data into the poster child of the foibles of big data. But GFT's failure doesn't erase the value of big data. What it does do is highlight a number of problematic practices in its use—what we like to call "big data hubris." The value of the data held by entities like Google is almost limitless, if used correctly. That means the corporate giants holding these data have a responsibility to use it in the public's best interest' (Lazer and Kennedy 2015, 721).

Kenya to address disasters and crises using citizen-generated data, as it will be further examined in the case study.

So far, I have examined how governments use data, whether it is to improve people's lives or to spy on them. However, they also are opening their vaults so other actors –individual citizens, companies and NGOs— can use their data. A variety of online platforms ease the task: for instance, the Open Knowledge Foundation has developed the open-source data portal Comprehensive Knowledge Archive Network (CKAN),[160] which is used by numerous governments, organisations and communities to release data catalogues; and Open Spending, which allows non-experts to delve into over 13 million government financial transactions from 66 countries. Forced by the open data movement, governments of all stripes are setting up open data platforms. However, this does not mean that all relevant data are always made available in a usable manner, as the open data movement principles command. An extreme example is the *open* data portal of Saudi Arabia.[161] According to the Global Open Data Index Survey, Saudi Arabia's data platform does not offer information on the Saudi's government budget or spending, and the information on other items is 'unsure.'[162]

There is a grey area where governments and businesses cooperate: the complex machinery devoted to spying on citizens and creating the informational state. An informational state requires 'corporate collaboration' (MacKinnon 2012, 8) to implement its laws and regulations. MacKinnon explores Google's and Yahoo's[163] history of censoring citizens at the request of China's government, and Facebook's submission of the identities of protestors to

[160] See ckan.org/ [accessed on February 23, 2016].
[161] See data.gov.sa/about [accessed on February 23, 2016].
[162] See global.census.okfn.org/place/sa [accessed on February 23, 2016].
[163] Yahoo is another US technology company known for its web portal, search engine and related services, including Yahoo! Mail, Yahoo! News, advertising and its social media website (Yahoo 2016).

several governments.[164] For all the dreams of liberty that technology has granted, it is also clear that the corporations that rule the cyberspace[165] are making decisions that have affected civil freedoms. In *Consent of the Networked*, MacKinnon contends that it is time for us to demand that our rights are respected before they are sold, legislated and engineered away (ibid.). In fact, governments gradually rely more and more on 'private sector entities as regulatory agents, turning private centres of power to state purposes' (Braman 2006, 34). Not only do tech companies participate in the monitoring of citizens without their knowledge, they also contribute towards an *appliancised* internet, whereby 'tethered' devices (which can be modified only by the manufacturer) endanger the generative qualities of the network, namely its capacity of finding ways out of government control at the end-user level (Zittrain 2008, 8). This state of affairs becomes more worrying when dealing with authoritarian governments. There is an ongoing debate about the degree to which firms chartered in freer societies should assist in censorship or surveillance taking place in less free societies. The argument considered here is one layer deeper than that debate: authoritarian governments are inheriting abilities to enforce their will more easily (ibid., 113). That is, as with propaganda, in a democracy that supposedly guarantees individual and collective liberties, governments have to disguise interference and manipulation

[164] Another example is how Google became a censorship tool at the request of the US government when the cheaply made anti-Islam movie *Innocence of Muslims* sparkled violence after being showed on YouTube in September 2012. Though it backs free speech, the White House advised Google to examine whether the video broke its own terms of service. In the end, Google blocked the video voluntarily in Libya and Egypt. When editing a story on this case for *Index on Censorship*, I had the chance to talk to a Google Europe's representative, who resented the fact of being 'used' by the US government to censor content. But he also added: 'people were being killed, put you in my place,' referring to the violence triggered off by the film (Gutiérrez 2012, 2).

[165] 'As an arena for civic engagement, cyberspace is two things: first, it is a "gym" in which to practise political participation and digital citizenry, where alternative and often contradictory views about society are articulated and shared. Second, it is a platform for collective action, like a public square would be: a site to articulate, organise and bring forward social struggles, a site where cyber-specific forms of collective action can take place. But far from being only a set of tools, cyberspace has become a site of struggle in its own right, because it has partially lost its original character as an e-commons and is threatened by increasing commercialisation, tightening state control, and restrictive legislation' (Milan 2012, 1).

better, whereas authoritarian regimes can be more offhand (Gutiérrez 2004). The governments of China, Cuba and Saudi Arabia have assembled intricate machineries of control that can branch out economic from political activity on the internet, for example (Kalathil and Boas 2003, 10). Because internet access in authoritarian regimes is frequently channelled via state-owned or approved service providers,[166] these authors contend that the internet does not 'inexorably' undermine authoritarian regimes (ibid., 136). Technology's freeing impact depends on the particular institutions and regulatory frameworks set up in specific countries. Extrapolating from Kalathil's and Boas's study of the impact of the internet on authoritarian regimes, some of the same things could be said about big data infrastructures in non-democratic countries: they pose an opportunity for governments to tighten up the control over citizens, making the thick coat of surveillance even thicker; at the same time, they are also a challenge for censors and propagandists, as they can empower people by granting them access to information and action they did not previously had.

This matter is extremely relevant for this study. As I have been saying, data activism takes two main forms: reactive and proactive. As the terms suggests, reactive data activism is precisely a reaction to governments' and corporations' intrusion, and mainly engages in protection from surveillance, censorship and manipulation (i.e. by developing encryption devices or sheltered, by-invitation-only networks and platforms). In summary, governments and intergovernmental agencies are progressively looking at how big data infrastructures can help them support, watch over and service citizens better on leaner budgets. Big data are also seen as an area to develop when it comes to understanding development challenges and assisting poor countries. But they are key tools when it comes to surveying and policing people as well. They

[166] In fact, this grey area becomes greyer in the case of authoritarian regimes, since the private and public spheres are in many cases the same thing.

have been so extensively used that an informational state has been created. And as MacKinnon suggests, governments are doing it with corporate collaboration. Corporates not only provide wider and better public services (i.e. Google Flu Trends and the subsequent exercises), but also create a net of surveillance and control, and facilitate censorship and manipulation.

4.2 Uses by the private sector

Together with public administrations, companies are making extensive use of big data infrastructures. In this section, I examine how the private sector employs these technologies in ways that are relevant to this dissertation. This is a pertinent issue because the application of big data infrastructures to generate profits has driven their development and accessibility, and some of the data solutions found by businesses have been applied or triggered applications in other sectors or have the potential to do so (i.e. including the cases of Walmart's tarts, as well as Bunt Planet's and Vestas's smart grid management examined later). Although deemed 'epic failure' (Lazer and Kennedy 2015),[167] Google Flu Trends is an illustration. Since 2008, Google's example elicited similar ventures, for example the flu-prediction project by the institute of Cognitive Science Osnabrück,[168] which improves the original model by combining social media (i.e. Twitter) with Centres for Disease Control and Prevention (CDC) data, and models that infer the spatial and temporal dissemination of the disease.

Let us start by pointing out that algorithms and big data infrastructures have been used by companies to manipulate people as well. In 2014, Facebook conducted a massive psychological experiment on 689,003 users, tweaking what appeared in their news feed to measure the impacts on their emotions (Kramer, Guillory and Hancockb 2014). Germany's largest credit agency, called SCHUFA, planned in 2012 to analyse what circles of friends say about solvency of

[167] GFT shut down its web site in 2015 (O'Connor 2015).
[168] See flu-prediction.com/ [accessed on June 19, 2016].

individuals to profile them and assess their creditworthiness, mining data from Facebook, Twitter, LinkedIn and other social-networking sites (Hawley 2012). Nonetheless big data are being put to work in the private sector for less sinister targets. Marketing professionals, whose main challenge is to fathom their customers' interests and market trends, are now progressively getting away from traditional surveying and favouring big data approaches to scrutinise customer preferences from the analysis of datasets. In doing so, corporations are finding new, useful information about their customers' habits. In 2004, Walmart – the world's biggest retailer— discovered, using correlation analysis, that Pop Tarts (a snack) were heavily purchased by US citizens preparing for severe weather events, such as hurricanes (Katrandjian 2011), for example. Bunt Planet[169] –a small company based in the Basque Country— uses data gathered by hundreds of sensors in the water service grids that they manage in order to make them smarter and more efficient. At a bigger scale, Vestas[170] –a global energy company dedicated to wind energy— has used big data platforms to enhance the modelling of wind energy production and identify the optimal placement for turbines, which involves a large number of location-dependent factors, including temperature, precipitation, wind velocity, humidity and atmospheric pressure. Whole industries, from energy and utilities to transport, are using big data technology to extract new information on production and delivery systems and customers' behaviour to tailor their services and products at lower costs, increasing their productivity. Telecom network operators make sweeping use of these techniques as well to come up with new customised services and to understand and 'optimise their management of customer relations, and manage their network quality and performance' (International Telecommunication Union 2014, 177).

[169] See buntplanet.com [accessed on February 23, 2016].
[170] See vestas.com [accessed on February 23, 2016].

Big data analysis is impacting the way companies are managed too. According to McAfee and Brynjolfsson, companies that use data-driven decision making intensely are 'on average, 5% more productive and 6% more profitable than their competitors,' a difference that is 'statistically significant' and 'economically important' (2012). 'Data-driven decisions are better decisions — it's as simple as that. Using big data enables managers to decide on the basis of evidence rather than intuition' (ibid.). For the management consulting firm McKinsey, big data 'will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus… Leaders in every sector will have to grapple with the implications of big data, not just a few data-oriented managers' (Manyika et al. 2011, 2). Other companies, for example, are utilizing the 500 million messages that are sent out every day on Twitter (Internet Live Stats 2015).[171] Several research groups have developed software to analyse the opinions and attitudes expressed in tweets to forecast social outcomes, such as election results, box-office revenues or stock market behaviour. MashApe, for instance, has published the List of 20+ Sentiment Analysis APIs,[172] some of which can analyse Twitter messages (Ismael 2013). Although the exactness of such predictions was a subject of dispute at the time, in 2012 Twitter started to publish a daily political index for the US presidential election based on these methods (Sharp 2012). Bollen, Mao and Zeng –a group of scientists investigating *meme*[173] diffusion—

[171] The first tweet was sent on March 21, 2006 by Jack Dorsey, the creator of Twitter. It took only three years to reach the billionth tweet (Internet Live Stats 2015).

[172] Sentiment analysis (also known 'opinion mining') refers to the use of text analysis, natural language processing and computational linguistics to track the mood of the public, and identify and extract subjective information in any digital content (Search Business Analytics 2010). An a*pplication* p*rogramme* i*nterface* (API) is 'a set of routines, protocols and tools for building software applications. An API specifies how software components should interact and APIs are used when programming graphical user interface (GUI) components. A good API makes it easier to develop a programme by providing all the building blocks. A programmer then puts the blocks together' (Beal 2016).

[173] A meme is an idea or behaviour –a unit for carrying cultural ideas, symbols or practices— that transmits from person to person within a culture through speech, writing, gestures, rituals or other imitable phenomena. The word is a shortening of *mimeme* (from Ancient Greek μίμημα Greek

have used similar software to hunt for associations between public mood, as conveyed on Twitter, and stock market fluxes (2011). Bollen et al. claimed an 87% accuracy in forecasting future movements in the FTSE 100, FTSE 250[174] and Dow Jones Industrial Average indexes, analysing tweets (ibid.,1). Their results were convincing enough for Derwent Capital Markets, a London-based investment firm, to license their techniques and create the Absolute Return Fund – a US$40.5 million hedge fund (Jordan 2010). However, sometime later they shut it down. In spite of it, today Bollen's Guidewave Consulting offers to 'track Fear and Greed [sic] levels of millions of investors to warn you of market trends, days in advance!'[175] In May 2013, Hawtin, the founder of Derwent Capital Markets, launched Cayman Atlantic, an investment management firm that uses social media data such as Twitter to look for valuable trading opportunities, 'but instead of focusing solely on sentiment analysis – which even the most optimistic early adopter would admit was a pretty wacky idea in hindsight – it seeks to use social media to detect market-moving news before it breaks' (Milnes 2014, 3).

Correlation analysis, optimal and tailored responses, data-driven decision-making, sentiment analysis, projection and forecasting... NGOs could benefit from the tactics put in place by businesses like Walmart, Bunt Planet, Vestas and others in order to respond more immediately and individually to challenges and threats, and establish a more efficient dialogue with members and supporters, decision-makers, donors, experts and researchers, practitioners and activists, volunteers and beneficiaries. For example, Amnesty International is currently looking into how to improve its data analysis on its supporters. In order to sustain AI's political and economic independence, this organisation needs 'better and more disaggregated data about

---

pronunciation: [mí:mɛ:ma] *mīmēma*, 'imitated thing,' from μιμεῖσθαι *mimeisthai*, 'to imitate'), and was invented by Dawkins in *The Selfish Gene* (1976).

[174] FTSE stands for Financial Times Stock Exchange.

[175] See guidewaveconsulting.com/ [accessed on February 23, 2016].

how to attract more members and about our current members in order to know them better... Data on this matter is of utmost importance,' says Beltrán in an interview for this dissertation

In summary, businesses are among the earlier user of big data infrastructures, mainly for marketing purposes and to improve their efficiency, augment their competition abilities, and increase their productivity and innovation capacities. And this pioneering has inspired and trickled down to other sectors as well.

4.3 Uses by journalism

News media as a sector deserves a section apart in this dissertation mainly for two reasons that overlap. As it will be shown, first, some journalistic specialisms (i.e. investigative journalism, data journalism and advocacy journalism) have a lot in common with data activism and share several characteristics. Second, some news media organisations are producing data-based advocacy projects; and vice versa, some ONGs produce data enabled journalism on a regular basis. I dedicate more space to the uses of big data by journalism too for its role in opening the path, together with reactive data activism, towards a more intense use of data tools in other sectors, especially within civil society and NGOs.

Whether composed of public or private companies, cooperatives, non-profits or associations, funded and sustained by philanthropists, private owners, donors, foundations, shareholders, state subsidies, subscribers, audiences or advertisers, whatever their business model or structure, journalistic organisations have always worked with data (the water, flour, salt and yeast of stories). In a process described in previous sections, the relationship between democracy and journalism is one of a symbiotic dependency. For democracy to function properly, citizens have to stay informed about those in power so they can make decisions (and vote) based on real information. In order to do that, common people rely to some extent on

journalists who analyse data and find stories hidden in them. Ideally, analysed data become information, which is then transmitted through media so people can acquire, reanalyse, use, share and alter it, and finally transform it in useful knowledge. Big data analysis and visualisations today offer new avenues for journalism. Journalists are not only mediators; they can be fundamental for activism as well. Baack acknowledges also a thought-provoking link between activism and journalism. In a series of interviews, he concludes that 'activists acknowledge that intermediaries (such as journalists) are necessary to make raw [sic] data accessible to the public. This leads them to an interest in transforming journalism to become an intermediary in this sense' (Baack 2015, 1). The picture gets more complex as some campaigning organisations produce journalistic data content in order to compensate for the lack of data journalism in some countries and situations, as it will be seen.

In this section, I look first at how journalists started to use data to the point that data journalism consolidated as a new brand of journalism; define this brand of journalism; examine the impact it has had; and finally, I compare the expectations it has raised against some of the challenges ahead, and the connections that exist between data journalism and data activism.

Investigative journalists were some of the first actors to realise the power of data analysis and visualisation technologies, and to put that power into action.[176] But before them, it was civic hackers and reactive data activists who showed the way. That is why 2010, the year that WikiLeaks released the collection of Afghan and Iraqi secret documents, could be a good starting point to put a date on the beginnings of data journalism (Sampedro 2014, 4). There had been previous signs of change. Between 2005 and 2007 some digital tools and technologies

[176] For the purpose of this study, we will define 'investigative journalism' as a form of journalism in which reporters reveal a secret by deeply investigating a single topic of interest –such as a serious crime, a case of political corruption or corporate wrongdoing (Hanson 2016).

began to emerge, at accessible prices, for data analysis in journalism. In 2007, the Knight Foundation awarded a prize to the Everyblock project, which allowed users to access information about their neighbourhood based on data analysis. In 2009, the *St. Petersburg Times* won the Pulitzer Prize for its Politifact project, which analysed data on the 2008 elections. Thanks to the interest these projects generated, the Hacks/Hackers movement was created in 2009. It is an international grassroots network that connects journalists ('hacks') and technologists ('hackers'), facilitating skill exchange between the two groups (Hacks/Hackers 2016). Then in 2010, the European Journalism Centre (EJC) organised the first conference on data journalism. The following year, in the MozFest held in London and organised by Mozilla, an initiative to write the first Data Journalism Handbook[177] was born. The text is itself an example of collaborative journalism, as journalists from *BBC*, C*hicago Tribune*, *Australian Broadcasting Corporation*, *La Nacion*. *The Washington Post*, *The Texas Tribune*, *Pro Publica*, *The New York Times*, *The Guardian*, among others, participated. In 2011, the Knight Foundation awarded a prize to 16 projects related to innovation in the field of data journalism, and US$4.7 million were allocated to develop them. Today, the Knight Centre for Journalism in the Americas, Poynter Institute and Investigative Reporters and Editors (IRE) permanently offer courses online and face-to-face train journalists in data analysis, visualisation and digital tools for online journalism.

In 2010, five media organisations –*The Guardian*, *El País*, *The New York Times*, *Der Spiegel* and *Le Monde*– became WikiLeaks' partners when they reviewed and published their own analysis of the US diplomatic cables leaks, consisting in hundreds of thousands of classified cables that had been sent to the US State Department by 274 of its consulates, embassies and diplomatic missions across the world, dated between 1966 and 2010. The cables contained

[177] See datajournalismhandbook.org [accessed on February 23, 2016].

analysis, assessment and commentary about the foreign governments and their leaders and officials, revealing the US government's view of the world. The five newspapers had agreed to publish cables, 'which had been subjected to a thorough joint editing and clearance process,' without including the identities of the sources in order to protect them (Ball 2011). In 2011, though, WikiLeaks make public the 250,000 diplomatic cables in their entirety, without protecting the identity of its sources, an action that was criticised for not adhering to proper journalistic standards. Although relying on undisclosed sources is controversial, it is a good journalistic procedure to protect the identity of the sources if the information has been confirmed independently. Reporters Without Borders, a press freedom group that had been maintaining a backup version of the WikiLeaks site, revoked its support for the whistle-blowing site in the wake of the decision (ibid.). In spite of this apparent breach in journalistic standards, Sampedro defends WikiLeaks and notes that the leaks had circulated already during a whole year without any consequences, since a *The Guardian* journalist –David Leigh— had published their password in his book *WikiLeaks: Inside Julian Assange's War on Secrecy* (2014, 193).[178] In *El cuarto poder en red*, Sampedro quotes WikiLeaks as example of how to put in practice a prototype of critical and free press for the common good, and argues that the new journalistic practices rooted in collaborative technology and data mark the dawn of a more innovative and better journalism. 'Hackers that reveal information following high ethical standards are

[178] Assange is the editor-in-chief of WikiLeaks, which he co-founded in 2006 after a career in hacking. WikiLeaks attained celebrity in 2010 when it published US diplomatic and military documents leaked by Chelsea Manning, who was convicted in July 2013 of violations of the Espionage Act and other offenses, and was sentenced in 2013 to 35 years' imprisonment. Assange has been under investigation in the US ever since. In the same year, the Swedish Director of Public Prosecution opened an investigation into sexual offences that Assange is alleged to have committed. In 2012, facing extradition to Sweden, Assange sought refuge at the Embassy of Ecuador in London and was granted political asylum by Present Rafael Correa's government.

recuperating the most valued ideals of the old muck-racking investigative journalism,'[179] according to this author. 'The trip that hacktivists[180] have initiated is a return trip: the journalism that is coming is going back to its roots.' Journalists must emulate civic hackers so they can regain 'democratic legitimacy' in their profession (2014, 14).

> Hackers have become a model for investigative journalists; and hacktivists, for militant activists and politicians. Before this very attractive language is trivialised, it is worth noting the radical change that these actors propose: a change that goes back to journalism's roots and holds very high aspirations. It is not easy to understand at first sight, and predicting its results is even more difficult. But we are all being affected by it (ibid., 14).[181]

The implications of what Sampedro is proposing are weighty. Access to data is a huge challenge, not only for journalists but also for activists, since the biggest data vaults are in governments' and corporations' hands. In places where governments do no practice open data and corporations guard their data, some activists are currently experimenting with methods to generate and appropriate data (i.e. community drones), as it will be discussed in the next section.

Then again, in 2013 Snowden –a computer analyst whistle-blower who had worked for the US Central Intelligence Agency (CIA)— provided several media outlets with top-secret US

---

[179] The term became popular to be applied to meddlesome journalists after US President Roosevelt referred to a character in John Bunyan's *Pilgrim's Progress* in a 1906 speech: 'It is very necessary that we should not flinch from seeing what is vile and debasing. There is filth on the floor, and it must be scraped up with the muck rake; and there are times and places where this service is the most needed of all the services that can be performed. But the man who never does anything else, who never thinks or speaks or writes, save of his feats with the muck rake, speedily becomes, not a help but one of the most potent forces for evil' (Roosevelt 1906, 2).

[180] The term is a combination of 'hackers' and 'activists.'

[181] 'El hacker se ha convertido en un modelo para el reportero de investigación. Y el hacktivista, en ejemplo del militante o el político. Antes de que se banalice este lenguaje tan atractivo, merece la pena señalar el cambio radical que estos actores promueven. Un cambio que va a las raíces y apunta muy alto. No es fácil de entender a la primera y, menos todavía, predecir sus resultados. Nos afecta a todos, aunque de modo diferente' (ibid.).

National Security Agency (NSA) documents leading to revelations about systematic massive surveillance on phone and internet communications in the US, with the connivance of European governments. Whether WikiLeaks is a journalistic organisation (it defines itself as a 'not-for-profit media organisation') or not, it a matter of debate. What several media organisations did with the information provided by WikiLeaks and Snowden was indeed journalism. In both cases, WikiLeaks and the Snowden papers, different media organisations had access to big amounts of data, analysed and published the findings of their analysis in an accessible manner. According to Sampedro, though, the newspapers that benefited from the leaks did not do much more than exploit the civic hackers, without taking any risks. They did not support Assange when he was in trouble either. In spite of profiting from the leaks, the media organisations that published them did not do anything to guarantee their continuity or protect the legal situation of their originators. Reaction to Manning's revelations was mixed. Nicks, one of her biographers, writes that the leaked material was widely seen as a catalyst for the Arab Spring (2012, 212). For Sampedro, the new data-, whistle-blower-based collaborative journalism constitutes a 'network fourth state,' defined as a series of practices and technological organisational models that, together, play the role previously associated with 'the press,' which is to offer a public control over the other three powers: parliament, judiciary and executive. These practices include new forms of transparency and participation. The resulting journalism becomes a civil society's platform for information debate and consensus, and an open tool box for communication (2014, 210-231).

Access to data, as it will be seen later, is a key issue. In some countries in Latin America, the relatively quick establishment of transparency laws have stimulated data projects (Ávila 2015). In Argentina *La Nación* has set up a joined team of programmers and specialised data journalists in order to produce data-enabled stories. The newspaper has been nominated for

several awards, also has a datablog, which has among its regular contributors Crucianelli, who has pioneered in data journalism in Spanish and spends much of her time training journalists. In Brazil, *O Estado de S. Paulo* has a coordinator, two journalists and a developer writes code for visualisations or applications (Zanchelli and Crucianelli 2012, 11).

News media organisations use big data infrastructures not only to investigate and denounce, but also to maximise their impact and the distribution of their content. Digital born media, such as the *Huffington Post*, with 22 million monthly visitors at the end of 2015, have utilised data to enhance and strengthen their business models. This media outlet is one of the most trafficked and fastest growing news and entertainment websites in the world (Stone 2014, 5). *HuffPost* uses big data to 'optimise content, authenticate comments, ensure efficacy of native advertising, regulate advertising placement and create passive personalisation' (ibid.).

But what is data journalism? For Congosto, what separates traditional, investigative and data journalisms is the different sort of sources of information these three brands of journalism use. Conventional journalism is 'based on people,' that is, it is based on interviewing and listening to people talking (i.e. in a press conference or a face-to-face interview). Meanwhile investigative journalism relies on documents that 'cannot be processed automatically,' whereas data journalism depends precisely on data infrastructures, which allow automatic processing of a vast amount of data until relevant information is extracted from them. 'This automatization permits to scale up the volume of data (that can be processed) to unimaginable levels' (2012). She thinks that the data visualisation tools accessible today to almost anyone owe a debt of gratitude to journalism as well. Cairo, interviewed for this dissertation, agrees. To his surprise, his book *The Functional Art* (2013) quickly became a text of reference not only for journalists, but also for 'scientists, statisticians, business intelligence people. ' That is, the fact that nowadays

data activists can access and use big data infrastructures and visualise them with relatively ease is due to a group of journalists who have pioneered and popularised the art of using them. 'Visualisation is infographic journalism's daughter,' and has inherited its visual language and the synthesis capacity to transmit information effortlessly (Congosto 2012). Having being an investigative journalist for a number of years, spoken to numerous sources and gone through countless documents, I have a clear idea of what distinguishes one type of journalist from another. It is not exactly the type of sources a journalist uses, although there is something in what Congosto says, but what the main purpose of the journalistic effort is. Investigative journalists' main objective is to shed light on a secret that some power –usually a government, a corporation or a criminal organisation— is interested in covering up. Conventional journalists mostly disseminate relevant and factual information (it could be based on any kind of source, whether documents or people), while data journalists use data technology either as an analytic or a communication tool, or both. Data infrastructures can be employed in an investigative story, but the fact that one uses data tools does not transform the product of it in an investigative story automatically. On the other hand, data journalism is usually associated with fancy interactive visualisations of data. In the same way that visual tools are not the whole story about data journalism, visualisations are not even a prerequisite for data journalism. One example of (advocacy) data journalism that does not resort to visualisations as a communication tool is the TV report broadcasted in 2012 by *ViveloHoy* –part of the Tribune Publishing of Chicago— about the suspiciously high number of jaywalking arrests of black people in Champaign and Urbana, Illinois, based on US Census Bureau data requested via Freedom of Information Act (known as FOIA). The *ViveloHoy* journalists who did the story first requested the official data, then analysed and visualised them on a map, realising, a bit as John Snow did, that there was a

concentration of jaywalking arrests in two black areas of the city. They got out of their office and went to these neighbourhoods to see for themselves how these areas looked like, and to interview their inhabitants. The report eloquently showed, doing traditional interviews on camera, that the communities where jaywalking was rampant simply had no sidewalks and were mostly black (Lowenstein 2012a). Is this a good instance of data journalism? Unquestionably so, even if *ViveloHoy* journalists did not use visualisations to communicate their findings, but good meat-and-potato TV journalism. *ViveloHoy* submitted a Freedom of Information Act (FOIA) request to the police departments in Champaign and Urbana asking for the records for each arrest in the years 2007 to 2011; and not satisfied with just getting the data, they forged an alliance with a civil rights group, and they also told their audiences about the process and the results. Their position is very clear from the start: 'We're privileged this year to be working on a project with our friends and colleagues at Citizen-Access[182] about demographic change in sixteen counties in Central Illinois' (Lowenstein 2012b), That is, with a series of journalistic data stories, *ViveloHoy* aimed at 'democratic change' in Central Illinois. Is it a case of investigative journalism? Not really, because the data in which the report in based were publicly available in the first place, and not a secret or concealed intentionally (i.e. relevant data can be buried deliberately in vast, but irrelevant, datasets). *ViveloHoy*'s virtue was to find a pertinent story hidden in accessible public data and to tell the story eloquently. Is this a case of advocacy? Certainly. *ViveloHoy*'s intentions to change the *status quo* were transparent from the beginning. What made this story different from other types of journalism? 'Perhaps it is the new possibilities that open up when you combine the traditional "nose for news" and ability to tell a compelling story, with the sheer

[182] See cu-citizenaccess.org/ [accessed on April 18, 2016]. This organisation practices data activism in food, housing, justice, education, public funding and health care issues. It provides data visualisations and applications, as well as the data behind the visualisations in Illinois. It says: 'CU-Citizen Access is a community online news and information project devoted to investigative and enterprise coverage of social, justice and economic issues in east central Illinois' (CU-Citizen Access 2016).

scale and range of digital information now available' (Bradshaw 2011). This hereafter data journalism will be understood as the craft of extracting stories out of datasets or telling stories using data visualisations, by means of social science and computational methods (Howard 2014).

This type of journalism has meant a breath of fresh air for tired and under siege journalistic organisations.[183] Data have come to journalism rescue. 'Let's admit that WikiLeaks has put us in our place… If we had continued this way, the professionals and professors of journalism were cadavers; or worse, zombies dragging their past across an apocalyptic landscape. From 2008 to 2012, (in Spain) almost 200 media organisations disappeared and 8,000 jobs were lost, and there were 27,443 journalists without work' (Sampedro 2014, 10).[184] This is very relevant because of the central role journalism has in any democracy. As we have seen, Habermas criticised the deterioration of the public sphere and the commodification of culture and private affairs, reinforced by the mass media, which turned citizens into passive spectators and created a pseudo-public sphere (1991, 171). In the same vein, Sampedro highlights the trivialisation of content in mainstream, conventional media, especially in the type of information disseminated by big media groups serving their economic interests. He criticises how big corporations controlling media organisations work, particularly as they become megaphones of generalists –experts with

[183] The media sector has experienced an enduring crisis as a result of the popularisation ICTs, which allowed anybody with a smartphone or a computer and an internet connection to become a *journalist* and create his or her own content, as well as access any other content, in a phenomenon that Castells has called mass self-communication (Castells 2009). In mature markets, such as Europe and the US, conventional circulation and audience figures have dropped on a yearly basis since the 90s. Parallel to this downward trend, traditional publicity has fallen as well in favour of digital ads. People and ads favour aggregators such as Google and Yahoo over media organisations' web sites (Mitchell 2015). Meanwhile newsrooms have withered, journalists have been fired *en mass* and foreign desks have been closed down in order to make media organisations economically viable. Many of them succumbed and went out of business, sometimes in connection with problems of credibility as well (i.e. *News of the World*).

[184] 'Reconozcamos que WikiLeaks nos puso en nuestro sitio… De seguir así, los profesionales y profesores de periodismo éramos cadáveres. O peor, zombis que arrastraban su pasado, por un escenario apocalíptico. De 2008 a 2012, desaparecieron casi 200 medios y se perdieron 8.000 empleos, con 27.443 periodistas en paro' (ibid.).

opinions for all intents. For Chomsky, mass media are interested in entertaining people and diverting their attention from real issues (1997, 1-2). Concurring, Castells criticises also the effects of the concentration of media ownership and the deterioration content quality as a result of it (2009, 56). Data journalism seems to me a remedy for some of these ailments: it deals with issues that matter (i.e. corruption or secret surveillance plans); it does not resort to the he-said-she-said sort of journalism than reigns in some media, but rely on data and analysis, and lets people make their own judgement and in many cases access the data directly; and it has been the basis for new alternative business models that challenge the way big media groups work. Data journalism represents good journalism's comeback, and more. Another profound change that data journalism has had is a return to reason and the best argument, where *opinionism* had the upper hand so far. Within journalism, Sampedro describes the appropriation of data as a process of empowerment and legitimisation of journalists too (2014, 211). Data have allowed journalists to take a 'trip' that has taken them back to their roots of muckraking[185] (ibid., 14).

Over the past few decades, news media organisations have suffered generalised and continued drops in circulation, readership, audiences and finances. This has compelled many media organisations to fire journalists, close down newsrooms and foreign desks, and ultimately fail as businesses. 'Journalism is under siege' (Lorenz 2012). Having stated that journalism is vital for democracy, the advent of data journalism is extremely significant. The ongoing impact of data on journalism has been intense to the point that some signs of regeneration can be observed. For example, data journalism has injected new blood into journalism as businesses, especially in the US. 'In many ways, 2013 and early 2014 brought a level of energy to the news

[185] The muckrakers where a group of US writers who published journalistic exposés of cases of corruption and abuse, and advocated for social reform before the World War I. The name *muckraker* was initially pejorative, but took on positive connotations of these writers' social concern and courage (Encyclopedia Britannica 2014).

industry not seen for a long time. Even as challenges of the past several years continue and new ones emerge, the activities this year have created a new sense of optimism – or perhaps hope – for the future of American journalism' (Mitchell 2014). Even if it does not attribute this 'level of energy' explicitly to the vigour of data journalism, 'The State of the News Media 2014' –which has been a witness of news media's failures of the past few years— points to factors such as dynamism of the new digital-native outlets, new money flowing into the industry 'fostering new ways of reporting,' social and mobile developments 'changing the dynamics of the process itself,' and 'new ways of storytelling' as reasons for it (ibid., 5). Examples of data-based 'new story telling' abound. A look at the Data Journalism Awards, granted by the Global Editors Network (2014), shows the variety and excitement of data journalism, from *La Nación*'s 'Declaraciones juradas' –about the private assets among Argentina's main public officials— (La Nación 2014) to 'Reshaping New York' –about how the city has changed 'in 12 years of Bloomberg,' the major— (The New York Times 2013). Other notable examples include, for example, 'Vidas contadas,' a project half way between journalism and activism.

This progress has accompanied the growth of tools for data analysis, which are also proliferating. 'You can find lists, guides and case studies on a number of journalism websites, including DataDrivenJournalism.net, Source and the Poynter Institute's Digital Tools Catalogue, among others' (Stencel, Adair and Kamalakanthan 2015). Conferences, workshops, *hackathons*[186] and other events around data and journalism have also increased. For example, attendance has jumped at data reporting workshops and seminars hosted by the Online News Association and the Investigative Reporters and Editors' National Institute for Computer-

[186] A hackathon is an event in which computer programmers, coders and others who collaborate intensively on a software project, typically for a brief period of time, between a day and a week. Hackathons tend to have a focus, which can include the programming language used, the operating system, an application, an application programming interface (API), or one issue (Technopedia 2016).

Assisted Reporting (ibid.). The copiousness and diversity of examples, the exuberance and dynamism that data are conferring to journalism show that data are indeed becoming tools in transforming this industry. And journalists are using data to move even beyond journalism, and jump from observation and analysis into denunciation, raising awareness and recommendations, a form of journalism called 'advocacy journalism' –which could be called 'advocacy data journalism.'[187] A clear example is 'Keep it on the ground,' a campaign where *The Guardian* openly mixes advocacy and data journalism to foster massive greenhouse gas (GHG) emission cuts (The Guardian 2016). Empirical evidence also shows that there is a plethora of new players in the field of journalism, partly fostered by data. In Spain, for example, the massive layoffs experienced in mainstream[188] newspapers and media organisations has inspired (or forced) sacked journalists to come up with new business models, some of them in the form of cooperatives where employees are shareholders as well. *Eldiario.es*, which is a spinoff of a spinoff (founded by former journalists of *El Público*, which, in turn, was created in 2007

[187] There is no accepted definition for advocacy data journalism. Advocacy journalism is to be understood hereafter as 'the use of journalism techniques to promote a specific political or social cause' (Jensen 2016). Advocacy data journalism is advocacy journalism that is enabled by data infrastructures (Milan and Gutiérrez 2015).

[188] Hereafter, 'mainstream media organisations' are to be understood as traditional media organisations that direct their content to mass audiences (Chomsky 1977). They normally disseminate the values and ideas of the elites, to use a concept developed by Marx ['the ideas of the ruling class are in every epoch the ruling ideas' (Marx and Engels 2000)]. In the liberal capitalist world, these ideas have capitalism as a context. In contrast, 'alternative' news media organisations are organisations that may include 'webzines, small-circulation religious bulletins, hobby listservs, cell-phone use in the demonstrations that brought down Philippines president Estrada in 2001, *graffiti*, as well as politically oriented national and international examples, such as *Il Manifesto*, global labour media, the Indymedia network and – at a radically different political location – neo-Nazi rock music or jihadist communiqués' (J. D. Downing 2004). Fuchs sees alternative media as 'critical media' or 'the communicative dimension of the counter-public sphere' (Fuchs 2010, 173). He further adopts the notion of 'proletarian counter-public sphere' as a 'radically different' public sphere from the 'bourgeois public sphere,' and as an expression of the emancipation of the working class (ibid., 176). However, I find this distinction redundant. The 'bourgeois public sphere' is an idealisation, whose characteristics can be applied to any *class*, once removed the bourgeois label. Fuchs definition is very useful: alternative media are considered 'self-organised, citizen-controlled, self-managed, self-owned, non-commercial, non-advertising media' (ibid., 188), and are media that produce 'content that formulates visions of an alternative world beyond capitalism' (ibid.).

profiting from a wealth of redundant seasoned journalists), was created in 2012 as a progressive online daily that uses data journalism to uncover cases of corruption, explore politics and report on human rights.[189] More than 50% of the company that owns this daily is on its journalists' hands. They practice data journalism often. Other media organisations emerged from the economic crisis include *lamarea.es* (created in 2012 as another cooperative of journalists); *Infolibre* (set up in 2013, owned partially by the journalists themselves); *Materia* (a media organisation dedicated to science created in 2012); *Sancho Panza* (another cooperative that was launched in 2012); *El Español* (founded in 2015); and Okdiario.com (founded in 2015). Some of them practice investigative journalism and some even practice data journalism; all of them are digital natives determined to refurbish and transform journalism in Spain. The same exuberance can be observed in other countries and markets. However, Spain may be a special case. In a context of a new corruption scandal every day and a deep crisis of government,[190] Spain was affected by the media sector with the lowest levels of trust, according to a Reuters Institute study (2016, 12). Only 34% of Spaniards trust their news media in comparison with two-thirds (68%) in Finland (ibid.). Although these new digital media organisations seem to apply higher standards of journalism, apparently, they have not done enough yet to change this perception. In this case, a mixture of factors may have coalesced to generate mistrust, among them the profound financial crisis that has hit media businesses hard; the corruption scandals that reveal a rotting political class; the youth of the democratic media organisations (i.e. *El País* was founded in 1976, whereas, for example, *The Guardian* was founded in 1821); the support that many media organisations offer to concrete political parties and big business; the low level of openness of the institutions and the newness of the transparency law; the concentration of media

[189] See eldiario.es/ [accessed on February 23, 2016].
[190] From December to October 2016, Spain remained without a government and had to endure ten months of failed deals and squabbling.

ownership in groups that control all aspects of content production and dissemination; and a generally slow answer to new demands of honesty, transparency and innovation by the citizens. However, by bolstering journalism, data infrastructures and techniques have made journalists recover some of their credibility and legitimacy (Lorenz 2012). And as I have been pointing out, a healthy journalism is the basis for a healthy democracy, whose citizens depend on journalists – salaried or not, citizen[191] or not— for access to information, analysis and debate to make intelligent decision. In short, data journalism seems to be the single best good news reaching the media industry in the past few decades. In a defective public sphere where opinions flow and clog the wires, where so called experts' ideas and positions triumph by acclamation, data and facts –once the substance of journalism— are handing the journalistic profession back to the journalists, to those who practice journalism and adhere to the rules of journalism as Kovach and Rosenstiel defined them in *The Elements of Journalism* (2007, 10). These values, together with the motto 'Comment is free, but facts are sacred' (Scott 2002), are back thanks to data journalism. The ideals behind the sentence coined by former *The Guardian* Editor C.P. Scott were the fundamental standards of a free press, and have been invigorated by data journalism today. A sign that these values are trendy again is that precisely *The Guardian* –an example of an open news media platform that is pioneering in data journalism— has made them its slogan.

An example of data journalism's influence on policy are the International Consortium of Investigative Journalists'[192] data stories on tax paradises that prompted 'high profile resignations, criminal and civil inquiries, policy changes and official investigations on four continents' in

[191] Citizen journalism is to be understood hereafter as the 'collection, dissemination and analysis of news and information by the general public, especially by means of the internet' (OxfordDictionaries 2016). Citizen journalism is part of what Castells has called 'mass self-communication communication' (Castells 2009).
[192] See www.icij.org/ [accessed on June 21, 2016].

2014 (Cabra 2014). Among the impacts, the UK Treasury announced that following the lead of the Cayman Islands, all British overseas territories agreed to share information about individuals holding bank accounts in their jurisdictions with the UK, France, Germany, Italy and Spain (Porteous, Hudson and Chavkin 2014). And Europe's five biggest economic powers — Britain, France, Germany, Italy and Spain — announced they would begin 'regularly exchanging banking and tax information as a way of identifying tax dodgers and other financial wrongdoers' (ibid.). Moving away from this specific example of impact on policy, McGreevy mentions several cases where data journalism is generating social impact, including the positive effect an initiative called Chequeando[193] —which employs volunteering fact-checkers verifying public utterances— is having on political discourse in Argentina; the support Famine Strikes Again[194] is providing to draught mitigation efforts in Kenya; and the catalyst role that a platform such as SocialTIC[195] plays in Latin America, enabling data journalistic projects. 'Through the strategic use of ICTs, data journalism trends and experiments in Latin America have included crowdsourcing[196] ventures such as Ojo Público,[197] user interfaces such as Poderopedia[198] and data driven narratives such as InfoAmazonia'[199] (2015).

As stated, there are new ways in which to practice journalism through collaborative methods and based on data. And with big data infrastructures, anyone can do it now, right? 'Data journalism is the new punk' (Rogers 2012). Then again, I do not necessarily agree with this

[193] See chequeado.com/ [accessed on June 21, 2016].
[194] See internewskenya.org/summaries/internews52e7747b74fff.pdf [accessed on June 21, 2016].
[195] See socialtic.org/ [accessed on June 21, 2016].
[196] Crowdsourcing is a process of creating something collectively (a service, an idea, research, content, data) by asking contributions from a large group of people, rather than from traditional stakeholders. Digital technologies make this process simple and easy, allowing almost immediate demands and responses, and accruing and rating positive contributions.
[197] See ojo-publico.com/ [accessed on June 21, 2016].
[198] See poderopedia.org/poderopedia/index/chapters [accessed on June 21, 2016].
[199] See infoamazonia.org/ [accessed on April 1, 2016].

assertion, since although access to data infrastructures has been lowered, doing a data-enabled story is not such a straightforward process. To start with, access to data is an ongoing problem in spite of the proliferation of open data platforms, and sometimes even journalists equipped with the last scraping and extraction tools[200] have difficulties in getting their hands on data. This can be challenging even in democratic countries where there are laws that grant access to information and have open data web sites, but especially difficult and dangerous in authoritarian or semi authoritarian regimes. One example of the former is simply accessing local city data in many Spanish cities. San Sebastian, for instance, publishes a pdf of its municipal budget with some degree of disaggregation, but this constitutes by no means open data, since there is no access to the data behind the pdf on a usable format. In terms of national open data platforms, Taiwan's is rated the most open government by the Global Open Data Survey[201] (with 78% of the information being open), while Spain is rated number17 among 122 governments (with only 55% of public information being open). Global government openness has dropped from 14% in 2014 to 9% in 2015. The most open information is usually about national statistics, containing mostly anonymised demographic information, while the least open information in all governments is about government spending, with the unique exception of Greece (Global Open Data Index 2014). This is shocking when one thinks this is information about how governments administer everyone's tax money. Once the data are obtained and polished, which is not an easy or a quick task, the person examining them needs the ability to use other tools to analyse and visualise them, which means, at times, dominating the art of statistical calculation, code writing and visual thinking. It is true that the barriers to do so today are much lower thanks mostly to the

[200] A web scraping tool is a software technique of extracting data from websites. Usually, such software programmes mimic human exploration of the World Wide Web by either implementing low-level Hypertext Transfer Protocol (HTTP) or embedding a web browser, such as Mozilla Firefox.
[201] See index.okfn.org/place/ [accessed on February 23, 2016].

collaborative work of pioneering reactive data activists and civic hackers. Recapping, complex data show patterns and tell stories, although not by themselves. Most journalistic tactics take a deductive approach to data. However, Cabra reveals in her interview for this dissertation that organisations such as the ICIJ, where she heads the Data & Research Unit, is looking at pouring complex, big and apparently unrelated datasets onto the *machine*, and let it do the job of identifying trends and correlations through machine learning, without any expectation, research question or hypothesis formulated in advance. This approach would involve even more sophisticated tools and analysis. The advantages are clear: to be able to extract correlations from datasets so vast that no human eye would be able to spot them, and to identify connexions so concealed that no human brain would be able to imagine in advance as research questions.

Nevertheless, data analysis and visualisation are more and more popular in journalism. However, even the pioneers of data journalism are lagging behind the promise. The authors of *The Goat Must Be Fed* conclude that 'many US newsrooms are not taking advantage of the emerging low-cost digital tools that enable journalists to report and present their work in innovative ways. Editors and producers cling to familiar methods and practices even when they know better' (Stencel, Adair and Kamalakanthan 2015). Stencel, Adair and Kamalakanthan also note that the proliferation of data journalism awards, events and conferences create 'a sense that the adoption of data reporting and digital tools is broader than it really is,' and that 'there is a still significant gap between the industry's digital haves and have-nots' (ibid.). The key obstacles for a general adoption of data infrastructures and methods, apart from budgetary concerns, are 'a lack of technical understanding and ability, and an unwillingness to break reporting habits that could create time and space to experiment' (ibid.). If US newsrooms have not adopted data infrastructures universally yet, especially in small mainstream

media organisations, the situation is even more precarious in other regions where the data revolution has been more gradual. 'While major newsrooms in the US were able to incorporate teams of programmers, statisticians and data visualisators, in the Spanish-speaking world there were only isolated efforts and, in many cases, at the initiative of civil society' (Ávila 2015). This last remark is really significant because, in the absence of scientific surveys, the impression from empirical observation is that civil society organisations have taken the role of media in exploring the possibilities of data analysis-based storytelling. Looking at concrete initiatives, there is, also, a great amalgam of organisations using data analysis for journalistic reporting. Further, data alone are not enough to 'save' journalism, and the data craze may be having negative impacts as well. Kaplan notes that databases, network analyses and code are not a proxy for investigative journalism, and that currently donors are so infatuated with technology that are no longer funding quality content (2013). Kaplan even says that this fixation with technology 'threatens to diminish what funding exists' for journalism (ibid.).

The Law of Freedom of Information (FOIA) has been a great incentive for data journalism in the US. This law was signed by President Lyndon Johnson in 1966, and it went into effect the following year. In contrast, in Spain, for example, the Transparency Law –which regulates access to data— came into force only in 2013. Time gaps of this magnitude translate in gaps in the production of data reports. Although not so extreme, other Spanish-speaking countries experience similar situations. 'Within the Spanish-speaking world, Mexico, while there is a transparency law that took effect in 2003, efforts to use the databases to do data journalism have been isolated' (Ávila 2015). Ávila highlights the fact that, in the Spanish-speaking world, civil society organisations, such as Medialab-Prado[202] and Civio,[203] in Spain, or Fundar,[204] in Mexico,

[202] See medialab-prado.es/?lang=en [accessed on February 23, 2016].

seem more proactive in using data and producing data stories than journalistic organisations (ibid.). And even if access to these technologies is by no means uniform or widespread within civil society, NGOs and campaigning organisations are currently exploring data with enormous creativity, bearing and energy, founding and encouraging proactive data activism, the focus of this dissertation.

Summing up, the impacts of data journalism in journalism can be précised as follows: data has given back journalism to journalists, since it demands a certain level of skills and standards that a simple blogger may not necessarily master.[205] It has reactivated the sector, making it easier to practice good journalism and is the basis for new ventures in a sector that seemed exhausted. As a result, there are more players and more enthusiasm in the form of new conferences, tools, awards and media organisations. Finally, it has given back some credibility to journalists after the fiascos that plagued the post-Iraq invasion, which is probably the starting point of a general deterioration trend.[206] Outside journalism, it has managed to change policy and public opinion, since by definition, many investigative data stories are very impactful. However, whatever the approach, doing data journalism is not for any kind of *punk*. It requires skills that have to be

---

[203] See civio.es/en/ [accessed on February 23, 2016].
[204] See fundar.org.mx/ [accessed on February 23, 2016].
[205] I refer to professionals that practice journalism on a regular basis. Whether they are salaried or not, whether they are employed by a mainstream media organisation or a community radio is of no relevance.
[206] This is a very complex issue that requires an explanation. Some media organisations in the US collaborated with the US government in disseminating a biased view on the invasion (Gutiérrez 2004). *The New York Times* even apologised for it. In a 2004 editorial, it admitted that, after reviewing their Iraq coverage, 'a number of instances of coverage that was not as rigorous as it should have been' had been found. 'In some cases, information that was controversial then, and seems questionable now, was insufficiently qualified or allowed to stand unchallenged' (New York Times 2004). Although I am not basing my opinion on surveys, my impression is that, together with a profound economic crisis during the past few decades, news media organisations in rich countries have experienced a parallel credibility crisis that is related, in part, to new phenomena such as embedded journalism and the biases found in the coverage after the Iraq war (Gutiérrez 2003). In the US, 'only 24% now think that the news media try to report on news without bias' (Newseum 2015, 5). Although it has been noted that big databases may hide biases and gaps, in the context of ta credibility and economic crisis, data journalism has meant a return to the basic values of journalism.

acquired. That is why, even if there is great enthusiasm around data journalism since the first WikiLeaks cables in 2010, the data promise has not been fully realised, as I examine in the next section.

5. Uses by civil society: Data-enabled activism

   Having hard numbers means that non-profits can no longer ignore the math of social change. The fight for social/environmental justice, regardless of the form the battle takes, is deeply imbued with emotion: we want to hear about the life changed, the forest preserved, the child educated, the disease eradicated, the war averted. And to be sure, whether one donates to a non-profit, works for an organisation that seeks to better the world, or volunteers for a cause, one's time, money, sweat and tears will almost always achieve a positive impact, however infinitesimal (West and Posner 2012).

   To date, stories of big data's progress and successes have tended to come from government and the private sector, but we've heard little about its relevance to social organisations. Yet big data can fuel big social change (Small and Anderton 2014).

So far, I have examined what big data are and how the new data infrastructures have made it possible to extract and analyse vast quantities of previously unfathomable data and metadata, whose volume, variability and value (once in a usable format) are unprecedented. I have gone through the proto-history of big data and their infrastructures, and defined all the key concepts needed in order to outline and contextualise proactive data activism, the focus of this dissertation. Finally, I have reviewed some of the critical approaches to data, with emphasis on those applied to humanitarian assistance, the sphere of action of the case study selected for this dissertation: the Ushahidi platform. In this part of the dissertation, I have examined so far some of the social uses of big data in the private and public sectors, as well as in journalism, since the lines dividing these different sectors are sometimes blurred when it comes to concrete cases, sometimes as a result of the use of data. Some of the solutions found in these sectors are being used, or have the potential to be used, by advocacy efforts as well. I have referred too to civic hackers, the first reactive data activists, because of their role in pioneering new data infrastructures, tools and practices that are being applied to both journalism and activism.

In the following section, I look more in-depth at the social uses of data, as well as their infrastructures, within civil society and campaigning organisations, including TANs, since this is the ecosystem in which data activism has emerged. Henceforth, I review the raise of the social movements that are connected to the proactive data activism case study included in this piece of research –the Ushahidi platform—, among others. I start with a detailed description of data activism and its varieties, and of the 'imaginative spaces' recreated by data activists to attend to 'the human need for sharing, bequeathing, consoling, mourning and hoping' (Berger 2008, 42) when they use data and data infrastructures in a critical and political manner for social change.

5.1 Reactive and proactive data activism

In 2014, after a series of lectures at Tilburg University, Stefania Milan and I embarked in the adventure of describing the emergence of a new phenomenon: what we called 'data activism.' The idea was to explore the opportunities for social change opened up by big data and how normal citizens were 'becoming increasingly aware of the potential of data for social change' (2015, 124). We also looked at how data activism had appeared, to conclude that it drew on social practices that had emerged from a 'critical approach to big data' and a movement of resistance to the massive data gathering and storing by corporations and governments –what had become a post-modern panopticon. That is, reactive data activism came first and showed the way to proactive data activists, among others. Accordingly, we identified two sub-fields of data activism: 'reactive' and 'proactive' (2015, 127). Data activism surfaced as a reactive movement 'in the fringes of society,' quickly advancing from a marginal to a dispersed one. This process happened as their technologies got more accessible and popularised, and they transferred from the skilled, elitist hackerism to the commoner user. However this happened also because these reactive data activists evolved into less antagonistic movements, the same way some *black hats*

became *white hats*, or hackers with a cause[207] (Gorenstein-Massa 2013, 63), and their values and methods became more widely acceptable. Reactive data activism, we concluded, relates to 'resistance to the threats to civil and human rights that derive from corporate and government privacy intrusion.' Proactive data activism, instead, takes advantage of the opportunities delivered by big data infrastructures and the new knowledge that data analysis can provide, and equips itself with the weapons of data and facts for civic engagement (ibid., 127). Both 'are enabled by software to manipulate data or to shield one's online interactions from intrusion and automatized collection' (ibid.). We locate this phenomenon in the intersection of other existing fields of human action, namely communicative processes, such as activism, advocacy and 'the subaltern counter-publics interested in empowerment through media and technology,' and information- and technology- related professions, such as data analysis and journalistic investigation. This intersection is explained as follows:

> When big data, and data analysis practices and tools, meet citizens' media and other grassroots experiences that put empowerment though active engagement with media at the centre, there emerge instances of "civic hacking". Civic hackers are technologists and open-data activists who engage with datasets to address challenges relevant to their community. On the contrary, when citizen's media encounters activism and advocacy, we witness the emergence of media activism. In turn, when activism and advocacy join journalistic practices and values we observe the emergence of advocacy journalism. When big data intersects journalism, and investigative journalism in particular, we have data journalism. It is, however, at the intersection of more quadrants that the most noteworthy trends come into being. When advocacy journalism and investigative journalism cross

[207] Black hat hackers are also known as *crackers* or dark-side hackers. While hackers build things, crackers enjoying breaking things and infringing rules (Urban Dictionary 2016).

paths with available data and data-analysis software, we have the sub-field of "advocacy data journalism", which combines the traditional values of journalism with crunching data and a social change ethos. At the juncture of the four circles, finally, we find "proactive data activism" (ibid., 134).

These distinctions are important here, since the case study chosen for this dissertation is an instance of a class of phenomena (proactive data activism) that provides an analytical frame that the case illuminates. The differences between reactive and proactive data activism, as well as other areas of related activity, are, therefore, significant.

Reactive data activism's representatives could be considered part of 'radical tech groups' focussing on 'counteracting commercial as well as state pressures on information content, media access and the privacy of media users, and the current drive towards pre-emptive mass surveillance' (Milan and Hintz 2013, 10). These groups extend protection to their members against the pre-emptive surveillance of the panopticon by offering web-based services such as 'website hosting and blogging platforms, e-mail and mailing list services, chat rooms and other tools such as anonymous remailers; or provide platforms for self-production of information and SMS[208]-to-website services in the occasion of protest events' (ibid.). Milan and Hintz note that some of these groups were among the pioneers of internet development in the early 1990s, and have contributed to innovation on the web. Examples of radical tech groups offering non-commercial[209] digital tools and platforms are the Spanish SinDominio,[210] the Italian Autistici,[211]

[208] Short Message Service (SMS) is a text messaging resource delivered by telephone, web or other mobile communication systems. In this dissertation, SMS can refer social movement studies as well.
[209] Hereafter, non-commercial will be understood as an enterprise or initiative that does not seek to produce an economic benefit. 'Non-commercial means not primarily intended for or directed towards commercial advantage or monetary compensation' (CreativeCommons 2014).
[210] See sindominio.net/ [accessed on February 26, 2016].

the British Plentyfact,[212] the North American Riseup.net[213] and the open-publishing platforms of the Independent Media Centre (known as Indymedia).[214] Some of these groups also engage in advocacy and raise awareness of issues such as privacy protection, intellectual property and access to knowledge, therefore they are also campaigning organisations of a kind (ibid., 12). Reactive data activism has been observed before, alas not terming it so, by authors including Milan and Hintz (2013), Gorenstein-Massa (2013) and Barnes (2000), among others. That is because reactive data activists are a subgroup of 'radical tech groups;' their idiosyncrasy is that they are 'tech activists' enabled by data and data infrastructures. One of the best studied tech activist group is Anonymous. In 2008, the group launched 'an ambitious nuisance campaign' against the Church of Scientology, which they accused of violating freedom of speech (Gorenstein-Massa 2013, 71). 'They unleashed distributed denial of service (DDoS) and Gigaloader[215] attacks on Scientology.org,' making them temporarily unavailable (ibid., 72). Gorenstein-Massa describes the evolution of Anonymous since its creation, which, in spite of the obvious differences, shows similarities with the process some civil society organisations go through when their activities are *threatened* and under siege by the lack of freedom or resources, and realise that they also have to campaign in order to subsist (Gutiérrez 2012, 225-230). Its evolution is highly pertinent as it embodies the evolution of reactive data activism as well. Anonymous first emerged as a 'highly permissive free space… enabling conditions for recreational *pranksterism* and the formation of an oppositional community where

---

[211] See autistici.org/es/index.html [accessed on February 26, 2016].
[212] See plentyfact.net/ [accessed on February 26, 2016].
[213] See help.riseup.net/ [accessed on February 26, 2016].
[214] See indymedia.org/or/index.shtml [accessed on February 26, 2016].
[215] 'Gigaloader, an internet-based tool that is now offline, was one of many tools used by Anonymous to overwhelm Scientology servers at the early stages of the conflict. It works by loading and reloading website images, thereby taking up massive amounts of bandwidth. The legitimate use for the tool is to stress-test servers' (ibid., 72).

experimentation with an emerging repertoire of new tools and tactics takes place' (Gorenstein-Massa 2013, 63). But in a second period, 'the then purely recreational community experiences becomes aware of censorship of online content by the Church of Scientology that offends users that have internalised free expression values in the previous period' (ibid.). *Anons* (as the members of Anonymous are known) subsequently 'mobilise and engage in hacktivism,' targeting the Church of Scientology (ibid.). Prompted by negative coverage from the media and an 'influx of risk-averse newcomers into community interaction spaces,' Anonymous experiences *de-radicalisation* during a third period (ibid.). Then, the community adopts traditional protest tactics and 'grows more inclusive of broader constituencies' (ibid.). A fourth period starts with a schism and 'demobilisation of anti-Scientology efforts,' and 'Anonymous adopts a project support platform which allows for the simultaneous incubation and support of diverse projects and co-existence of traditional and hacktivist protest' (ibid.). What is in store for this group? In an article entitled 'Operation troll ISIS,' Parkin suggests, for example, that Anonymous is the origin of both anti-terrorism, vigilante hacktivism and terrorist activity (2016, 133-139). Whatever the case, it is my presumption that if there has been a *de-radicalisation* of Anonymous, it has made it more palatable and its methods more acceptable to majorities. Similarly, data activism engendered a more propositive branch –proactive data activism—, propagating within less secluded communities as a result. Anonymous is one example of 'a wider range of activism for which the internet offers a (relatively) new battleground and a laboratory for experimenting with online protest practices and alternative cyber infrastructure' (Milan and Hintz 2013, 11). These forms of cyber activism comprise 'collective action that addresses network infrastructure or exploits the infrastructure's technical and ontological features for political or social change' (ibid.). But what exactly is tech activism and what are its links with data activism? Tech activism

'has roots in the free software movement[216] but has cultivated its own ethically grounded and socially informed focus;' it has its origins in the 'movement that emerged in the 1960s as a digital counterculture' involving hackers working in Artificial Intelligence (AI) at Massachusetts Institute for Technology (MIT), and 'developed the habit of sharing source code based upon a cooperative spirit and a belief that information should be free' (Milberry 2005, 21). Although they are not exact synonyms, tech activism has been called 'internet activism,' 'online activism,' 'digital activism,' 'digital campaigning,' 'electronic advocacy' and 'cyber activism,' among other terms. For the purpose of this dissertation, I will refer merely to tech activism and cyber activism, and will use them as alternative expressions. Tech activism uses electronic communication systems to propagate messages, communicate with followers and sympathisers, lobby, generate collective identities, build communities, summon and mobilise them, coordinate and organise action (including targeted action addresses at a concrete authority or decision maker), crowdsource and crowdfund initiatives, among other purposes, using communication technologies, especially social media (i.e. Twitter, Instagram, YouTube, Facebook). Data activism could be considered a kind of expanded form of tech activism in the sense that it is activism enabled by a highly sophisticated form of technology (i.e. big data infrastructures). I say 'expanded' because, as ICTs develop, tech activism finds new ways of accessing and mobilising its followers, and increasing their numbers, enlarging and enriching it repertoires of action, as well as intensifying its impact. When Milberry described tech activism in 2005, data

[216] Free Software is not the same as open software. Although there is a lot of controversy and debate around the difference between the two, free software could be considered a social movement that believes the free access to all software as an ethical imperative, vital for the users' freedom. Meanwhile, open source is a methodology (which is advocated for as well). The vision of open source activists is to make software better (Stallman 2010, 83-88). As an open software activist told me: '(the free software movement) saw its apex in the 90s and 00s, but today is regarded as too idealistic and a lost cause. As if a hippie of the 70s comes and tells you now: "make love, not war." Its precursor is Richard Stallman, who would be regarded today as a Don Quixote if he found a Sancho Panza. Open software, instead, should be understood as the open access to the source code.'

infrastructures had not been employed yet for activism. Indeed, cyber activists and data activists have a lot in common. Data activists often use the internet as a platform for disseminating data-based information, creating awareness and mobilising. When they do that, they become part of the wider movement of cyber activism. As they rely on the internet, therefore, they conform to Vegh's definition of online activism as 'a politically motivated movement relying on the internet' (2003, 71).[217] Data activists could also be considered examples of Milan's cyber activists as they are 'part of the organised civil society,' or the realm of 'non-state and non-business actors, organised in formal (nongovernmental organisations) or informal (social movements, networked collective action) groupings and networks' (2012, 1). That is, data activists relate to cyber activism in that they often hinge on the internet and mobile technologies to share their data-based contents, and in that they are part of the organised civil society.

How is the internet used in cyber activism? Action tactics can include 'e-petitions' to be sent to a concrete government, decision-maker, public or private organisation to protest against 'a wrong' and urge action for positive policy or practice change. This is a widespread method that has been labelled *clicktivism*. According to clicktivist.org, clicktivism 'is not exclusively the support or promotion of a cause online;' it is the use digital media to support an organisation or cause in a 'quick and easy' way and to facilitate 'social change and activism' (2016). It can include a whole a range of activities, for example: organising protests, facilitating boycotts, signing petitions, crowdfunding, online parody, among others (ibid.). Many non-profits use these methods, for example, emailing petitions and asking people to pass them on. Vegh's concept of organisation/mobilisation, for example, can refer to activities taking place solely offline but

[217] Vegh classifies it in three main categories: awareness/advocacy, organisation/mobilisation and action/reaction (ibid., 72-73). Data activism could also be categorised taking these categories as a reference as well.

organised online, solely online, or a combination of both. In addition, denial-of-service attacks, the vandalising of a website and sending out *email bombs* (i.e. mass emailing) are examples of cyber activism, in this case reactive. Drawing on the concept of cyberspace as a battleground, cyber activists can offer autonomous infrastructure for secure communication and action, including hardware infrastructure (i.e. servers, wireless networks and free internet access points), software infrastructure (i.e. mailing lists and websites), and open source tools (i.e. content management systems and encryption software) (ibid.). Proactive data activists can engage in these practices and tactics as well, based on data infrastructures and with a proactive approach.

5.2 Data activism from different perspectives

In the mentioned article, Milan and I set out to observe data activism from different theoretical perspectives. In this section I offer a review of such exercise, as well as more cases and reflections on what different perspectives can say about proactive data activism. Hereafter, I provide a polyhedral description and analysis of data activism using the knowledge generated around associated realms of action, namely activism, campaigning and advocacy, journalism, alternative media, international relations and social movements. I start with a brief outlook of how data activism looks like from the angle of the Habermasian communicative action.

5.2.1 A critical look at data activism

Any data activist initiative always involves the management and transfer of knowledge. As such, it can be examined through the lens of how knowledge and action are understood by Habermas. In this sense, any data activist initiative has an empirical-analytical character, because when dealing with data analysis, it entails a scientific or technical exploration. It is hermeneutic as well, because not only any dataset integrates a political, economic and/or social framing, as discussed, but it also because any data activist project includes and is based on a political

interpretation of an historical, social or cultural context. And finally it is critical because it is related to a cause that usually interprets a social and political situation, and has to do with our common human interest in freedom and emancipation (Habermas 1971).

From the point of view of the Habermasian types of action, a data activism initiative is a full-blown communicative action, since it includes teleological or purpose-oriented elements (i.e. linking actors and facts in an objective world); norm-regulated elements, which place actions in the social world (i.e. any campaign, explicitly or implicitly, is always based on a theory of change and a political construal of the social world); and elements of the dramaturgical action, because actors leading campaigns bring along their subjective worlds, present their personas in front of an audience, and by showing something of their subjectivity, actors wish to be accepted and followed by the audience. The same can be said about the *collective persona* a campaigning organisation is interested in projecting. As a communicative action, any data activism initiative integrates the objective, social and subjective worlds.

5.2.2 Journalism studies perspective

From the journalism studies perspective, Milan and I conclude that proactive data activism acquires the values of objectivity, fairness and accuracy of good journalism (2015). The 'comment is free, but facts are sacred' *motto* emerges in this context as a statement of an ultimate value for data activism. Henceforward, I examine more in detail the similarities and disparities between data activism, and investigative, advocacy and data journalism, which are the closest branches of journalism. I start by examining the role of research –the empirical-analytical element— in both realms of action, which is something they have in common, followed by an examination of whether *objectivity* is possibly a difference between them, and a comparison

between data activism and the abovementioned types of journalism. The aim of this exercise, again, is to explore data activism from the maximum number of angles.

I have already defined activism and journalism. Henceforth, research is to be understood as the disciplined investigation into and study of materials and sources so as to discover facts and ascertain new conclusions using the scientific method. Researchers can be found mainly in research centres and universities, but also in NGOs, news media organisations and even businesses. Research can be both a tool and an objective. When I talk about research as a separate practice from journalism and activism, for example, I mean the activity that takes place mostly in research centres and universities, which has no other purpose than arriving at scientific conclusions. However, there is no journalism and, ideally, no activism without research, which then becomes an instrument. The meaning is clear from the context. And research is always based on data, which is the main common trait of data activism and data journalism. The difference between journalism and activism, when it comes to research, data and information, is one of objectives. Journalism –as Bill Kovach and Tom Rosenstiel conceived it— is supposed to aim at revealing the truth about something socially relevant, whatever it is, because the first obligation of journalists is 'to tell the truth' (2007, 23). To be exact: for journalism the research findings are the final objectives in themselves. What is the difference then with research as an activity? The difference is that journalism looks for news and stories, whereas this is not necessary so for researchers. All advocacy campaigns are preferably based on research as well, and data-activist campaigns even more so. However, information and research in advocacy are usually regarded as tools to support the final objectives of a campaign, which is always to correct or eliminate a *wrong*, or to propose or bring about a *right* in social policy or practice. Similarly to data activism, 'information activism' (or *info-activism*) is what Tactical Technology Collective

(known as Tactical Tech) defines as 'the strategic and deliberate use of information within a campaign' (2009). A key concept in this definition is *strategic and deliberate use*. In journalism, there is no other strategy (in an ideal world) than simply revealing a relevant truth and convey it well, even if that truth contradicts the initial assumptions (the same way as in the scientific method). In info-activism, instead, the information needs to be used in a strategic manner; otherwise it is of no use. That is, the information resulting from research is mainly teleological. In my view, by adding 'data' to 'activism,' this notion goes beyond *info-activism* and the tactical or strategic use of information as well, making data central. By using data analysis and visualisation –including data mining,[218] statistical analysis, machine learning and interactive graphics —, data activists may not even know what lies ahead; they may not have elaborated a postulate to start with, much less a put a strategic campaign in place beforehand. Prest, interviewed for this dissertation, mentions a project she is working on with Global Witness, 'in which activists and campaigners analyse and process a bunch of data on who owns companies in the UK to examine offshore tax havens.' The approach to this project is abductive.[219] The analysts have not formulated any clear hypotheses. If they find notable correlations and associations, 'there is a potential to shape a campaign' around them; if not, that possibility ceases to exist. Prest admits, though, that the opposite is more frequent in campaigning: first, you usually first create a campaign (based on previous research), and then find the data and analysis to support it. Likewise, Cabra tells about how the ICIJ is doing something similar throwing large datasets of unrelated data into the *machine* to see whether correlations emerge. However, this

[218] A particular data analysis method that focuses on modelling and knowledge discovery for identifying patterns and establishing relationships. It focusses on predictive, rather than descriptive, forecasting (Tech Target 2016).

[219] When talking about the scientific method, I defined deductive and inductive reasoning. Abductive reasoning 'usually starts with an incomplete set of observations and proceeds to the likeliest possible explanation for the group of observations' (Bradford 2015). 'It often entails making an educated guess after observing a phenomenon for which there is no clear explanation' (ibid.).

exercise is exploratory and not standard yet. Nevertheless, the role of research and data are fundamental for both data activism and data journalism. Seen from this perspective, the difference between both practices is one of purpose. There are different approaches to data activism as well in terms of objectives. A common element between info-activism and data activism, for example, is that both aim at removing barriers and providing access to knowledge for people. However, for me, data activism goes beyond info-activism as well in that it uses data infrastructures, maps and visualisations to communicate findings. Visualisations facilitate the communication of complex ideas. So, data activism is not only about access to information; it is about using data tools to communicate research, employing practical knowledge for decision-making, as well as inciting social change. In other words, info-activism seems more focused on providing access to and using information, while data activism is more focused on the actionable use of data-enabled knowledge. Ushahidi's projects, again, are good examples, since they geoposition data, making them available through maps that are easy to access, interpret and use.

As seen, journalism and data activism are very close. Both are supposed to be based on data and research. This is also a key issue here, since by assimilating the values of journalism, data activism obtains the objectivity ethos that dominates good journalism and research. So, what separates data activism from journalism, apart from a more teleological view on research on the part of data activism? Perhaps a key difference is that data activism, even if based on data and analysis, then takes a subjective turn of the dramaturgical nature, and becomes non-objective when it uses data and research to connect with people to mobilise them. In fact, there is still a general misgiving among journalists and researchers with regards to any kind of activism because of its supposed lack of objectivity. Alternative media –as media guided by a political position, therefore non-objective—can be examined to observe how, and whether, this *non-*

objectivity influences how journalism reports reality. Downing sees alternative media's values as a challenge that has both 'a normative' and 'an epistemological' aspect. 'Alternative news values are bound up not just in terms of what is considered as news but also in approaches to news gathering, who writes such news and how it is presented. These values present a direct challenge to the objectivity ethos that dominates professionalised journalism' (2011, 18). Mainstream journalism is based upon the 'empiricist assumption' that it is possible to identify facts that occur, 'accurately and without bias;' in contrast, 'the normative ideal of alternative journalism argues the opposite: that reporting is always bound up with values (personal, professional, institutional) and that it is therefore never possible to separate facts from values' (ibid.). Does this set of values and practices invalidate alternative journalism's truths? Far from it. Alternative journalism argues that different forms of knowledge may be generated, representing 'multiple versions of reality from those of the mass media' (ibid.). Mainstream media, nonetheless, do not always adhere to the rules imposed by an objectivity ethos, and transgressions are less and less tolerated in a digital environment where everything is immediately accessible and checked, and reactions are instantly disseminated. A paradigmatic example is the blunder that put an end to CBS anchor Dan Rather's distinguished career in 2004, when a bunch of unknown bloggers raised questions about documents he used to support a report on CBS-TV related to US President George W Bush's service during the Vietnam War (Hagan 2012). Examples of such blunders abound. Apparently 2004 was not a good year for US media organisations' credibility. In 2004 too, *The New York Times* admitted it had been less that objective when publishing information about Saddam Hussein's government based on 'a circle of Iraqi informants, defectors and exiles bent on regime change in Iraq' without, basically, confirming it (2004). The paper was referring to information that corroborated Bush's

government stand on Iraq, that is, that Hussein was behind the 9/11 terrorist attacks against the US, and that his government was hiding so called 'weapons of mass destruction.' Except none of it was based on verified facts. This type of gaffes should not revoke the power of journalism practiced by mainstream media, though. When they occur, honest organisations will admit to them, apologise, and review and improve their processes, and honest journalists will step down. However, this kind of missteps has not humbled mainstream media organisations. Alternative and citizens' media[220] have drawn criticism from traditional media institutions, which reproach alternative and citizen journalists of forsaking the traditional goal of *objectivity*. But it is a fading belief among conventional journalists that only trained, salaried and professional journalists within traditional media organisations can understand the exactitude and ethics involved in reporting news. It is true that citizen journalists may also be activists within the communities they report on. But the reality is much more complex than a simple dualistic vision that distinguishes *objective* conventional journalism from the biased reporting practiced by dilettantes who are emotionally involved in the causes they write about. At the beginning, Radio Vallekas,[221] for example, 'was more a mobilisation medium than a communication medium' (Rato 2016). But with time, Radio Vallekas became more professional and journalistic, and some of its journalists went on to work in mainstream media. Besides, there is a growing field of action within civil society organisations that specialises in fact-checking stories, when mainstream media organisations who published them did not (i.e. Accuracy in Media).[222] In the face of lack of access to information, citizen journalism has also been a source of information for mainstream media. An example is the Syrian civil war; the ban on international reporters

[220] The difference between alternative and citizen journalism will be explored in connection with data activism in a separate section.
[221] A community radio station in a workers' quarter in Madrid, which was founded in the 80s by an ecologist-pacifist group. See radiovallekas.org/spip/ [accessed on June 28, 2016].
[222] See aim.org/ [accessed on April 18, 2016].

generated a 'growing dependence on citizen journalism' (Bahja 2013).[223] Talking about the role of the ordinary citizens in news reporting, Allan introduces the concept of 'citizen witnessing' (2013, vi) as public service, in order to reconsider traditional standards lying under conventional distinctions between the assumed *amateur subjectivity* and *professional objectivity* (ibid., 10). Allan pays special attention to the actions of ordinary people who feel bound to partake in the making of news forced by events. In bearing witness, they get involved in distinctive types of journalistic activity, producing direct reporting, frequently making a key input to news coverage and reinvigorating journalism (ibid., 8). Allan's concept may be close to what Ushahidi, among others, do, facilitating and providing witness-based information in cases of crisis, provided also by ordinary people who feel forced by events to participate in the generation of information. In fact, this kind of *public service* may be considered 'citizen witnessing' as well, even if not of journalistic nature.

The claimed objectivity in mainstream journalism and research has been questioned as well, and not in relation only with a few isolated blunders. Lynd has been quoted extensively when he asked: 'knowledge for what?' (1967). This question still looms over social sciences in the big data society. Some techno-utopians pretend a state of asepsis exists, or at least it is possible now, in the age of big data. After all, when a scientist starts researching with a cause, he or she has had to develop a theory of change first, responding to Lynd's question, 'for what;' therefore, he or she begins marching into the realm of advocacy, and any pretentions of objectivity disappear with it. At present, this concern is very much alive in relation with big data

[223] However, Bahja argues also that, in the case of the Syrian war, 'the reliance on citizen journalism increasingly threatens the credibility of the mainstream media. While the use of a disclaimer to indicate unverifiable information may to some extent protect credibility, it could also be argued that its overuse makes reputable news organisations more inclined to publish information without going through the necessary verification process' (ibid., 9).

infrastructures, and some critical voices have emerged in the field of journalism to denounce a sort of pernicious obsession with technology (Kaplan 2013), when the real question is *for what*. Conversely, it has been established earlier in this dissertation that data are never objective or *raw* (Gitelman 2013), that they emerge from a set of values and biases, that the can contain errors, gaps and preconceptions. As a researcher and an activist, I subscribe to Lynd's challenging query and Kaplan's suspicion of tools for tools' sake. Today's complex, post-modern problems (i.e. poverty, climate change) require first-class thinking, research, collaboration and instruments, with clear purposes. That is why DataKind, among other organisations, strives to transfer data science into social organisations fighting for a cause. DataKind places volunteering data scientists in social organisations to work together in research and social projects (2015), providing these scientists with a 'for what.' My stand is that neither research nor journalism of the most serious kind are, or can be, ever totally objective. Their methods can be objective instead. Their aim should include attaining objectivity as well, even if it is never to be achieved. As Downing says of alternative journalism, different forms of knowledge may be generated by analysing datasets, representing 'multiple versions of reality' (2011, 18). In my view, no version of reality –whether obtained using the journalistic method or the scientific method— is monolithic and absolute, especially when it refers to the complexity of post-modernism. However, one of the differences between research and journalism is that you can say more (valid) things based on anecdotal evidence as a journalist, whereas in science you need to examine relevant samples revealing the whole picture, and can sustain scientific truths only based on scientific evidence and other scientifically accepted truths. The standards of verification in science are, therefore, higher than in journalism, which does not mean that good journalism is

less true. As said, different forms of knowledge may be generated by both, although neither the scientific nor the journalistic method makes them *objective*.

Fortunately, some voices have emerged in favour of being more public about this thorny issue. Gillmor, among others, is asking for a more open discussion about objectivity in journalism. He even proclaims: 'Maybe it's time to say a fond farewell to an old canon of journalism: objectivity' (2011). In a similar vein, Media Ethics calls for a 're-evaluation of objectivity and advocacy as ethical values in journalism,' and highlights the fact that some mainstream media that are supposed to practice objective journalism are, nonetheless, perceived as biased (2013, 10). Quoting a survey, Media Ethics says that 'some have asserted that a lack of objectivity has been responsible, at least in part, for journalism's reputational decline. Others declare that objective journalism has grown stale and that advocacy journalism offers an opportunity to make it fresh and relevant in a society overloaded with information' (ibid.). Media Ethics concludes that 'news organisations need to understand what members of the public mean when they express a desire for the press to be objective, the extent to which the public is aware of the advocacy movement within journalism, and their reaction to the virtues of both objectivity and advocacy' (ibid.). Maybe, what the *public*[224] calls objectivity is really honesty, transparency, fairness and a sense of proportion when dealing with several versions of a fact. Activism and advocacy, on the other hand, are never objective (although they include objective and empirical-analytical elements). And they do not try to appear objective. Kovach and Rosenstiel do not talk about objectivity as a goal in journalism either, but about seeking the truth and being independent, exposing the facts confirmed by the method of verification. 'The concept of objectivity has been so mangled it now is usually used to describe the very problem it was

[224] I use this word in the context of Media Ethics' musings. However, as discussed, *the public* or *the audience* no longer exist in the world of mass self-communication.

conceived to correct' (2007, 26). These authors say that 'the conventional explanations by journalists about how they get at the truth tend to be quick responses drawn from interviews or speeches, or worse, from marketing slogans, and often rely on crude metaphors' (ibid.).

> (When asked, journalists say the press is a) mirror of society… or a reflection of the passions of the day… These explanations make journalists seem passive, mere recorders of events rather than selectors or editors. It's as if they think truth is something that rises up by itself like baking bread. Rather than defend our techniques and methods for finding truth, journalists have tended to deny they exist… (This is) one reason why the discussion of objectivity has become such a trap… Originally it was not the journalist who was imagined to be objective. It was his(/her) method. Today, however, in part because journalists have failed to articulate what they are doing, our contemporary understanding of this idea is mostly a muddle (ibid.).

Should journalism relinquish objectivity, then? Gillmor also says: 'It will never be time to kiss off the values and principles that undergird the idea (of objectivity in journalism)' (2011). The principles that withstand objectivity in journalism should not be relinquished. Objectivity is an aspiration that, in order to produce good journalism, has to be combined with the journalistic principles articulated by Kovach and Ronsenstiel, and more recently by Gillmor, who talks more specifically about citizen journalism. Objectivity is in the journalistic method of verification. In citizen journalism, says Gillmor, it is thoroughness, accuracy, fairness and transparency (ibid.). Objectivity could be also in *disclosure* of issues such as financial conflicts of interest and personal biases in journalists and media organisations, and making source material available so the audience can make their own mind, as in the open movement (ibid.). Similarly, talking about the use of data infrastructures in the study of history, Schmidt sees a great opportunity for

historians in creating a new 'source criticism' that explains what is behind the datasets (2012). In brief, journalism is never totally objective, the same way data are never *raw*. I totally agree with Gillmor's view: let us stop pretending journalism can be objective and be transparent about it. Total objectivity in journalism –or in research— does not exist. However, the fact that data and data infrastructures have a central, enabling role in data activism, along with their channelling function as communication tools, make data activism closer to journalism and its values. But what sort of journalism?

Once established the relation between data, research, data activism, journalism and objectivity, I examine the three types of journalism that have more in common with data activism: investigative, advocacy and citizen journalism (Milan and Gutiérrez 2015, 129). As noted before, investigative journalism –an old form of journalism— has turned to data and data infrastructures of late. Data and data infrastructures, therefore, are the point of juncture between both. Advocacy journalism intentionally and openly embraces a political perspective in reporting. It can be fact-based, but it explicitly takes sides (Careless 2000). However, 'being an advocate journalist is not the same as being a full-blown activist. No matter how dear a cause is to journalist's heart, there are lines which should never be crossed by a professional journalist' (ibid.). That is, in the convergence of journalism and advocacy, the former prevails. Advocate journalists should follow the same principles as any other journalist. An example could be *The Guardian's* coverage of climate change issues, mentioned earlier. This British paper does not hide the fact that it is campaigning towards global and substantive greenhouse gas emissions cuts, and *uses* journalism, and more often than not data journalism, in order to support its political stand to foster social and political change. Here the point of intersection is having a cause and a theory of change: both advocacy journalism and data activism always have one.

Finally, citizen journalism and proactive data activism coincide mainly in that they are usually bottom-up, grassroots, participative initiatives. Internet has revolutionised journalism practices (as well as activism). People formerly known as the *audience* or the *public* now generate and disseminate their own contents. This phenomenon has 'busted open the system of gates and gatekeepers' that previously regulated access to information (Rosen 2006a), a model that was based on scarcity of information, the lack of access to it and the shortage of space to communicate it. In the process of 'gatekeeping' that existed until the start of the post-modern society, journalists were the sole custodians of what was debated and known in the public sphere (Park 1922, 328). Intelligence was scarce, and citizens depended on the journalists' ability to transmit relevant information and analysis so they could make sensible decisions and appoint governments. But this all changed when the internet opened the gates and citizen journalism emerged. Is what citizen journalists do journalism? Of course it is. As long as this type of journalism follows the guiding principles of journalism, whether journalists who produce journalistic content are salaried or not, employed or not, certified by universities or not, citizen or not is not of relevance when it comes to the value of the content. This is even less and less relevant in the current environment in mature economies, where purported professional journalists are being fired by the thousands. However, citizen journalism 'deliberately positions itself in the realm of subaltern counterpublics rather than seeking integration with the mainstream' (Milan and Gutiérrez 2015, 129). Another point of contact between citizen journalism and data activism is collaboration. 'Data journalism calls for the development of specific skills in programming and advanced statistical methods that are not typically part of the skillset of old-school journalists. Thus, the journalists' engagement with data prompts unprecedented alliances, which are of paramount relevance for observers of data activism' (ibid.,

130). In summary, proactive data activism is similar to three varieties of journalism: investigative journalism (because of its analytic nature), advocacy journalism (due to its political stand) and citizen journalism (because of its collaborative, rank and file character).

5.2.3 Alternative media studies perspective

Milan and I used the alternative media studies perspective as well in order to look at the empowerment capacities of data infrastructures in activism (2015, 131). We conclude that there are 'evident links' between data activism and alternative media, which empower and 'bring citizens at the forefront of media production' (ibid., 128). But what are alternative media to start by? We note that, over the years, media scholars have proposed different labels to describe non-commercial grassroots media: Rodríguez prefers to call them 'citizens' media' (2009), and Downing, 'radical media' (2001); Hackett and Carroll refer to their 'oppositional communication practices' (2006, 14), focussed on 'lifeworld change' (ibid., 56) and looking for alternative public spheres, 'bypassing mass media gatekeepers to communicate directly with the broader public' (ibid., 47). Thus, alternative media are manifestations of what Fraser would call *subaltern counterpublics*, an expression of marginalised groups' forming their own post-bourgeois public spheres, as a challenge to hegemonic domination and views (1990, 58-61). This circumvention is a direct antagonist of the *status quo,* since alternative public spheres do not only constitute alternative channels for content, but also alternative channels for *alternative ideas and value systems*. Fraser suggests that alternative media defy the dominant hegemonic thinking (1990, 61). This could be considered a rebellion against mass media as vehicles of the 'ruling class ideas' –borrowing this concept from Marx and Engels (2000, 21)—, and the hegemonic value systems (Chomsky 1997). Chomsky discriminates between 'elite media' or 'agenda-setting media,' and mass media (1997, 1-2). The elite media 'set the framework in which everyone else operates,' and include media organisations such as *The New York Times* and CBS. 'Their

audience is mostly privileged people. The people … are actually involved in the political system in an ongoing fashion. They are basically managers of one sort or another… or other journalists who are involved in organising the way people think and look at things' (ibid.). The 'real mass media,' instead, are 'trying to divert people' and 'get interested in professional sports, for example… Let everybody be crazed about professional sports or sex scandals or the personalities and their problems or something like that. Anything, as long as it isn't serious' (ibid., 2). The goal that lies beneath is that the economic elites' value system is not questioned and perpetuates itself. Chomsky goes on to argue that other institutions, such as universities, support this value system as well. They are not independent, although independent elements may populate them, the same way there may be independent elements in a Fascist state (ibid., 2-3). For Downing, 'alternative media are produced outside mainstream media institutions and networks' (2011, 15), and the term is the most 'all-encompassing definition' (2004), which can include from media of protest groups, dissidents and fringe political organisations to fans and hobbyists (ibid.) They tend to be produced by amateurs 'who typically have little or no training or professional qualifications,' and they seek to redress 'what their producers consider an imbalance of media power in mainstream media, which results in the marginalisation (at worst, the demonization) of certain social and cultural groups and movements' (ibid.). In contrast, Rodríguez proposes to abandon 'alternative media' for 'citizens' media,' because the first concept 'rests on the assumption that these media are alternative to something' and highlights 'the type of oppositional thinking that limits the potential of these media to their ability to resist the alienating power of the mainstream media' (2001, 20). Whereas the second concept implies that 'a collectivity is enacting its citizenship by actively intervening and transforming the established mediascape;' that these media 'are contesting social codes, legitimised identities and

institutionalised social relations;' and that their communication practices ' are empowering the community involved, to the point where these transformations and changes are possible' (ibid.). I would distinguish between *alternative* and *citizens'* media in that the term 'alternative media' puts the accent in either new ways of communication with audiences, or the new perspectives brought to reporting and the unconventional and unorthodox ideas these media funnel, or both. Meanwhile, 'citizens' media' puts the accent in the collective, participative, grassroots, non-hierarchical ways of producing information and analysis, which can be, and often is, alternative in nature as well, but not necessarily. In citizens' media, citizens, either as individuals or as a collective, act upon a situation and become content producers, alternative or conventional. Seen in this light, *Al Jazeera* –at least at the beginning— and *Inter Press Service* (*IPS*)[225] could be considered alternative, but not citizens' media organisations, since their content is produced in a hierarchical structure of specialised roles –reporter, correspondent, anchor, contributing writer/commentator, editor, fact-checker, etc.— by salaried journalists. On the other hand, Radio Vallekas —where community citizen journalists report on locally relevant news— and Antonio French's Twitter account –with pictures and videos of police actions captured on mobile phone (Gillmor 2014)– are examples of citizen journalism producing alternative information you cannot find anywhere else. It is noteworthy that some citizen journalists that learnt the ropes at Radio Vallekas went on to become mainstream journalists in commercial channels (Rato 2016).

[225] Inter Press Service was founded in Rome in the 60s by a group of South American dissident journalists who thought alternative information could be a weapon against the authoritarian dictatorships of the time. Its structure is an association and it is funded by grants, donations and subscriptions. It evolved from a confrontational media organisation to a professional news agency dedicated to covering development and human rights issues with a from-the-ground perspective. I was its editor in chief from the beginning of 2003 until the end of 2009, period during which it experienced a great online expansion. Some of the unique characteristics of IPS are: each story should include at least three exclusive and relevant sources; and the perspective of all stories should be that of the people affected by the issue, trend or event that is being analysed. Rodríguez considers IPS as a 'Third World press agency' informing about their own societies 'from their own point of view' (Rodríguez 2001, 6).

That was only possible because, as said, processes and content production can be comparable in alternative/citizen and conventional journalism. The perspectives and issues covered, as well as the business models that enable them, make this type of journalism alternative or citizen. Interestingly, Rodríguez also reveals a departure from the notion that the media landscape is inhabited either by the powerful (mainstream media) and the powerless (alternative media), (2004), in a way that alters 'historical power configurations' (2001, 64). The *mediascape* is, in fact, much more complex and fertile than that this bipolar picture, as shows the example of Radio Vallekas, a school of radio journalism for many who would later become salaried journalists in mainstream media organisations.

Rosen defines citizens' media implicitly as media that are vehicles for content generated by 'people formerly known as the audience' (2006a; 2006b). This is a key concept, as it is the nexus between citizens' media and data activism: in both the control and the content generation is in citizens' hands, and that very fact transforms them. Rodríguez, precisely, theorises about alternative media in terms of the transformative processes they bring about (2001, 20). This transformation occurs also in data activism, which not only empowers individuals and organisations; it also 'alters the power dynamics within the transnational civil society' (Milan and Gutiérrez 2015, 133). 'What makes the notion of citizen's media so appropriate for the subject matter (data activism) is its focus on empowerment and the politics of daily life. Within the communication sphere, empowerment is the process through which individuals and groups, by taking active part in the actions that reshape their communicative processes, exercise control over their communication resources' (ibid., 128).

To summarise, data activism and alternative and citizen's media have many points of contact. Similar to citizens' media, 'proactive data activism involves a politics of the quotidian,

as it alters the ordinary relationship between citizens and automatized data collection. As such, it brings back into the data collection machine the fundamental elements of agency and politics' (Milan and Gutiérrez 2015, 134). Citizens, whether individually or collectively, are at the forefront of production both in citizens' media and in data activism (ibid., 128), and that this fact constitutes a powerful transformative and empowering process (Rodríguez 2001, 2004).

5.2.4 International relations studies perspective

Most data activists deal with international issues (i.e. deforestation in the case of InfoAmazonia) and even global issues (i.e. climate change in the case of *The Guardian* campaign or illegal fishing in the case of the ODI/porCausa project), and therefore operate internationally and in some cases, globally. Big data infrastructures are especially suitable to study global trends, since they offer the possibility to zoom in and out on them in order to get the most complete picture possible both of the reality as a whole as well as of details of microscopic dimensions. Big data analysis and visualisations are also perfect tools to communicate campaigning messages at a global level. Maps and images are universally understood; numbers are a universal language too. In this section, I examine some of the facets of *globality* in relation with big data first, and how data activism could be seen by international relations studies, second.

In modern times, the sphere of action of any given state was restricted by territory, and the articulation of state power was maintained by the claim of the monopoly of violence within a given territory (Castells 2009, 15). Nation, state and territory demarcated the limits of society until the arrival of globalisation and liquid post-modernity, when the territorial boundaries of the exercise of power were redefined (ibid., 17). The result is that the big data society lives in a globalised world where contact occurs in the space of flows, and in the space of flows, national

authority gest diluted. But what does this mean? Habermas acknowledges the dilemmas emerging by the coming of what he calls 'the post-national constellation' for the course of democratic legitimacy; constitutions are national, while the sources of power –as well as the challenges humanity faces— are progressively based at a supranational level (2001, 61). The result is pervading perception that national politics have shrunken to the management of 'a process of forced adaptation to the pressure to shore up purely local positional advantages,' which 'deprives political controversies of their last bit of substance' (ibid.); or to the communication of messages (Barandiaran 2015). This is possibly one of the factors determining the dwindling numbers of apathetic voters. However, national and local politics deal with global threats. An example of a global challenge that demands global, and local, answers is climate change. The problems that climate change generates are being felt locally [i.e. sea level upsurge is not homogenous across the world (Surging Seas 2016)]. 'We're already experiencing extreme weather that could become the new normal, and villages like Kivalina, in Alaska, will soon be lost to the sea. The relative "nimbleness" of small communities has allowed them to experiment with climate action' (Reddy 2015). Most developing countries have not contributed to creating the problem, but they are more vulnerable to and are less prepared for climate change impacts than rich countries –big emitters of greenhouse gasses (GHGs). Many are already experiencing the effects of global climate change very locally. Another example is Bangladesh, whose efforts have to be directed to adapt to the irreversible changes generated by climate change. Among other impacts, Bangladesh is at risk from rising sea levels and salinization of fresh water, as pumping water causes the land to sink and, when sea level rises, the risk of flooding increases. The country's climate scientists conclude that 'by 2050, rising sea levels will inundate some 17% of the land and displace about 18 million people' (Harris 2014, 5). Bauman points out that 'one

of the most bewildering paradoxes revealed in our time' is that, in spite of a fast globalising process taking hold of our planet, politics is mostly local (ibid., 82). I would add that some of these challenges are long term issues that can be counted in decades, but the horizon of politicians generally expands just four of five years. Local politics –often urban politics— has to deal with global challenges, an impossible equation. Cities everywhere are facing new weather patterns, making adaptation strategies at city-level ever more important. Adaption challenges in Quito, for example, are demanding: With a population of 2.24 million (expected to double by 2025) and located at 2,800 meters above sea-level, Quito's streets are steep and disrupted by ravines. Regular floods, earthquakes and landslides produce widespread damage, mainly in informal settlements on hillsides (Gutiérrez 2015). This is because, in a globalised world, most of human experience is local (Borja and Castells 1997, 44-47). The classical theory of power, focused on the nation-state government structures, 'lacks a frame of reference from the moment that key components of the social structure are local and global at the same time rather than local or national' (Castells 2009, 17).

This is not to say that the nation-state disappears, but it becomes 'just one of the dimensions in which power and counter power operate' (ibid., 18). Ultimately, this affects the nation-state itself. Even if it does not fade away as a specific form of social organisation, it changes its role, its structure, and its functions, gradually evolving toward a new form of state, a 'network state' (ibid.). Due to the communication and data infrastructures, this is a globalised society made more global, as well as more *tribal* (McLuhan 1964, 380). Although a lot has been said about the disappearance of the nation-state in this global context, there are human rights spaces which are still dependent on national policies and decisions too, where governments nonetheless retain a measure of power (i.e. national laws on migration in European countries

versus a common European policy). And the nation-state is very much alive when it comes to the internet regulation. In fact, state regulation and control of the internet are becoming prevalent both in authoritarian regimes and in democratic ones as well (Kalathil and Boas 2003, 173). However, the controls are not being created only at a national level. 'In contrast to its anarchic beginnings, the internet is increasingly being regulated and overseen by governmental, intergovernmental and nongovernmental organisations. National governments are pressing for a greater say in determining acceptable content within their borders (as in the French case against Yahoo)'[226] (ibid.). National data access –mostly via internet— is also regulated by national policies. This issue is even more apparent in authoritarian regimes, which are in favour of 'strongly delineated national borders online' (ibid.).

In this globalised society, the possibility of competing internationally is controlled by a privileged few individuals, companies and countries that have access to knowledge, data and technology –the supreme values in the Information Age. Bauman calls this a 'negative globalisation; that is, a selective globalisation of trade and capital, surveillance and information, violence and weapons, crime and terrorism, all unanimous in their disdain of the principle of territorial sovereignty and the lack of respect for any state boundary' (2007, 7).[227] According to Bauman, the ultimate cause of negative globalisation is due to the fact that the owners of 'Capital' are 'invisible and shifting,' having the power to invest locally without making

[226] Ligue contre le racisme et l'antisémitisme et Union des étudiants juifs de France c. Yahoo! Inc. et Société Yahoo! France (known as LICRA *v.* Yahoo!) is a court case decided by the Paris High Court in 2000. The case concerned the sale of memorabilia from the Nazi period by internet auction.

[227] This brings to mind the oceans again. Illegal, unreported and unregulated (IUU) fishing is a type of environmental crime that does not respect territorial boundaries, and that not only creates catastrophic and often irreversible damage that compromises a local community's ability to survive, but also growing evidence indicates that it undermines governance, perpetuating weakness in institutions; funnels funds away from areas like health, education and food security; fosters corruption, money laundering and organised crime; generates tax avoidance; displaces communities; and endangers species and depletes resources. All of which have direct and indirect impacts on development (Daniels et al. 2016, 6-9).

commitments. The global elite are globally free to move on to lower tax systems elsewhere.[228] Meanwhile, the experience of 'negative globalisation' for the rest of us who are 'doomed to be local' is one of increasing anxiety, which derive from living in an unpredictable world over which we have no control (ibid., 5-27). Globalisation forces all countries, whether industrialised or not, 'to press their advantages and thereby make the best use of their human resources, their skills, and to raise their level of production' (Touraine 2007, 65). Touraine talks about the 'decomposition of society,' which happens 'when the link between system and actor is broken, when the meaning of a norm for the system no longer corresponds to the meaning it has for the actor' (ibid., 63).

This means that transnational organisations dealing with global issues must campaign at a global level, and deal with the post-modern panopticon as well to protect activists and witnesses from prying eyes and the dangers of repression. This means also that big data analysis, with their capacity for both an all-inclusive and an infinitesimal grasp is especially adept to generate the kind of research needed to campaign at that level. This is the scenario where *transnational* advocacy networks (known as TANs) operate and where they implement their international campaigns to influence, and ultimately change, global normative systems. Ultimately, *transnational* advocacy networks aim at subverting the established power relations and norms at a global level. Fraser says that TANs aim to make a difference to international regulation or practice that affects political, social and economic conditions within and across states on behalf of those who are affected by global governance and trends. If they are successful, they may increase the porousness of international political institutions to less influential and non-state

[228] Investigating money laundering from Panama in the 90s, I often learnt that when a Caribbean island cracked down on money laundering, another one was opened for business, bringing to mind the high degree of flexibility and resilience that (dodgy) international finances exhibit.

actors and the possibility of 'translating' dissatisfaction with those institutions upwards, but generally this is a kind of side-effect (2014, 72-73). Another key element is that the ideals that surround data activism generally run counter to individualism and consumerism, and are close to the communitarianism described in previous chapters. Data activists normally work collectively, and favour open source software and open data, for a common good. 'Advocates of principled causes, ideas and values (like human rights or the environment) do not work alone. Nor are they limited by national boundaries. The last several decades have witnessed significant growth in the number of loose coalitions or networks of advocates building bridges across borders to bring about social change' (Keck and Sikkink 1998). Data and collaboration are then perfect allies in the face of global challenges.

International relations studies offer 'hints on how to integrate transnational power dynamics into the analysis of a phenomenon that is local and transnational at the same time' (Milan and Gutiérrez 2015, 134). By means of the space of flows (Castells 1992, 126-172), the emerging data activism networks 'take the form of trans-border de-localised communities, while operating as transnational advocacy networks' (Milan and Gutiérrez 2015, 134). Similar to other instances of internet activism, these 'new forms of networked action and informal collaboration are challenging traditional notions of civil society' (Milan and Hintz 2013, 14), and are different from traditionally organised collective action. This perspective is very useful, since some of the data-based campaigns that examined here are shaped in the configuration of an international organisation and are, therefore, observable from an international relations perspective. But what are TANs exactly? *Transnational* refers to advocacy efforts that include at least two countries (Bandy et al. 2004, 231). Keck and Sikkink define TANs as networks of social actors across countries characterised by sharing values, who create, exchange and utilise information

strategically. They seek to influence policy, but also to 'transform the terms and nature of the debate' (1998, 90). Meanwhile, the *network* concept 'stresses the fluid relations among committed and knowledgeable actors working in specialised issue areas' (ibid., 91). And *advocacy* encapsulates what is distinctive in these transnational networks: they are work support of principled causes (ibid.). TANs comprise an array of groups, sometimes beyond the reach of concrete national governments. One example that Milan and I identify is 'The Price of Oil'[229] (ibid.), a campaign lead by Oil Change International, of Washington. It seeks to expose 'the true costs of fossil fuels exposing the true costs of fossil fuels and facilitating the coming transition towards clean energy' everywhere (2016b). This transnational campaign takes the shape of a TAN, and has data and research as its core, since it analyses large datasets of information on fossil fuel subsidies in several rich countries and forms the campaign messages around the findings of data analysis. In this TAN, other organisations cooperate, including ODI, of London, and the Global Subsidies Initiative (GSI), of Geneva. Each of these organisations contribute something different to the campaign: OCI and GSI offer their specialised research and knowledge of the oil sector; ODI contributes its experience of linking all sort of data (i.e. trade, climate, economic, consumer data) to social and development issues, as well as its communications machinery and network of international contacts. This alliance is described as follows:[230] Through ODI's existing relationships with key international organisations, the proposed action builds on and links to existing work on FFS (fossil fuel subsidies) by the OECD, World Bank, IMF (The International Monetary Fund), IEA (International Energy Agency), GSI and OCI. In addition, local partnerships are established at country level with

[229] See http://priceofoil.org. This is a campaign in which, as responsible for communications at ODI's Climate and Environment Programme, I have worked with.
[230] This is based on the Terms of Reference that the three organisations agreed to forge the alliance (Overseas Development Institute 2014).

research institutions, advocacy groups and media organisations. The coordinated action links the often separate processes of research, advocacy, campaigning, media outreach and independent journalism with the goal of developing critical awareness and active engagement of European citizens. This approach develops bespoke research and analysis on FFS in order to unlock information to producers and consumers, and enables effective outreach to the public and key stakeholders with the aim of shifting attitudes, behaviours and policies. By joining up these different processes of research, campaigning and media, a combined capacity is generated to make a difference in multiple constituencies, spaces and processes (Overseas Development Institute 2014). The findings of this research indicate that the combined annual spending on FFS production in the G20 governments is US$444 billion; in comparison, G20 support for the Green Climate Fund –an international binding commitment designed to help poor countries fight against climate change and adapt to its irreversible consequences— was US$7.5 billion in 2014 (Doyle 2014), including up to US$3billion by US President Barack Obama (which is now in question by the current President Donald Trump). The enormity of the difference is the basis for the campaign messages. From my privileged access to how this alliance was forged, I consider it a successful example of a TAN. Shaping the alliance was not easy, and many misgivings had to be overcome among its members during the process. For example, ODI was a newcomer to FFS issues, but considerable bigger than the other partners in terms of size and influence. Instead of focussing on weakness and what separated each of the other consortium partners, the efforts were unremittingly directed to focus on the common goals and the strengths each of the partners contributed to the consortium, and remove reasons for scepticism. This was a task that took many months. Since this campaign was launched in 2013, there have been policy changes (although, as it often happens with campaigns, it is difficult to

attribute the results only to the TAN's work). 'While governments have been slow to act on this point, there is a growing recognition among major world powers that subsidies are in need of a radical rethink: according to the IEA and IMF, almost thirty countries introduced reforms in 2013 and 2014, with many more in the pipeline' (Timms 2016). The pressure on OECD countries to change their FFS polices after the climate summit in Paris of 2015 is even greater now. Another example of a TAN practicing proactive data activism is the illegal, unreported and unregulated (IUU) fishing big data–based campaign put together by ODI and porCausa –a civil society organisation that practices investigative journalism—, supported by FishSpektrum, the provider of the massive data, and CARTO, visualising the data. Data analysis and visualisations, research, lobbying and social and conventional media outreach are some of the elements of this campaign.

The Greiner's Growth Model, which shows how an organisation develops over time, is interesting to explain how one sole organisation alone, even if it fairs well and survives all the growth crises that are likely to hit it, cannot tackle alone a complex challenge such as climate change. The Greiner's Growth Model places 'growth through alliances' at the last stage of development in any organisation, which is then followed by a final growth crisis[231] (Exponential Training 2016). That is, the stage where organisations are able to work in alliance with others, not just cooperate with them, is the most mature stage of the organisational life circle of any

[231] The other stages are: growth through creativity (characterised by not many staff, informal communication channels and long hours); growth through direction (more formal structures and communication channels, and a demand for autonomy in top managers), followed by a leadership crisis; growth through delegation (top management allowing managers to make decisions, increasing decentralisation), followed by an autonomy crisis; growth through coordination (pressure to fall in line with central management strategies and maximise synergies), followed by control crisis; growth through collaboration (this phase emphasises spontaneity in management actions, cross functional team-working, increased collaborative working and co-operation), followed a 'red tape' crisis. Once survived the 'red tape' crisis, the organisation grows through alliances (Exponential Training 2016).

organisation. Organisations of the civil society have always cooperated, since the problems they confront are always bigger than themselves, although there are different levels of cooperation and cohesion. One of the main strengths of such alliances is precisely their collective reach and influence. A loose alliance or platform for cooperation on a given struggle does not require a formal central structure coordinating the alliance, and can include *ad hoc* coalitions, strategic cooperation or concrete campaign groupings. Einarsson calls them 'umbrella associations' (2009). These could include TANs. One example of this type of alliance is the cooperation between Greenpeace, Amnesty International (AI), Medicos sin Fronteras (MSF) and Intermon Oxfam in the Spanish campaign 'Armas bajo control' (Weapons under control) in favour of the control of arms and against their export from Spain to countries at war.[232] This campaign had a data element, as its messages emerge from the investigation of weapons exports from Spain into other nations. With a higher degree of centralisation, some campaigns embrace different types of organisations across countries and this collaboration usually needs a secretariat and a minimal centralised structure, with a legal status, in order to coordinate action. These can include informal networks coordinating action, exchanging information and learning from each other, and they have more formal structures around thematic issues or a concrete geography. They could include Global Action Networks (GANs), such as the Global Call to Action Against Poverty (GCAP)[233] and the Global Reporting Initiative (GRI) (2016)[234]. Meanwhile, the highest

[232] This campaign started in 2004 and it is still active. It does not require any formal secretariat, and it is coordinated by *ad hoc* meetings.

[233] The Global Call to Action against Poverty is a worldwide alliance consisting of national campaigns focused on fighting against poverty. In 2005, it involved some 38 million people in different actions in over 75 countries, and in 2006, 23 million people in over 85 countries. It defines itself as the 'largest civil society movement calling for an end to poverty and inequality' (GCAP 2016).

[234] As of 2015, 7,500 organisations used GRI Guidelines for the sustainability reports. This is an 'international, independent organisation that helps businesses, governments and other organisations understand and communicate the impact of business on critical sustainability issues such as climate change, human rights, corruption and many others' (Global Reporting Initiative 2016).

level of cohesion is in transnational or global organisations that agglutinate national chapters within the same brand name, with a varied degree of political, financial and action independence, depending on their resources, philosophy and financial contribution to the central structure. This last group includes three types of configurations: confederations, federations and organisations with national delegations. Confederations have a common brand name and a central headquarter receiving a small percentage of their national incomes (devoted to sustain the central structure). These comprise MSF and Oxfam International, for example, whose national organisations have a high degree of independence. Federations, with a common brand name and a central headquarters, group national chapters that contribute a bit more of their income to the international or global structure. Amnesty International is an example of this second type.[235] International organisations with national delegations, such as Greenpeace, usually absorb more money from their national chapters, which also have the lowest degree of independence in relation to their headquarters.[236]

The globalisation process has created challenges and opportunities for social action, and the growth in global problems, as well as their interconnection and complexity, has required a growth in global alliances with the civil society as well. 'The expansion of international institutions, international regimes, and the transfer of the resources of local and national actors to the international stage, producing threats, opportunities and resources for international NGOs, transnational social movements and, indirectly, grassroots social movements' (della Porta and Tarrow 2004, 235). Waddell –who coined the term Global Action Networks (or GANs)— and

[235] Amnesty International does not define itself as an organisation, but as a global movement. At the time of writing, AI is undergoing a descentralisation process that is transferring research capacity from London into regional and national offices all over the world.

[236] Elaboration by the author based on personal observation as Greenpeace Spain Executive Director and member of several international alliances, and also on Business Dictionary (2016), Basic Law (2003), Einarsson (2009; 2012).

Khagram believe that the complexity inherent in the creation and management of public goods requires networks that provide the capacity to act collectively to create a global future together (2007). Global pandemics like AIDS and bird flu, transnational crime, pollution of the seas and climate change are 'only a few examples of challenges that indicate new approaches to global action, policy and governance are needed' (ibid., 261). These authors define Global Action Networks as entities that mobilise multi-stakeholder resources, bridge divides among people, and promote the long-term change and innovation work needed to address complex global challenges. And they describe them as having these characteristics: they are active globally and on multiple levels; embrace diversity and engage in boundary spanning; are inter-organisational networks; are systemic change agents; are entrepreneurial action learners; engage in action, but also need to learn the lessons of their errors; create coalitions of the willing, but they can exert great social pressure; and finally, are producers of global public goods (ibid., 263-265). The origins of GANs can be traced to the Red Cross Red Crescent Movement (1963) and the International Labour Organisation (1919), while more recent examples for such networks include the Forest Stewardship Council and the Fair Labour Association (ibid., 261). The difference between TANs and GANs seems to be in geographic scope only, although, as seen, their description includes specific characteristics. Nevertheless, hereafter, I will use one or the other depending on their geographic reach.

Recapitulating, collaboration seem to be in any data project's DNA, both at a micro level (people cooperating to generate data, analysis and mobilisation) and at a macro level (organisations working together across borders to marshal different sets of skills and influence international or global trends, processes and policies). While, as seen, alternative media studies offers a glimpse of the first ambit, the international relations perspective offers an insight of the

second one. Based on the Greiner's Growth Model, their collaboration and alliance is also a sign of maturity, since organisations seem to be ready to work together only when they have outlasted a series of crises first. Proactive data activist initiatives seem to thrive in partnerships and alliances of organisations with different levels of cohesion, which are transnational in nature and objectives, as seen for example in the case of the subsidies and IU fishing campaigns.

5.2.5 Social movement studies perspective

The study of social movements and their technologies can help to understand data activism as 'a new form of grassroots engagement with technology, one that entails direct action, cultural forms of resistance and coding' (Milan and Gutiérrez 2015, 134). This is another perspective that is useful when observing a data-based campaigning organisation.

Defining what a social movement can be tricky. It seems easier to define them based on what they are nor. A social movement is neither a political party nor a campaigning organisation; neither an association nor an interest group, which usually are stable political entities; neither a mass fad nor a trend, which are unorganised, fleeting and without goals (Christiansen 2009, 2). They are somewhere in between. Social movements 'are involved in conflictual relations with clearly identified opponents; are linked by dense informal networks; share a distinct collective identity' (della Porta and Diani 2006, 20). Van de Donk et al. say that, to speak of a social movement, 'generally four elements should be present: (1) a network of organisations, (2) on the basis of a shared collective identity, (3) mobilising people to join mostly unconventional actions, (4) to obtain social or political goals' (2004, 88). For Shirky, social movements include many groupings of diffuse –'previously uncoordinated'– actors with common interests (2008, 163). Social movements generate 'raw materials' for future movements as well, such as songs and slogans (i.e. protesters during the 2011 movement Occupy Wall Street in New York signing

'*Ciao, bella*');[237] create new tactics and 'political know-how' that future protestors can utilise (i.e. Ghadian hunger strikes and other non-violent resistance practices); leave behind 'social ties that that can be used to ignite new efforts in the future;' generate 'a moral voice' that people can then use to articulate values that they do not have time to think about in their daily lives (Goodwin and Jasper 2014, 381). For the purpose of this dissertation, hereafter, social movements can be understood as organised, yet informal and horizontal, social entities that are engaged in disruption and conflict (della Porta and Diani 2006, 20; van de Donk et al. 2004, 88), which is oriented towards social change –whether specific and narrow policy or practice or more broadly aimed at change—, and usually networks smaller organisations of diverse nature (ibid.). Their coordination and mobilisation is often based on ICTs (i.e. social networking via smartphones), (Milan 2013), and enabled by and as fluid as the space of flows. Campaigns in the digital age are frequently leaderless and decentralised, with diverse actors and groups of actors functioning self-sufficiently or with minimal coordination, sometimes with different targets, but with the same defining goal. 'For example, groups of supporters may implement tactics in support of the campaign goal without the campaign initiator's knowledge (or consent)' (Joyce 2013). It is the origin of 'connective action,' defined as 'large-scale, sustained protests are using digital media in ways that go beyond sending and receiving messages,' which become a prominent part of the organisational structure (Bennett and Segerberg 2012, 739). Some of these action formations can embrace well-established advocacy organisations; others only afford them a small role. The main concept here is that ICTs enable 'personalised public engagement.' Gabriel separates the logic of connected action from the logic of collective action (ibid.). 'The familiar logic of collective action associated with high levels of organisational resources and the

[237] *Bella ciao* is an Italian partisan song originated during the Italian civil war (1943-45), which is used worldwide as a hymn of freedom and resistance.

formation of collective identities, and the less familiar logic of connective action based on personalised content sharing across media networks' (ibid.). In the former, using digital media does not change the core dynamics of the action; however, in the latter, they do because it is about individuals mobilising as individuals. People uttering their views online do not need to be part of an organisation to engage in political activity; but this dispersed political force has to be packaged and bundled to be impactful. It is still a matter of discussion, but connective action has had *net* positive impacts on people's participation, emancipation and freedoms. For example, Howard argues that across the universe of Muslim and partly-Muslim countries, ICTs diffusion has had a generally progressive impact on democratisation (2010, 37-57). Although he admits that authoritarian governments are using ICTs to spread propaganda, censor and monitor citizens, Howard also claims that ICTs allow for monitoring of the state by civil society organisations and individuals, the same way that the active participation of crowds in software development acts as a continuous collective improvement and correction mechanism[238] (ibid. 154-156). Besides, ICTs offer journalists and activists tools to defy the state's monopoly on information, and allow diaspora communities to network and share information (ibid. 40-41). Speaking about the high-tech surveillance of governments in China and Iran, Diamond still highlights the liberating potential of ICTs, since even activists in repressive countries have managed to leverage the properties of digital technologies to organise online campaigns of dissent (2010, 80). It is clear, then, that connective action has made protest and the monitoring of the powerful by the governed easier as well.

[238] 'Open source projects, products or initiatives are those that embrace and celebrate open exchange, collaborative participation, rapid prototyping, transparency, meritocracy and community development… Programmers who have access to a computer programme's source code can improve that programme by adding features to it or fixing parts that don't always work correctly' (opensource.com 2016).

Touraine explains social movements in relation with the phenomenon of globalisation and their socio-economic context, discussed earlier. This is a relevant question, since many of the data-based initiatives that have emerged within civil society seem envisioned to redress forms of exclusion that the globalisation has brought about. 'Anti-capitalist movements came to dominate an important section of public opinion, resulting in a capacity for massive mobilisation of discontented wage-earners and consumers' (2007, 21). Touraine initially seeks for signs of a new, unified social movement, different from the old social movement that represented the class clash between workers and owners of capital in the industrial society. A now classical example of a wide, if not unique, social movement is the World Social Forum (WSF), which emerged as the antithesis of the World Economic Forum (WEF) and the alter-globalisation movement. It met annually around the same time of the Davos summit of the powerful.[239] It was created in 2001 in Brazil around the idea of developing alternative futures and 'another globalisation' (Savio 2008). Some considered it as the physical manifestation of global civil society itself, as it gathered non-governmental organisations, TANs and GANs, campaigns, formal and informal social movements seeking international solidarity and the strength of numbers. Della Porta and Diani consider it 'the most visible example of network organisation,' which has then been extended to coordinate in a flexible and negotiated way the multiplicity of actors involved in global justice campaigns at the continental, national and local levels' (2006, 160). The most prominent WSFs took place in 2003 (Porto Alegre), 2004 (Mumbai) and 2005 (Porto Alegre), where up to 155,000 registered participants attended. The WSF attracted well-known thinkers and writers, including Galeano, Saramago, Ramonet, Savio, Cassen, Ali and Bello. In spite of the fact that there was

[239] Whitaker does not consider it a social movement but an open space 'for horizontal meetings –therefore without leaders, spokespersons or hierarchies– of civil society movements and organisations aiming at, autonomously in relation to political parties and governments, the overcoming of neoliberalism' (2004). Whitaker proposes the creation of a 'new movement,' not to be confused with the space (ibid.).

never a final statement or demand, its influence has been significant, as Savio points out. 'When we started in 2001, we were considered a fringe movement... Now, seven years later, nobody defends any longer the Washington Consensus. The damages it did worldwide have prompted the IFIs (international financial institutions) to do some significant corrections, and even the Bush administration is having several changes of route' (Savio 2008). From 2006 onwards, the movement became decentralised and its prominence started to fade away. It is my impression that, in rich countries, the financial crisis that started in 2007 had the effect of forcing social actors to look around their immediate environs harder and withdraw from the global arena to focus much more on the local dilemmas created by growing unemployment, poorer wealth distribution, austerity measures, decaying public services and massive evictions in their own countries, giving birth to more a different movement, global in values and methods but local in nature (i.e. 15M and Occupy Wall Street movements). That is, the mass of 'discontented wage-earners and consumers' (Touraine 2007), together with the new unemployed or unemployable, became involved in local protesting. In the Greiner's Growth Model's terms, forced by circumstances, the WSF succumbed to a growth crisis, even if it was not a standard organisation that had not resolved earlier leadership or control crises.[240]

Although the data activist organisations studied here are not social movements *per se*, they share some of their values and characteristics. In the paper published by Milan and me, we note that 'a few key concepts derived from the literature on social movements are particularly useful to study data activism: for example, one can study data activists' collective identity,' or what Melucci calls 'the interactive and shared definition' by individuals who recognise that they have

[240] One could say that the 15M –or indignados movement— overcame leadership and control crises when the political party Podemos emerged in 2014 with a founding manifest that prompted people to transform their indignation in political power and votes, and generated a very strong leadership.

certain alignments and ideals in common, and, on that basis, decide to act together (1996a, 70). One of the features that defines data activists' collective identity is technology –not only data infrastructures, but also the information and communication technologies employed to mobilise and coordinate action. Milan talks about 'tech-savvy activists who put their skills at the service of social change and social movements' (Milan 2014b). Social networking, for example, identifies many groupings using social media tools, which are creating new opportunities for collective action. These tools allow for 'ridiculously easy group formation,' which in turn connects and unites diffuse actors with common interests (Shirky 2008, 54). So much so that authorities with despotic strands frequently obstruct social media in order to prevent collective action. An example is DR Congo, where authorities shut down internet access and SMS services for mobile phones throughout the country after nationwide demonstrations in January 2015 (Committee to Protect Journalists 2015). In digital humanitarianism, this type of platforms is called 'activist mapping' or 'critical cartography,' which is a form of 'tech activism,' or activism that 'concerns itself with creating and shaping digital technology infrastructure and tools to facilitate communication and networking for social change activists, and free likeminded individuals from the constraints and threats of commercial communications' (Milan 2014b). The systematic use of the internet in activism is in fact another trait that can characterise these groups. Talking about the role of the internet in activism, van de Donk, Loader, Nixon and Rucht say that it is seen by activists 'as a way to bypass the professors of the scientific council' in order to access information (van de Donk et al. 2004). That is, as data-based knowledge and the internet –as a vehicle for knowledge— can be a social equaliser as well. 'The internet provides each activist with the same data, documents and arguments and contributes to the process of becoming an expert, that is, it raises new matters of concerns on the agenda. The internet allows

each activist to become an expert, to spread knowledge about globalisation and to diminish the differences between experts and non-experts' (ibid., 150). These authors arrive at valuable conclusions, even if their study took place before the popularisation of data infrastructures and intensification of social networking. In my opinion, it does not matter whether it is transmitted via mobile technologies, the internet or ham radio: data-based knowledge empowers activists, provides them with the same arguments and allows them to become experts, as these authors propose about the internet.

Similar to social movements, proactive data activists work towards long-term norm change. As social movements, data activists are identified by the specific use they make of technology to form a collective identity, communicate, convey, mobilise, demonstrate and act. In the long run, data activism is likely to influence the way citizens approach computational politics and the informational state, as well as the way we see and practice social change (Milan and Gutiérrez 2015, 135), as it can be seen in some of the cases examined in the next section. Proactive data activism emerges from citizens' realm, and as with citizens' media, paraphrasing Rodríguez, it rejects tight definition and classification, since each interaction 'takes a unique profile' and exists in 'a polymorphic ensemble' (2001, 164).

Next, Table 2 summarises what these different perspectives and theories can say about proactive data activism.

Table 2: Proactive data activism seen from different perspectives

Perspective	Description
Critical theory	A communicate action, with empirical-analytical, hermeneutic and critical elements
Journalism	Similar to three varieties of journalism: investigative journalism for its analytic nature; advocacy journalism for its political stand; and citizen journalism for its collaborative,

	rank and file character
Alternative media	Involves a politics of the quotidian, as it alters the everyday relationship between citizens and automatized data collection. Places citizens at the forefront of production, which constitutes a powerful transformative and empowering process
International relations	It is local and transnational at the same time, it is based in collaboration across borders
Social movement	Works towards long-term norm change and gathers people around the specific use they make of technology to form a collective identity, communicate, convey, mobilise, demonstrate and act

Source: Elaboration by the author.

In the case study, I employ some concepts originated in critical theory, and journalism, alternative media, international relations and social movement studies in order to examine Ushahidi more in-depth from a multidisciplinary viewpoint, and corroborate whether it conforms to an effective case of proactive data activism. What follows is a description and classification of actual cases of proactive data activism and connected initiatives from an empirical perspective. The objective is to distil the main *sine qua non* characteristics of effective proactive data activism, based on a description and analysis of separate cases, in order conclude with a model that can be used in exploring other cases in the future.

6. The shapes of proactive data activism

> Data socialization is a term that's being heard more frequently, and it's an attempt to tackle the ongoing (and often mission-critical) process of making sure the right data is in front of the right person – a decision maker – at the right time. Get this wrong, and despite all of your good intentions it is likely that your strategy for data-driven transformation will become one of the many Big Data projects that fail (Marr 2017).

Kennedy, Poell and van Dijck argue about the need to ground studies of big data in real-world practices, and conclude that the 'debate on the power of the algorithm… leaves little room to explore the small-scale actors who are making organisational adjustments to accommodate the rise of data's power' (2015, 2). Big data not only are forcing organisations to make adjustments, as for example Prest notes in her interview for this dissertation; they are also facilitating and inspiring the creation of new organisations and projects at a small scale in the 'real world.' But what else can be said about them? How can they be categorised and further studied? What are the characteristics typical of proactive data activist initiatives in the 'real world'? What elements can be observed in order to single out this type of activism from others? What are the action repertoires typical of these proactive data activist ventures? In this section, I try to answer these questions by looking more in-depth into some data activist initiatives from different angles, that of tech activism and data journalism; identify the presence of certain common features; and finally examine the type and origins of the data employed by these projects. I first examine an existing classification of a related phenomenon ('tech activism') in order to apply it to proactive data activism. Subsequently, I pay special attention at civil society organisations that produce data journalism, since by filling a gap where media organisations are not doing their jobs, they also respond to an implicit social demand. Next, I include my own classification of proactive

data projects and related activities, which I divide in four main categories: the organisations that produce data journalism, the 'skills transferers,' the 'catalysts' and the actual 'data activists.' Skills transferers are themselves divided in three groups: the ones that transfer data or social sciences skills to other organisations lacking those abilities; the ones that facilitate match-making opportunities; and the ones that develop data infrastructures and tools, which can foster the realisation of data projects. The catalysts usually fund and facilitate social data projects by providing resources, both monetised or in kind. And the actual proactive data activists engage largely in creating alternative mapping systems and geoactivism –the most extended form of proactive data activism observed so far. I also define key related concepts, including 'map,' 'geoactivism,' 'crisis mapping' and 'activist mapping,' provide examples of these activities and the action repertoires of geoactivism, and describe the usual technological platforms employed in geoactivism, as said, the main form of proactive data activism. And I finally look at them from the perspective of the origins of the data employed in each project can be a criterion for classification. Generally, organisations specialise in using only one or two methods to obtain data, and therefore, looking at where data come from is a good measure for the classification of empirical cases. This way, proactive data activist projects can be divided in a) the ones that rely on whistle-blowers for data; b) the ones that resort to open-source datasets; c) the ones that use crowdsourcing tools; d) the ones that turn to appropriating data; and e) the ones that get data from primary research that can be datafied. To finish I look at the associations among data activist organisations brought about by the use of data infrastructures. I complete this section looking a key concept: the emancipation properties of big data through data activism, and the empowerment process that takes place by using data infrastructures for advocacy and action. This is possibly the most important process in data activism, and therefore deserves attention.

Data infrastructures in data activism are not only social equalisers; they can also enable action; help people overcome exclusion; and foster real participation.

Let us look first at an existing classification of the action repertoires of a related activity, such as 'tech activism,'[241] in order to determine whether it can be applied to proactive data activism as well. Taking into consideration that tech activism is not for your ordinary user, and looking at how cyber activist action repertoires differ from those of other civil society groups, Milan and Hintz propose that tech activists are considered 'beyond-ers,' as opposed to 'insiders' –who 'interact directly and cooperatively with power-holders through advocacy, lobbying and, in some cases, participation in multi-stakeholder forums.' Beyonders, instead, 'question the legitimacy of power-holders and address them through protest and disruptive action' (Milan and Hintz 2013, 20-21). Beyonders have 'no interaction with policy processes,' hold 'no dialogue with institutions,' bypass regulation and expand 'unregulated spaces,' and focus on 'prefigurative action, envisioning and creating a different system, both at the material and symbolic level, setting up alternative infrastructures and generating alternative sets of rules' (ibid., 22). Reactive data activists respond perfectly to the description of beyonders, and could be considered a subcategory –that is, a reactive data activist is a beyonder that uses data infrastructures as his or her main tool for analysis, communication and action. As reviewed earlier in this dissertation, many campaigns could also be considered as part of the beyonders' realm, as they are based mainly on confrontational tactics (i.e. Greenpeace's campaign in the Arctic). Proactive data activists, on the other hand, do not conform to the same description. As

[241] Hereafter, I will use these 'tech activism' and 'cyber activism' as synonyms. I understand tech activism as the process of using internet-based social networking and communication techniques 'to create, operate and manage activism of any type' (Technopedia 2016b). It allows individuals and groups to use social networks and other online technologies to reach and gather followers, disseminate messages and support a cause. Cyber activism is also known as tech activism, internet activism, online activism, digital activism, online organising, electronic advocacy, e-campaigning and e-activism (ibid.).

beyonders, and more specifically reactive data activists, proactive data activists can create parallel systems, and alternative communication infrastructures and public spheres for action, both at the symbolic and material levels. This is the case, for example, of Ushahidi, which generates independent software-based platforms for data activism. However, proactive data activists lack the confrontational, elitist, out-of-the-system nature that defines beyonders and reactive data activists, who 'seek to act regardless of the logics and rules of the game of the known social system' (ibid.), creating and recreating their own rules and spaces. Then again, proactive data activists are more concerned with proposing, enhancing, acting and doing, than with shielding, defending or protesting. Milan and Hintz described beyonders in 2013,[242] and that label could have applied to reactive data activists at the time. However, since then, data activists have evolved in a similar way to how Anonymous developed, becoming more propositive (Gorenstein-Massa 2013, 63); in other words, as 'hacker culture' became more 'mainstream' (Barnes 2000), reactive data activism turned into a more popular and accessible way of activism. As a result, a different type of data activism branched out of it, that is, proactive data activism, which does not comply with the beyonder activism's criteria any longer. Proactive data activists seem to be data- and technology-centred; generate data analysis, maps and other visualisations; feel comfortable crossing sectorial lines dividing activism, journalism, research, and unreservedly promote cross-fertilisation between them; generate theories of change; create their own independent public spheres of communication and action when needed; work better in alliance or collaborating with others; and tend to be less confrontational than reactive activists and more cooperative with authorities, establishment and norms. I look at these characteristics

[242] Based on Milan's PhD dissertation in Political and Social Sciences from the European University Institute.

more in-depth later, and use them as a way of categorising data activism and come up with a model for effective data activism.

Now I examine what actual forms proactive data activism takes. What follows is by no means a comprehensive or detailed account of existing proactive data activist organisations and initiatives around the globe. I am biased by the languages I speak, the experience I have had and the interests I pursue, so the examples included here come mainly from Europe, America and Africa, are operated mainly in Spanish, English and Portuguese, to the detriment of other regions and languages, and focus on issues such as human rights, humanitarianism, climate change and the environment. The objective, though, is to outline a classification of proactive data activist organisations, propose criteria that tell them apart, and offer a classification, which should be enriched and perfected in the future, and a applied to other cases in other regions.

6.1 Doing data journalism and filling a gap

I have already mentioned the civil society organisations dedicated to producing data journalism. I am including them here because, as said, even if the product is of journalistic nature, the objectives behind its production always include an implicit demand or a statement, a theory of change, and are part of a strategy within a campaigning logic. Former Greenpeace International Executive Director Kumi Naidoo, for example, criticises how NGOs have taken up the role of signalling the difference between the academia's scientific consensus on climate change and doubtful information published by 'climate sceptics,' whose research is mostly funded by fossil fuel companies (Goldenberg 2015). Naidoo is not comfortable in that role. 'We shouldn't have to do the job of journalism in sorting out real research from paid agents' (2012,

113).[243] In the specific case of climate change, and given the vast amounts of money at stake, there has been an increasing 'politisation of the triple interface of climate science-media-policy,' ascribed to 'the coalescence of a small group of influential spokespeople and scientists emerged in the news to refute scientific findings regarding human contributions to climate change,' who have 'received funding from carbon-based industry interests' (Boykoff and Roberts 2007, 6). This situation has generated an avalanche of 'climate sceptical' coverage, which campaigners feel must be neutralised, or at least balanced, with facts and analysis, which sometimes comes in the form of journalism. Research and analysis is a strategic element of any serious campaign, but as we have seen, they are the *meat and potatoes* of journalism as well. This is precisely what many civil society organisations are doing: producing data journalistic content where media do not. In this section, I am going to attempt a proposed classification of proactive data activist organisations, including the case of the ones producing data journalism, for two reasons: data (advocacy) journalism is possibly one the closest activities to proactive data activism (even taking into account the differences examined earlier), and mostly because some organisations practice both advocacy data journalism and proactive data activism at the same time. This initial classification includes four main groups: activist organisations producing data journalism; organisations transferring skills to other organisations in order to foster proactive data activist projects; organisations making data activist projects possible by providing resources (as opposed

[243] 'We at Greenpeace therefore often ourselves shed light on the murky world of fossil fuel corporations buying influence and preventing effective climate action. Many journalists have thanked us for websites such as www.exxonsecrets.org –which show the links between, in this case, Exxon (Mobil) and climate deniers, including politicians… The fossil fuel industry managed to delay action on climate change with a deliberate and heavy assault on the free press, insisting for years that a "balanced view" be presented of what was in fact an unmitigated cry of alarm from the scientific community. As voices of dissent in the real scientific community became fewer and fewer, they simply funded pseudo-scientific institutes to market untruths as science. We did our best to expose some of these "Paid for by big oil" spokespersons through our www.exxonsecrets.org website as a service to the media, but we shouldn't have to do the job of journalism in sorting out real research from paid agents' (Naidoo 2012, 110-115).

to transferring skills), acting as *catalysts*; and the actual proactive data organisations, projects and contents. I will also revisit the initial questions of how proactive data activist organisations can be classified, and what groupings are more conductive to data activities and practices.

As noted, one of the shapes that data activism is taking is that of the civil society organisations practicing data journalism because they are deliberately filling a gap. By performing an atypical public service, they are also bringing attention to the very fact that journalists and media organisations are not doing their jobs, especially in countries where data journalism is at its early stages and transparency laws are yet to be voted for or are very recent. Some, as Naidoo, do explicitly signal it; others do not. Although their main objective can be simply producing good data-based stories, data journalism originating from a campaigning organisation, however, comprises an underlying objective as well, which goes beyond just journalism and enters the sphere of civic action. This is something that (advocacy) data journalism and data activism share. However, as discussed earlier, the differences are that data activists are mainly activists working away under a campaigning logic, while advocacy data journalists remain always journalists, who should work within the rules, constraints and ethics of journalism, as defined by Kovach and Rosenstiel (2007). In fact, a feature that seems to be common to all proactive data activism is brazen hybridisation. Data activists do not explain themselves when they cross lines between journalism, activism, training or research, and have not bothered yet with discussions of whether they are losing the journalistic objectivity ethos when they are selective or vocal about social issues. An example could include the graphic non-fiction journalistic report *Vagabundos de la chatarra* (Scrap drifters),[244] a year-long journalistic

[244] See normaeditorial.com/ficha/012034515/barcelona-los-vagabundos-de-la-chatarra/ [accessed on April 16, 2016]. The economic crisis in Spain is generating a plethora of interesting artistic and journalistic

project that has produced a graphic book and a web site with some data visualisations (although this is not a full-blown data undertaking). The editorial project –put together by journalist Carrión and graphic designer Sagar— comprises observations, interviews and accounts of people who gathered and sold scraps metal for a living during the economic crisis in the edges of Barcelona. The content of this project has been produced using journalistic techniques, including the standard interview and observations from reality, as well as data analysis and maps. However, the way of communicating this journalistic research is innovative because it is very hybrid. As a result, the project can be interpreted from several approaches: the journalistic angle, the data angle, the social angle and the comic design angle. Carrión defends the formula saying that objectivity in journalism does not exist anyway, and that in graphic non-fiction what is important is social commitment (Carrión 2016).[245] Interestingly, the book which is the product of this endeavour includes at the end a comic-stripped interview with Sacco, in which he says that 'what is worthwhile of journalism is the commitment. Facts matter, reality matters, victims matter. You have to question power. These are the moral essentials you have to defend' (Carrión and Sagar 100).[246] As I have pointed out at the beginning of this section, in countries where data

---

manifestations, which combine social and campaigning elements as well. Another example is *Cerca de tu casa* (Close to home), a musical film about the evictions emergency.

[245] He is following in other journalists' footsteps. Journalists that practice graphic non-fiction –minus the data angle— include Joe Sacco, who has published books on Palestine; Dan Archer, who has worked on a number of social justice topics, from homelessness and the financial crisis to human trafficking; Matt Bors, who has talked about the proliferation of weapons and the death penalty in the US; and Susie Cagle, who has published work on wildfires and the effect of climate change on drought cycles, among others. All of them share the fact that they are committed to social causes, that they apply journalistic rules to their job of gathering information, and that the stuff of their stories is real life. Maybe in the future we will see more non-fiction graphic works that combine data journalism as well.

[246] [Lo que importa del periodismo es el compromiso. Los hechos importan, la realidad importa, las victimas importan. Hay que cuestionar el poder. Esos son los fundamentos morales que hay que defender (ibid.)]. In the same interview, the participants, Carrion, Sagar and Sacco, defend also fiction in TV series as a way of revealing some truths, and say that, 'without fiction, you have the facts, the data, but not always can connect them. You have facts A and B about a corrupt politician, but cannot link them. Fiction allows you to do that' [Sin ficción tú tienes los hechos, los datos, pero no siempre los puedes

journalism has not been yet embraced by news media organisations, others, such as individual citizens, artists and civil society organisations, are filling the gap, as well as exploring new ways of communicating journalistic findings and content. This is the case of Spain, where –judging by empiric observation— NGOs are doing more data journalism than journalists themselves. In a context where nobody else can or dares to practice data journalism, where stale business models continue to render benefits,[247] where opinions fly right and left, data journalistic projects are becoming a social statement, sometimes a social struggle. At an individual level, Cabra explains how, when moving from the US to Spain with the idea of simply doing data journalism, she realised that she had to campaign for a transparency law before she could put her hands on public data, because *public* data was not public, much less open. She then got involved in the campaign for a transparency law in Spain. Cabra has participated in many data-enabled investigative reports and is one of the journalists behind the coverage of biggest leak in history, the 'Panama Papers,' described as 'a giant leak of more than 11.5 million financial and legal records exposes a system that enables crime, corruption and wrongdoing, hidden by secretive offshore companies' (International Consortium of Investigative Journalists 2016). A total of 376 journalists in 70 countries reviewing and analysing 11.5 million documents, recording 40 years (from 1977 until 2015) of 214,488 transactions from financial entities in 200 countries, in which at least 140 politicians and active high-ranking officials were involved (ibid.). Without the assistance of big data infrastructures, this investigative report would not have been possible. In an interview, project's coordinator Walker reveals the strategy behind the leak: 'Because part of our model is that we all publish together. That's because we want to create a commotion when

---

conectar. Tienes el hecho A y B sobre un político corrupto, pero no puedes relacionarnos. Con la fiction sí puedes hacerlo (ibid., 96)].

[247] *Diario Vasco*, a local paper based in San Sebastian which is part of Vocento Group, still delivers two digit benefits, according to one of the paper's managers, with a conventional formula of print-based local news catered for an aging readership.

we publish. We want to have global impact. We don't want the story to drip and to start turning out in little pieces that make the news for one day or two in a country and then go away' (2016). Although it does not explicitly make the connection, with this giant joint and well-timed effort, the ICIJ is looking to shine a light on culprits and illegal practices, and ultimately change policy. 'These latest disclosures dramatically expand on previous leaks of offshore records that ICIJ and its reporting partners have revealed in the past four years. ICIJ's stories have triggered official inquiries, high-profile resignations and policy changes around the world' (Center for Public Integrity 2016). This statement implicitly gives away an unspoken theory of change: pernicious policies that allow corrupt practices can be changed by publishing exposés of their consequences, generating attention and outrage, and forcing governments to initiate legal processes. If public data are not available, laws that grant access to relevant data are needed. As it is discussed later, public data should be open by default. This is an area where organisations of the civil society have been active too: demanding access to public data, as the case of Civio shows in Spain. In countries that have passed laws that grant open access to public data, these legal instruments have been 'a great incentive, but they are not enough' (Ávila 2015). There are other factors that influence the ability to access data. Ávila notes that Mexican journalists who have dared to publish data-based reports on organised crime, for example, have been attacked (ibid.). A general context of lingering authoritarian practices and opaque relations between politicians and criminal groups make things difficult for data journalists and activists (ibid.). Not all civil society organisations have the ability to engage with big data either. But the shift towards big data infrastructures may end up shaping the structure of the organisational ecology of civil society. An example of this is presented by Prest, interviewed for this dissertation. Although not common, she for example talks about 'the case of an organisation that has changed its strategies,

has hired an eight person team and now in every meeting, they ask themselves "What does data tell us?"' Civio, which has been one of Cabra's playgrounds,[248] is another example of this phenomenon as well. Civio (see the next section too) publishes data-based information in several thematic sections, among other things. For example, it its project 'Quién manda' (Who is in charge), there is a journalistic story that reveals that 249 majors were seeking re-election in 2015, in spite of the fact that they broke the law when they had repeatedly failed to submit their municipal budgets to the Court of Auditors (Gavilanes 2015). Other story-producing projects associated to Civio include: 'El Indultómetro' (The Pardonmeter),[249] dedicated to register all the inappropriate legal pardons favouring corrupt officials and the like; 'Dónde van mis impuestos' (Where does my money go),[250] devoted to visualising where tax money comes from and how it is spent; 'El BOE nuestro de cada día,'[251] which digs out stories hidden in the massive official Spanish state gazette, the so called BOE; and 'Tu derecho a saber' (What do they know),[252] a web site dedicated to channelling public information inquiries sent by citizens or groups, which also published analyses and expert commentary on transparency issues. All the stories, commentary and information published by Civio include an implicit social demand. In Mexico, Morlan[253] is a private company dedicated to data analysis with three criteria: scientific

[248] In 2012, Cabra was Civio's executive director.
[249] See elindultometro.es/index.html [accessed on March 30, 2016].
[250] See dondevanmisimpuestos.es/[accessed on March 30, 2016]. This is a version of the application 'Where does my money go' (see wheredoesmymoneygo.org/), part of Open Spending, was developed by Open Knowledge Foundation. In a similar way, 'Where does my money go' publishes stories, analyses and visualisations on public spending.
[251] See elboenuestrodecadadia.com/ [accessed on March 30, 2016].
[252] See tuderechoasaber.es/ [accessed on March 30, 2016].This is a version of the application 'What do they know' (see whatdotheyknow.com/), developed by mySociety. 'What do they know' includes stories from around the globe on freedoms and transparency. In December 2015, 'Tu derecho a saber' was discontinued due to the 'unwavering administrative and technical barriers' created by the government to circumvent citizens' inquiries (2015).
[253] See morlan.mx/#trabajos [accessed on April 1, 2016]. Morlan does not reveal how it funds its activities, although it works in joint projects with other organisations.

robustness, social context and design. At the time of writing, Morlan is to put together a project, for example, analysing videos of the electoral campaign in order to respond to questions about how well candidates present their proposals and respond to attacks during debates. Morlan works together with journalistic organisations, such as *El Universal*, in joint data projects (Redaccion El Universal 2016).

In the specific area of environmental journalism, open data from satellites is transforming environmental reporting, many times in association with or enabled by campaigning organisations producing journalistic-type outputs. An example is InfoAmazonia, which combines citizen participation and journalism, and data analysis. InfoAmazonia provides news and reports of the endangered Amazon region, based on the work of a network of organisations and journalists delivering updates from the nine countries of the forest (2016). Since its launch in 2012, InfoAmazonia has been training Brazilian journalists to use satellite imagery and collect data related to forest fires, carbon monoxide and deforestation, as well as publishing information and interactive maps. This approach to storytelling using maps has been dubbed *geojournalism* or the practice of telling stories with geographic information systems (GIS) data generated by the earth sciences. This sort of experience is analogous to *geoactivism*, a modality of proactive data activism which uses interactive maps to communicate and enable action, as seen later. Environmental News Lab[254] is a facility supporting InfoAmazonia. 'Based on experience of using digital tools to report on deforestation in the Amazon rainforest,' Environmental News Lab 'experiment frequently, promote more training, and share with other professionals what it means to innovate' (2016). This organisation not only trains journalists, but can 'also deploy sensors, mobile apps and interactive maps to tell the stories' (ibid.), generating their own data in order to

[254] See lab.oeco.org.br/ [accessed on April 3, 2016].

put in practice geojournalist projects in Brazil, in collaboration with O ECO.[255] It also launched a *Geojournalism Handbook* (Geojournalism.org), in partnership with the International Centre for Journalists (ICFJ),[256] Internews' Earth Journalism Network,[257] and the Flag It! Project.[258] The online handbook explains how to use open source, data infrastructures and web-based tools to collect, organise, visualise and publish data. Other examples of geojournalism include the Oxpeckers Centre for Investigative Environmental Journalism in South Africa, where journalists track poaching of rhinoceroses in the country's national parks (Code for Africa 2016).

A classification of proactive data activist organisations can be established based on what they do and what their objectives are, although this is not an easy task due to the fact that, as said, many organisations cross the lines between activism, journalism and training without any qualms. Data activist organisations can be focussed on skills transfer in order to, again, fill a gap; they can be focused on catalysing or making data activist projects possible, providing funds and resources; and they can be focused in actually producing proactive data activist content, and using it in campaigns and social action. Let us examine them next.

---

[255] See oeco.org.br/ [accessed on April 3, 2016]. This is another Brazilian NGO dedicated to skills transfer in environmental journalism. It was born with a grant from the Avina Foundation, which supports projects of social or environmental interest in Latin America. Today, its main supporter is the Grupo Boticário. But throughout its existence, it has received support from the Blue Moon Foundation, Google, Porticus, Hewlett Foundation and sponsorship from various companies such as Natura, Cetrel, Braskem, Bradesco, Hi, Siemens, xTAM, Petrobras and Vale.

[256] See icfj.org/ [accessed on April 3, 2016]. The ICFJ launches programmes directed at empowering journalists and engaging citizens with new technologies and best practices (International Center for Journalists 2016). 'We believe that better journalism leads to better lives' (ibid.).

[257] See earthjournalism.net/ [accessed on April 3, 2016]. Developed by Internews 'to empower and enable journalists from developing countries to cover the environment more effectively' (Earth Journalism Network 2016). EJN 'establishes networks of environmental journalists in countries where they don't exist, and builds their capacity where they do, through training workshops and development of training materials, support for production and distribution, and dispersing small grants' (ibid.).

[258] See flagit.youthpress.org/ [accessed on April 3, 2016]. The 'Flag it!' project 'aims to train young media makers from all over the world on how to report on environmental issues at local level with a global impact by using digital visualisation tools in open source' ('Flag it!' project 2016). The project is funded by the European Commission (ibid.).

6.2 The skill ‘transferers’

As said, proactive data projects and related activities can be classified taking into account some of the elements, tactics and behaviours that have been observed so far (i.e. do they produce data journalism? do they provide funds to create data projects? do they generate digital tools for data activism?). That way it can be said that these projects and organisations can specialise in transferring skills or incentivising data projects or actually practicing data activism. From empiric observation, each of these cases shares some characteristics that typify and group individual them together. The first observed cluster is the one that congregates organisations and initiatives that specialise in transferring data skills or social research skills, precisely because of the gaps there exist as a result of the failure to fulfil the data promise in many areas. I have considered this group as part of the proactive data activist family because, although their main objective is not to produce proactive data activist content, they are making an underlying statement by solving a problem –lack of skills in the data activist field—, building networks where there were none, and bridging a gap between the skills-holders and the unskilled. As said, data activism is not for punks, since it requires the acquisition of a sophisticated set of abilities. This group can be pigeonholed in four assemblages –data skills and social skills transferers, match-makers and data tools developers—, although in some cases there are not many representatives just yet or I am not aware of existing projects in other languages and communities. It is to be expected that more organisations will join in as this activity expands.

6.2.1 The data skills ‘transferers’

Some organisations concentrate on transferring data infrastructures skills, of which many are in universities. I will refer here only to those that work as independent organisations (even if

a university is behind them), but not to the formal education centres dedicated to data science.[259] DataKind, already mentioned, is the perfect example. As said, it specialises in selecting social projects where their volunteering data scientist can be deployed in order to work together in identifying opportunities for data analysis and developing data projects. Others include Bayes Impact,[260] which is 'a group of full-time data scientists, engineers and academics who believe data science can be used to solve the world's most ambitious problems' (2015). This group of experts builds 'operational solutions to social problems by building software for governments and non-profit organisations' (ibid.). School of Data,[261] with its motto 'evidence is power,' describes itself as a 'network of data literacy practitioners, both organisations and individuals, implementing training and other data literacy activities in their respective countries and regions' (2016). Its mission is to 'empower civil society organizations (CSOs), journalists, governments and citizens with the skills they need to use data effectively in their efforts to create better, more equitable and more sustainable societies' through training (ibid.). Again, here we see the hybridisation of journalism and activism put together at the same level. The School of Data is led by the Open Knowledge Foundation (OKFN) and Peer 2 Peer University (P2PU), and supported by organisations such as Tactical Technology Collective[262] to produce courses geared towards their human rights audience.

---

[259] I, for one, head a postgraduate programme called "Data analysis, research and communication," which precisely transfers skills within the formal framwork of the University of Deusto.

[260] See bayesimpact.org/ [accessed on March 8, 2016]. Bayes Impact is supported by Microsoft, Tableau or LinkedIn, among other companies.

[261] See http://schoolofdata.org/ [accessed on April 3, 2016].

[262] Mentioned earlier, Tactical Tech works with an international network of partners 'to help rights, accountability and transparency advocates and the communities they work with to use information and digital technologies effectively in their work' (Tactical Tech 2009). Thematically, Tactical Tech works on the politics of data, digital security and privacy, and 'new forms of finding, creating and representing evidence by advocacy and activist groups and individuals' (ibid.). In 2013, it published *Visualising Information for Advocacy*, a handbook about how advocates and activists use visual elements in their campaigns. It features 'nearly 50 case studies from around the world to provide an introduction to

Tactical Tech, already mentioned when dealing with info activism, deserves another parenthetic reference here, as this organisation is dedicated to increasing the security of humanitarian workers and operations as well through the transfer of skills. Together with the opportunities for digital humanitarianism and data activism, digital technologies 'have simultaneously created new points of weakness: exposing (human rights defenders) HRDs' whereabouts, activities and networks, and creating evidence against them through data leakages, digital traces and direct surveillance and interception' (Hankey and Clunaigh 2013, 536). Meier acknowledges that, just because social media messages shared during crises and disasters are freely available, utilising them is not inevitably ethical — social media users' safety might be compromised in some cases (2015a, 91). Mitchell, from Ushahidi, admits as much in an interview carried out for this dissertation, in which he says that he is working ways to protect eyewitnesses reporting cases of abuse. But the fact remains that facilitated by technology, attacks on HRDs have escalated, with an 'increase in the number of entrapments and networks being compromised through the use of computers, cameras, mobile phones and the internet' (Hankey and Clunaigh 2013, 536). These authors conclude that 'the integration of digital tools, information and communications technologies into the work and everyday lives of HRDs is ever deeper and cannot be ignored in any intervention,' and that 'given the resource differential between HRDs and many of their adversaries and the rapidly developing technological climate, the community of practitioners trying to build HRDs' capacities in digital security can no longer afford to adopt an approach that fosters dependence upon their direct advice in order to stay safe' (ibid., 546). Fittingly, Tactical Tech also transfers skills to handle these situations.

---

understanding visual information and a framework for using images for influence' (ibid.) (Emerson 2008).

Apart from working with data journalists, as seen earlier, another data skills transferer, Code for Africa,[263] liberates 'deadwood' data, runs a fellowship programme for media and civil society representatives, does data training (including seminars, policy roundtables, skills *bootcamps*, advanced masterclasses and Code4Democracy hackathons), and supports the 'development of citizen-driven solutions, such as the GotToVote[264] toolkit that has already been used in Kenya, Malawi and Zimbabwe' (Code for Africa 2016). In order to maximise existing tools, *Code for Africa* 'champions re-use and replication wherever appropriate, by supporting the work of pioneers such as mySociety[265] and the Open Knowledge Foundation, and others' (ibid.). Data4Good is 'a social business' based in Brazil that aims to 'promote the usage of data for social good' connecting the corporate pro-profit world with 'driving action in the right direction' (2016). Data4good is attempting to use data 'as an instrument of change and solution of social problems,' based on three spheres of action: mobilisation ('creating a network of volunteers and data fellows to support NGOs and social causes'); education '(producing content to generate education and awareness around the usage of data for public issues'); and consultancy ('providing consulting services to support social organisations with their biggest data challenges') (ibid.). Data4Good also has a skills transfer component in its activities as well.

There are many more examples of data skills transferers, especially if projects producing advocacy data journalism are included. This variation, in fact, might be one of the most common

---

[263] Code for Africa is funded by organisations such as World Bank Group, Hivos, Bill & Melinda Gates Foundation, the International Centre for Journalists (ICFJ) and the Praekelt Foundation, among others.
[264] See gottovote.cc/ [accessed on April 5, 2016].
[265] See mysociety.org/ [accessed on April 5, 2016]. MySociety is s a non-profit 'social enterprise' that builds 'online technologies that give people the power to get things changed,' and shares 'these technologies so that they can be used anywhere' (mySociety 2016). Although not specialised in data infrastructures, mySociety does facilitate and implement data projects. One example is EveryPolitician, which via crowdsourcing, gathers and visualises data from politicians around the globe. At the time of writing, it had amassed data on 67,855 politicians from 233 countries, making it dataset on politicians. However, the data are not very granular.

ways data skills are being acquired and absorbed by civil society organisations. However, Prest, interviewed for this dissertation, notes that DataKind has difficulties in finding social organisations that can absorb data skills, while there is no shortage of volunteering data scientists. It is my belief that the most important thing in a campaign is not the technology, even if it is not for everybody, but its theory of change and how information is strategized and mobilised for social change.

6.2.2 The social skills 'transferers'

The first phase in the growth of any organisation, according to the Greiner's Growth Model, is growth through creativity (Exponential Training 2016), and in the case of non-profits, this is often based on strong charismatic leaders that attract large numbers of volunteers, but not so much on robust knowledge and competences. There are many programmes and projects that are focused on transferring skills from other fields of action (i.e. accounting or marketing) into campaigning organisations that may have growth led by charismatic leaders out of the comfort zones. As they have gone through a professionalization process, advocacy organisations now produce budgets, indicators and strategies, and recruit professionals in marketing, research, management, communication, business development and fund-raising. Whether data skills are harder to absorb by non-profits or not, the truth is that they have been taking in and adapting skills from the private sector for decades now. However, I have always thought that businesses and research centres had a lot to learn from campaigning organisations in terms of the 'what fors,' and the inspiration and the motivation that drives their efforts. In fact, there are a few organisations that are dedicated to do precisely this: training data scientists and researchers in social issues to provide them with the 'what fors.' Your average data scientist may know how to 'play' with data and get correlations, but he or she may also lack the skills to identify, communicate, narrate and utilise the findings behind the data. And that us because the 'what for'

element is often missing. That is why some projects, mostly incubated at US universities, specialise in training data scientists to work directly with non-profits. Although the approach is from the other side, DataKind's ability to find numerous volunteering data scientists willing to work *pro bono* with social organisations reflects the appetite that many data scientists feel to be socially relevant and useful. This field of activity may be on the rise. According to Marín Miró, interviewed for this dissertation, 'since big data and related technologies are being commoditised, the focus will shift to the social sciences, and data teams will be led by people with social and communication backgrounds.' Frustrated with the corporate approach to data science, Marín Miró –a telecommunications engineer— set up Outliers Collective in order to analyse and visualise data for social causes. 'These tools are like microscopes and telescopes, but at the other end of the lens what you always find are human behaviours (atoms as opposed to bits and datasets).' This telecommunications engineer says Outliers Collective's main strength is basically filling 'the gap between social and data sciences. It is really difficult to find a place where you experience both worlds, so we had to create it.' Another example is Data Science for Social Good, which teaches 'data scientists to tackle problems that really matter' through a fellowship programme.[266] The programme brings dozens of data scientists to the University of Chicago to work on analytics projects with non-profits, local governments and federal agencies. This organisation has a fellowship programme that allows students to 'solve high impact problems' (2016). The programme makes data students work directly with those 'at the forefront of public policy decision-making,' and perform data analysis tasks with real data together with the experts that collected and created them so fellows are not just conducting research on theoretical problems, 'but rather are working with a real team with solid goals' (ibid.). This modality has other followers, including eScience Institute of the University of Washington, in

[266] See dssg.uchicago.edu/ [accessed on March 8, 2016].

collaboration with Urban@UW[267] and Microsoft. It organises programmes modelled after Data Science for Social Good, which bring together 'data and domain scientists to work on focused, collaborative projects that are designed to impact public policy for social benefit' (eScience Institute, 2016). The programme selects fellows to 'work with academic researchers, data scientists and public stakeholder groups on data-intensive research projects' (ibid.) Georgia Tech has a similar programme as well.

6.2.3 The match-makers

These organisations do not focus on transferring skills from one space (the technical realm) to another (the social realm), or the other way around; they put together representatives of these communities, stimulate and facilitate contacts between them, and see if magic happens. Medialab-Prado[268] specialises in summoning all sorts of profiles –from artists, performers and designers to journalists, social workers, campaigners, activists, data scientists and engineers—, and enabling and stimulating collaboration between them around concrete data projects. Its methodology works like this: For two weeks, gathered in Madrid, the representatives of a number of pre-selected projects with a common theme present them to an audience of strangers summoned for the occasion –including data scientists, journalists, engineers, journalists, artists and the like—, and make a case trying to marshal support and resources. The participants in the audience have the opportunity to ask questions and debate the projects, and then make a decision about which project to work on for the remaining of the two weeks. Medialab-Prado pays for basic expenses. It defines itself as a 'citizen laboratory of production, research and broadcasting of cultural projects that explores the forms of experimentation and collaborative learning that

[267] See urban.uw.edu/ [accessed on April 6, 2016].

[268] See medialab-prado.es/?lang=en [accessed on April 3, 2016]. Medialab-Prado but also trains journalists. Medialab-Prado is part of the Department of Culture and Sports of the city of Madrid and is funded mainly by the city.

have emerged from digital networks,' and talks about promoting a 'culture of data' (2016).[269] The already mentioned Hacks/Hackers,[270] although focussed on skills transfer for journalism, has nonetheless an activist attitude and a match-maker aspiration. It is defined as 'a network of people interested in web/digital application development and technology innovation supporting the mission and goals of journalism' (2016), and a 'digital community of people who seek to inspire each other, share information (and code) and collaborate to invent the future of media and journalism' in order to 'make sense of the world' (ibid.). This organisation is often described as a 'social movement,' although, according to what has been discussed so far, this would be up for debate. In any case, it aims to help 'members find inspiration and think in new directions, bringing together potential collaborators for projects and new ventures' (ibid.). For example, Hacks/Hackers Africa, part of the wider global movement,[271] works in alliance with other organisations, including the ICFJ[272] and the World Bank,[273] as 'a match-maker, with chapters bringing together technologists (who are seeking content) with content-creators (who are looking for technological platforms or solutions) and data liberators (who curate data of public interest)' (Hacks/Hackers Africa 2016).

6.2.4 The data tools developers

The main areas of development of digital tools for non-profits are to be found in the areas of a) crowdfunding, social fund-raising and crowdsourcing; b) communication, networking,

---

[269] Elaboration by the author based on participant observation. I have taken part in two of these exercises as leader in two different projects, one with ODI on climate finance in 2014, and another with University of Deusto precisely to visualise data skills transferers in 2015.

[270] See hackshackers.com/ [accessed on April 3, 2016].

[271] Code for Africa provides each African chapter with 'a seed grant to kick-start activities, followed by a two year subsidy to underwrite meetings, and a pan-African coordinator to help chapters, secure international speakers and provide fundraising support' (ibid.).

[272] The ICFJ helps the organisation 'network with chapters elsewhere in the Global South, including synergies with Hacks/Hackers chapters in Latin America, Pakistan, India and the Middle East' (ibid.).

[273] The World Bank 'co-funds our d|Bootcamps and other training initiatives, helps underwrite chapter meetups and helps develop resources and tools' (ibid.).

collaboration and co-working; c) media monitoring, including social media; d) metrics for actionable analytics; and d) mobile apps and mobile activism (SocialBrite 2016). In a subordinate position, tools to enhance privacy and protection can be found as well, responding to concern expressed before for example in connection with digital humanitarianism (i.e. Tor).[274] Some of the abovementioned organisations do produce data tools for social causes, but specific data technology tools for non-profits are probably at the rear end of what is on offer. There are new businesses, such as SocialBrite[275] and Morlan, among others, that are specialised in offering this kind of tools and services to NGOs and other social organisations. For example, enthused by the open government movement, there is a whole sector being developed in the US of businesses providing data tools and analysis capacity for the state administrations and the businesses that work with them, including Govini[276] and OpenGov,[277] among many others. However, these organisations are for-profit entities and do not work with non-profits. Vizzuality[278] is a private company that creates tools and applications 'with a lasting benefit to society and the environment' (2016). This company has developed Global Fishing Watch[279] for the campaigning organisation Oceana,[280] in collaboration with Skytruth[281] and Google, a big data technology platform that leverages satellite data to create a global view of commercial fishing. Some

[274] Tor is a 'free software and an open network that helps you defend against network surveillance, enhancing personal freedom and privacy and protecting confidential activities' (ibid.).
[275] See socialbrite.org/ [accessed on April 2, 2016].
[276] See govini.com/ [accessed on April 6, 2016].
[277] See opengov.com/ [accessed on April 6, 2016].
[278] See vizzuality.com/ [accessed on April 2, 2016]. The original CartoDB is actually a spin-off of Vizzuality, based in New York and Madrid.
[279] See globalfishingwatch.org/ [accessed on August 2, 2016]. Although launched in 2014, this interactive web tool was still in prototype stage at the time of writing, and only contained some 40,000 fishing vessels, severely limiting its usefulness. In comparison, the IUU fishing project developed at ODI, in which I have participated, relied on the most comprehensive fishing vessels database in the world, with more than 820,000 fishing units on it, and terabytes of data. This example demonstrates that putting in place big data projects for development issues is not such an easy task. The ODI project took more than two years to launch.
[280] See oceana.org/ [accessed on August 2, 2016].
[281] See skytruth.org/ [accessed on August 2, 2016].

organisations combine *pro bono* work for worthy causes with consultancy work, among them, other tools developers, such as CARTO,[282] Kiln or Populate.[283] CARTO supported the ODI/porCausa *pro bono*, for example. Populate describes itself as an organisation dedicated to 'conceptualise, design and build products for civic engagement,' which 'enable citizens and organisations to better understand and interact with the world around us' (2016). These tools include data analysis and visualisations. Populate is behind the data analysis and design of the 'Panama Papers' investigative story.

General tool developers such as Tableau[284] provide great vehicles for data visualisation and analysis to anyone that can pay for these services, but include open, simpler versions as well. Tableau proposes a free version as well to anyone that wishes to use it. CARTO, specialised in cartographic visualisations, offers a free version of its interactive mapping tools, and also supports social causes by offering the 'pro' version and technical support to NGOs that seem worthy. Other businesses, such as Kiln, Populate and Vizzuality, do work with non-profits, as well as with media organisations (the project is a spin-off of *The Guardian*), to produce data projects, some of them of journalistic nature. And some of the organisations included in other categorisations provide data tools as well. The already mentioned Civio, which defines itself as a 'non-profit organisation that fights for a stronger democracy,' and 'works towards the creation of authentic transparency and free access to public data for every citizen and organisation' (2016). Civio develops tools that 'both reveal the civic value of data and promote transparency' (ibid.). As seen, Hacks/Hackers Africa, among others, is in the business of providing tools for projects

[282] See carto.com/ [accessed on July 11, 2016]. In fact, CARTO –previously CartoDB— is a partner in the big data illegal fishing project led by ODI that I have already mentioned.
[283] See populate.tools/ [accessed on August 2, 2016].
[284] See tableau.com/ [accessed on April 6, 2016].

too. The Statistics Division of the United Nations[285] provides data and tools to analyse development data, including visualising tools, as well as the data and metadata behind the visualisations. GapMinder also delivers tools focused on analysing and communicating on development issues.[286] Other useful tools and data repositories –although not so specialised in activism— include the World Bank's Data Visualisation Tools page,[287] and FAO's datasets, which provide data and tools for the analysis of food production and consumption around the world, for example (Food and Agriculture Organisation 2016). This is a growing field where there is a newcomer –whether a for-profit, a non-profit or an institutional organisation— almost every day.

6.3 The catalysts

In this cluster, there are organisations that specialise in making data activist projects possible. Here I have included the organisations that provide the physical means and resources, as opposed to tools and skills. As said, Code for Africa, as enabler of several activities and organisations in Africa; Open Society Foundation, Hewlett Foundation, Shuttleworth Foundation, to mention just a few, could be included here as well, since there are donors, funders or conduits of resources for data activism. Less known organisations that can be classified as catalysts include Nominet Trust,[288] among others. Events and awards have become a popular way of enabling data projects, and demonstrate the data boom. For example, the NY City's BigApps awards, where all sorts of data projects –not-for-profit and for-profit—can get some financing, is among a growing number of awards that do it. The projects emerging from these type of awards

[285] See devinfo.org/libraries/aspx/Home.aspx [accessed on April 6, 2016].
[286] See gapminder.org/ [accessed on April 6, 2016].
[287] See data.worldbank.org/products/data-visualization-tools [accessed on April 6, 2016].
[288] See nominettrust.org.uk/who-we-are/about-our-work [accessed on March 8, 2016].

are mostly for-profit, though. In the realm of the inter-governmental arena, UN Global Pulse,[289] an initiative of the UN that could be considered a catalysts as well, aims at harnessing data for public good, by accelerating 'discovery, development and scaled adoption of big data innovation for sustainable development and humanitarian action' (2016). The initiative was established based on a recognition that 'digital data offers the opportunity to gain a better understanding of changes in human well-being, and to get real-time feedback on how well policy responses are working' (ibid.) The type of projects supported by this UN programme include 'Using mobile phone data and airtime credit purchases to estimate food security,' in which the UN World Food Programme (WFP), Université Catholique de Louvain, Real Impact Analytics participate; 'Using Mobile Phone Activity For Disaster Management During Floods,' in which Digital Strategy Coordination Office of the President of Mexico, UN World Food Programme (WFP), Telefonica Research and Technical University of Madrid (UPM) participate; or 'Measuring Poverty with Machine Roof Counting,' in which Uganda Bureau of Statistics and the University of Edinburgh collaborate. This is another growing field of action for data activist organisations.

6.4 The proactive data activists and geoactivists

This cluster groups the most interesting organisations and initiatives, and therefore more space is dedicated to them here. As discussed, there are many organisations focussed on transferring data or humanities skills for social good, or facilitating funding, resources and match-making opportunities to foster data projects. However, finding actual proactive data activist projects is harder. This is possibly because, although there is a general recognition of the potential of data for social causes, development and proactive campaigning, this potential is far from being satisfied at the time of writing. Pigeonholing the actual cases of data activism is a

[289] See unglobalpulse.org/about-new [accessed on March 8, 2016].

challenging task as well. A difficulty is that some of the selected cases are organisations dedicated to other tasks concurrently, so they cannot be classified as purely data activist organisations (i.e. Civio). Meanwhile, in other instances the case included in the study is an isolated data project (i.e. 'Vidas Contadas'). Hereafter, I will take into account both organisations that practice proactive data activism, either occasionally or regularly, and data activist projects and initiatives. Those organisations that practice proactive data activism as their core endeavour and strategy will be considered, hereafter, proactive data activist organisations. What follows is an attempt to group some proactive data activist initiatives, organisations, projects and contents, establishing what unites them (i.e. the centrality of data and data infrastructures with a proactive focus, and a social purpose, mostly done in alliance with others), and what separates them. At the end of this approach to data activism, I offer a classification of cases according to a set of criteria derived from this empiric observation. It is not intended as a systematic or comprehensive effort, but as a way of producing a model of effective proactive data activism. This model will be tested afterwards in the case study.

As in geojournalism, digital and interactive maps provide a fertile area where most examples of proactive data activists and organisations can be found. Ushahidi, among other organisations, can be said to practice *proactive geo(data)activism*, geoactivism for short, in the field of digital humanitarianism, offering interactive maps visualising data for action in cases of disaster and crisis. But geoactivism is not limited to humanitarianism. Other organisations and projects that put geoactivism into practice could include 'Vidas Contadas,' 'España en llamas' and a campaign set up by the Plataforma de Afectados por la Hipoteca, known as PAH (Platform for People Affected by Mortgages), among others. De Soto calls this activity 'the art of cartography of connected multitudes' (de Soto 2014). The commonalities among most

geoactivist examples can be grouped in three areas: a) they tend to rely on collective participation to generate data contributed by users and witnesses, who are coordinated via online applications. These data are then transformed into layers of information in maps; b) they employ interactive maps for communication, coordination and activism; and c) they aspire to directly improve or solve a social problem or generate support in an emergency. Of these three characteristics, only the second and third are essential: by definition, geoactivism will depend on maps in all cases and will entail a campaign to be activism. However, proactive data projects can rely on non-crowdsourced data as well. The campaign focussed on illegal, unreported and unregulated (IUU) fishing mentioned before is an example of a data project without crowdsourced data, as it relies on data extracted for a myriad of registries and sources that make the biggest database on fishing vessels on earth. The already mentioned Kiln – a for-profit company providing visualisations of all types— mixes journalism, research and campaigning in its visualisations and maps in projects such as 'Keep it on the ground.' It works with international organisations' campaigns as well. For example, the geoactivist project 'Death on the roads,' together with the World Health Organisation (WHO),[290] shows data on an interactive map to support the campaign against accidents involving cars, pedestrians and cyclists. The data are not crowdsourced, but gathered in collaboration with governments.[291]

The specific use of geoactivism in cases of crises and disasters –because of its potential to immediately improve the situation on the ground and aid humanitarian assistance— has given rise to what is being called 'crisis mapping.' That is, when geoactivism is applied to crises and used to support humanitarian efforts, then it can be regarded as 'crisis mapping.' Interviewed for

[290] See who.int/violence_injury_prevention/road_traffic/death-on-the-roads/en/.
[291] WHO's data repositories are brought together by WHO and partners, in consultation with member states (World Health Organisation 2016).

this dissertation, Mitchell differentiates between 'crisis mapping' and 'activist mapping.'[292] Crisis mapping is 'the act of receiving reports during a crisis or an emergency about a request or location of services, and plotting them on a map.' Activist mapping is different. 'Many times it is not that urgent, although it can be if victims need help. But it is usually a longer term activity, more focused on gathering bodies of evidence that can be used for lobbying and advocacy,' he says. This differentiation becomes handy when examining the projects that use the Ushahidi platform. Considered under this lens, Ushahidi's deployments are used in both areas. However, I do not particularly think the term does justice to geoactivism because it confers an inert character to it. That is, it may be inferred that crisis mapping it is just about charting a catastrophe, not about triggering action, whereas 'geoactivism' suggests a dynamic and proactive engagement with an adverse situation. Both activist mapping and crisis mapping can be considered geoactivism.

In geoactivist projects, dedicated software is often used, for example, based on technology developed by OpenStreetMap,[293] Ushahidi or Sahana Software Foundation.[294] In order to power

---

[292] 'It is called crisis mapping because actually is quite a specific set of abilities that are needed in crises. Obviously, when a crisis occur, the first 24-48 hours are very important. So you want people trained to deploy all the right information into the report. If there is translation needed in a foreign country, you need to be able reroute the information to translators and get it back, so people can act on it. So it is the act of managing crisis reports and mapping them, and channelling the information to responders on the round, and being able to do that very quickly.'

[293] See openstreetmap.org/#map=5/51.500/-0.100 [accessed on May 4, 2016]. The site is supported by the OpenStreetMap Foundation, a non-profit organisation registered in England and Wales. These crowdsourced data are then made available under the Open Database Licence. Thousands of volunteers in the OpenStreetMap community are using imagery from DigitalGlobe to populate their maps (Bullock 2015). See also digitalglobe.com/ [accessed on August 16, 2016]. OpenStreetMap has been used for crisis response, mainly in locations that are inadequately mapped. After the 2010 earthquake in Haiti, teams of digital volunteers used OpenStreetMap and satellite imagery to quickly improve the maps of existing infrastructure and record the damage from the earthquake. Out of this experience, the Humanitarian OpenStreetMap Team (HOT) –a global network of volunteers— emerged.

[294] See sahanafoundation.org/ [accessed on May 3, 2016]. 'Sahana software was originally developed by members of the Sri Lankan IT community who wanted to find a way to apply their talents towards helping their country recover in the immediate aftermath of the 2004 Indian Ocean earthquake and

early warning systems for rapid response to humanitarian emergencies, crisis mapping applies different tools and methods, including participatory maps and crowdsourced data, geospatial platforms, mobile applications, aerial and satellite imagery, visualisations, and computational and statistical models. Crisis mappers are often expert volunteers themselves.

However, what can be understood as a *map*? The question is very pertinent, as I started this dissertation precisely professing my supposition that big data infrastructures allow us to construct practical maps to navigate and understand complexity in post-modern times. No map can be as perfect as to build a complete image of the reality it is designed to represent, as the cartographers in Borges's tale yearned for.[295] Big data cannot be a flawless mapping system, since –no matter how big or complex— they can hide gaps, biases and errors. And a wide variety of cartographic approaches, as well as political and social perspectives on maps, exist around this matter, which has been revived by the exuberance of options that the new interactive data visualisations have brought to the field of cartography today. An example is worldmapper.org, an organisation specialised in cartograms or maps in which some thematic mapping variable – such as population, agricultural production or Gross National Product – is replaced for land area, distorting the geometry of the map in order to convey the information of the chosen variable in a

---

tsunami. The word *Sahana* means "relief" in Sinhalese, one of the national languages of Sri Lanka' (Sahana 2014). Sahana is supported by aidIQ, Geeks Without Bounds, Google, IBM, ISCRAM, World Food Programme, among others.

[295] In Jorge Luis Borges's short story *On Exactitude in Science* (*Del rigor en la ciencia*), the author tells about an empire where, seeking faithfulness, maps become as big as the territory they represent. It is a metaphor of the futility of seeking total perfection in science. Later generations 'deemed a map of such enormousness impractical... In the western deserts, tattered fragments of the map are still to be found, sheltering an occasional beast or beggar' (Borges 2007). 'En aquel Imperio, el Arte de la Cartografía logró tal Perfección que el Mapa de una sola Provincia ocupaba toda una Ciudad, y el Mapa del Imperio, toda una Provincia. Con el tiempo, estos Mapas Desmesurados no satisficieron y los Colegios de Cartógrafos levantaron un Mapa del Imperio, que tenía el Tamaño del Imperio y coincidía puntualmente con él. Menos Adictas al Estudio de la Cartografía, las Generaciones Siguientes entendieron que ese dilatado Mapa era Inútil y no sin Impiedad lo entregaron a las Inclemencias del Sol y los Inviernos. En los Desiertos del Oeste perduran despedazadas Ruinas del Mapa, habitadas por Animales y por Mendigos; en todo el País no hay otra reliquia de las Disciplinas Geográficas.'

more compelling manner (Worldmapper 2013). Mapping has always been a highly political activity, the monopoly of states. That is why *political maps* exist as well, as opposed to physical maps, which are no simple matter either, given the popular Mercator versus other projections, such as Gall-Peters, dilemma. And that is why 'counter-mapping' movements have emerged in resistance to governmental initiatives that *remap* land to hand over pieces of the commons to companies, for example, in Indonesia (Radjawali and Pye 2015). 'Challenging state power over maps and its categorisation of land uses by counter-mapping indigenous and local claims to territory has developed into an important movement in Indonesia' (ibid., 4). As the authors of the study correctly observe, 'mapping needs to be understood as a political process rather than a merely technical tool… It is important to always ask who produces the maps, how people can access the maps and how the maps can be used for emancipatory purposes' (ibid.). Map Kivera, for example, is another initiative at smaller scale to fill the blank spot on the map that the neighbourhood of Kivera, Nairobi, was until November 2009, when a group of Kiveransu created the first free and open digital map of their community, which became interactive community information project.[296] Meanwhile OpenStreet is another example. Counter-mapping is therefore a political process as well as a reaction to mainstream, conventional cartographic interpretations. This counter-mapping movement is even testing 'grassroots UAVs' (Unmanned Aerial Vehicles or drones) to generate data, and strengthen their datasets and their political actions (Meier 2015b), a real case of a data activism producing its own data.

Box 2: A definition of a 'map'

Definitions of the concept of a *map* always conceal questions of misrepresentation

[296] See mapkibera.org/ [accessed on November 10, 2016].

and social inequality as well. The Gall-Peters projection, for example, named after James Gall and Arno Peters, came to reconfigure world maps in the 20th century, and generated great controversy about the political implications of map design. The Gall-Peters projection aspired to substitute the conventional Mercator projection, a creation of geographer Gerardus Mercator in the 16th century, which had become the standard map projection for nautical purposes because of its ability to show lines of constant course. The Mercator projection distorts considerably the real size of objects so, for example, landmasses such as Greenland seem much larger than they truly are in comparison with land masses near the equator, such as Central Africa, leading to discussions about the misrepresentation of 'Southern' nations. There are many other projections.[297]

Google Maps, among many other online map providers such as OpenStreetMap, Mapbox or Mapquest, uses the Mercator projection (or more precisely, a Web Mercator formula). CARTO, however, encourages people to 'free' their maps from Web Mercator (Akella 2015). 'What projection you choose depends on your map's extent, the type of data you are mapping, and as with all maps, the story you want to tell. Well, get excited because with a few lines of SQL[298] in CARTO, you can free your maps from Web Mercator!' (ibid.). This matter is so important that some organisations campaign in favour of against concrete models, such as the Organisation of Cartographers for Social Equality in the US, which supports the Gall-Peters projection.

In order to define what a map is in the age of technopolitics, de Soto, who studies the cartography of the 15M movement in Spain, refers to the conception of *rhizome*, coined by Deleuze and Guattari (concept taken from botany). Rhizomes conform to several principles: The principle of connection and heterogeneity, since 'any point of a rhizome can be connected to any other, and must be' (1987, 7); the principle of multiplicity, since 'it is only when the multiple is effectively treated as a substantive, "multiplicity," that it ceases to have any relation to the One as subject or object, natural or spiritual reality, image and world' (ibid., 8); the principle of 'a signifying rupture,' since 'a rhizome may be broken, but it will start up again on one of its old lines, or on new lines' (ibid., 9); and the principles of cartography and decalcomania, since 'a rhizome is not amenable to any structural or generative model; it is a "map and not a tracing"' (ibid., 12).

Based on these concepts, de Soto understands a map as 'action' instead of representation, and cartography as 'the identification of new components, the creation of new relationships and territories, new machines' (2014, 361).

[297] See 'Comparing Map Projections,' a mashup of map projection distortions [bl.ocks.org/syntagmatic/ba569633d51ebec6ec6e, accessed on November 10, 2016].
[298] Structured Query Language or SQL is used to communicate with a database.

> One conclusion is that in any representation, interpretation or illustration of the world there are many factors to be taken into account, which include social and political constructions and biases, power relations, as well as practicalities. Whatever the mapping model, maps in the context of data activism will be understood henceforth as not limited to the illustrations of a bi-dimensional geographic territory, but as a knowledge and production tool, which can represent complex situations, relationships and heterogeneous objects, social, political or technological processes, events and places, and mutable interactions and networks.[299]

'Vidas contadas' is an example of an initiative of geoactivism. It started as a data project during a *hackaton*, as hacker Labarga recalls during a workshop organised by Medialab-Prado (2015). The purpose of the project is recovering the 'historical memory'[300] of the times of the Spanish Civil War and the subsequent dictatorship, during which many people disappeared, were killed, or subjugated and discriminated against. It does it through the systematic integration and harmonisation of all the data, information, testimonies and documents around the matter, and making them interactive, collaborative and accessible using open source resources. The project agglutinates information and documents from listings, websites, documentation centres and registries, and physical memorials. It tries to bridge different repositories as well, connecting them and making them searchable in the four Spanish official languages (Spanish, Basque, Galician and Catalan). 'Much of this work, gathered after a monumental, collective and years-long effort, serves as a basis for (data) visualisations and interactions' (Vidas Contadas 2013). I have included this as a case of geoactivism because the project includes, among many other

[299] Luiselli also talks about imagined 'maputopias' versus the Custom Maps of Migrant Mortality that geolocate the deaths of Mexican migrants who died trying to cross the border with the US (2017).
[300] The Historical Memory Law passed in 2007, and recognises the victims on both sides of the Spanish Civil War, and gives rights to compensation to the victims and the descendants of victims of the Civil War and the dictatorship.

content and analysis (i.e. from interviews to datasets), interactive maps showing where human remains and mass graves linked with the Civil War and Franco's repression have been found, and these maps are very central to the project. As is central the activism of their frontrunners, who demand that the Historic Memory Law, based on a bill proposed by Spanish Premier Jose Luis Rodríguez Zapatero's socialist party when it was in power, is fully budgeted and implemented. The conservative Popular Party, who succeeded the socialists in the government and had voted against passage of the law arguing that it would reopen old wounds, has been reluctant to put the law into effect ever since. 'España en llamas' (Spain on fire), another project facilitated by Civio and crowdfunded via Goteo.org,[301] shows where and when fires happen, quantifies the loss of life and how much area has been burnt, and estimates the economic loses, the resources employed to put them out and the costs of the operations, and reveals whether they were deliberate or not. The web site provides also analysis of the numbers, bringing attention, for example, to the fact that 91% of the fires in Spain are premeditated or due to human error.[302] Combining visualisations and analysis, the underlying objective is precisely to fight against deliberate fires. For example, the project publishes analyses explaining how, soon after launching 'España en llamas,' its initiator s set out to find all the sentences related to forest fires in recent years, to realise that there was no unified record compiling or coordinating them, making it impossible to analyse them (García Rey and Garrido 2013). After nine months of work, Civio developed 'the most comprehensive database of both official or unofficial court sentences, data and documents around forest fires in Spain' (ibid.). The authorship of 420 fires between 1996 and 2012 was attributed to 393 people. They also discovered a culture of secrecy, which combined with the lack of coordination or dialogue among the authorities in charge of

[301] See en.goteo.org/ [accessed on July 11, 2016].
[302] At the time of writing, the last entry in this project's website was from July of 2015 [accessed on July 11, 2016].

dealing with this massive problem, make this matter opaque to scrutiny and only manageable in a case by case basis. PAH employed the Ushahidi platform for its 'Stop Desahucios' (Stop Evictions) campaign, tackling the same issue as the Anti-Eviction Mapping Project[303] does in the US. 'Stop Desahucios' visualised crowdsourced information about evictions taking place in Spain, and issued warnings to connected users. At the height of the financial crisis in 2012, Spain endured an average of 517 evictions per day (Diario Jurídico 2012).[304] The 'Stop Desahucios' campaign was created by hacktivists at tomalaplaza.net,[305] in collaboration with the PAH, who developed a collaborative map evictions alert so that anyone who wanted to contribute information about ongoing evictions could do it easily, and operations resisting dislodgments or asking for help against evictions could be organised quickly on real time. By law, there must be one whole day between the eviction is published and executed; meanwhile 'Stop Desahucios' would issue an 'eviction alert' and the network of solidarity would try to organise resistance and prevent it. While I consider this a proactive data activist project, it has a strong reactive and confrontational element to it, although in the physical offline world of atoms. As said, there are many technological platforms that cooperate in geoactivist projects. CARTO is the preferred option for many geoactivist initiatives. One example is the already mentioned Anti-Eviction Mapping Project, which documents 'dispossession' in San Francisco Bay area (2016). They include maps, testimonies, pictures and narratives on the issue, gathered in part via crowdsourcing. For example, there is a map visualising the locations of evictions related to the

[303] See antievictionmap.com/ [accessed on April 20, 2016].

[304] Millions of families acquired their homes during the years of the property bubble, however, when the 2007 crisis hit the country, they were left very exposed. Unlike most of the US, Spain does not recognise mortgage loans as *nonrecourse debt*. Since during the crisis property prices dropped enough for most foreclosures to only account for 60% of the loan, those evicted continued being responsible for large debts for property they no longer own.

[305] Literally 'take the plaza.' The last entry on this web site is from May 20, 2015. This movement dedicated to the coordination of Indignad@s protests in Spain and around the globe is associated with DemocraciaRealYA, 15M and Occupy.

use of a certain state law (the Ellis Act), which gives landlords the power to evict residents and has been linked to growing volumes of more expensive deals. San Francisco's housing market has been in the news with many residents outraged about the fast mounting costs of living and gentrification that is causing great swings in demographics in the City. The initiative also organises mobilisations against evictions.

Abridging, digital interactive maps are extremely important for proactive data activist projects, since most of the cases examined are examples of map-based geoactivism. That is why so much attention has been paid here to the use of cartographic representations for data activism. In the next sections, I examine more in-depth the preferred platforms for data activism and I resort to the origins of data in an attempt to classify individual cases in this cluster.

6.4.1 The mapping platforms

Given the abundance of proactive data activist projects employing digital maps and the collective generation of relevant and usable content, platforms that sustain both cartographic representation and crowdsourcing have flourished and deserve some attention. What follows is a quick overview of some of the options available and examples of how they are being used.

CARTO is a Software-as-a-Service (SaaS)[306] cloud computing platform that provides global information system (known as GIS) and web mapping tools. Its software was first released in Beta in 2011, and officially debuted as a final release in 2012. This Spanish start-up

[306] Software as a service (SaaS) is a software licensing, delivery and distribution model in which applications are hosted by a provider and made available to paying subscribers over a network, typically the internet. It is sometimes referred to as 'on-demand software' (Search Cloud Computing 2016). SaaS has become a common delivery model for many business applications, since it can reduce IT support costs by outsourcing hardware and software maintenance and support to the SaaS provider. In case of CARTO, it offers a simpler version to non-paying users and collaborates with many non-profit organisations, providing them with the 'pro' version and technical support, as in the example examined here. This way CARTO (previously CartoDB) users can use the company's free platform or generate their own version of its open source software. CARTO is offered as freemium service, where accounts are free up to a certain size. For larger accounts, a subscription fee is applied.

raised US$7 million from a consortium of investors in 2014, and in 2015, it received another $23 million in a second round of financing. Since its creation, CARTO has provided a platform to organisations and companies, such as the US National Aeronautics and Space Administration (NASA), Nokia, Google Trends, Deloitte, Wall Street Journal, BBVA and Twitter. CARTO is also a partner in the proactive data activist project against IUU fishing mentioned before, together with the ODI, FishSpektrum and porCausa (Daniels et al. 2016, 6-9). This project was conceived from the start as a full-blown proactive data activist effort, combining many of the elements of a campaign, including research, data analysis and visualisations, geoactivism and interactive maps, media outreach plans, alliances and direct lobbying. The central theme is IUU fishing, a great contributor to overfishing, which has reached catastrophic levels: many major fish stocks are in decline, and some species are being pushed towards extinction. As much as one fifth of the world's fisheries' catch may originate from IUU activity, linking consumers in Europe and the US with a practice that is fuelling a global tragedy of the commons – a tragedy that is leading to the over-exploitation of a common resource. The project[307] is based on the western African coast, which is at the epicentre of the tragedy. The region's coastal waters include some of the world's most abundant fishing grounds, creating a magnet for commercial vessels that supply rapidly growing markets in Europe and Asia. The profits generated are substantial. However, as highlighted by the former UN Secretary-General Kofi Annan in the 2014 Africa Progress Panel report *Grain, Fish, Money*, the over-exploitation of western Africa's fishery resources has produced devastating social, economic and human consequences (2015, 14). Recent years have seen a renewal of international efforts to combat over-fishing and IUU activities, and strengthened regulatory frameworks have been put in place for monitoring and

[307] See odi.org/publications/10459-western-africas-missing-fish-impacts-illegal-unreported-and-unregulated-fishing-and-under-reporting [accessed on July 13, 2016].

reporting through Port State Measures.[308] These moves are encouraging – but they are failing to tackle IUU fishing practices. Far too many governments in Europe and in emerging markets adopt encouraging principles at international meetings, while enacting policies at home dictated by powerful fishing lobbies. The report that is the basis for this campaign identifies two practices at the heart of the disjuncture between sustainable fishing principles and reality: The first practice involves 'reefers' – vessels that collect catch from fishing fleets at sea through transhipment. Analysing the FishSpektrum Krakken® UVI database – the world's largest fishing vessel tracking resource—, the authors tracked reefers operating in western African coastal waters to discover that most of the fish leaves the region on refrigerated containers, which are subject to less stringent reporting requirements. Tracking signals from most of the identified vessels point towards another suspicious activity: the signals are consistent with widespread and systematic transhipment activity occurring in some cases in Exclusive Economic Zones (EEZs) of two countries – Senegal and Côte d'Ivoire – where transhipment is prohibited by law. In other cases, the transhipment activity appears to be unauthorised or inadequately monitored. The impact of the report was immediate: within days more than 150 international and regional media outlets –including *The Guardian*, *CNN*, *BBC* and *El País*— had covered the story, FAO felt compelled to release a statement supporting some of the recommendations of the campaign and referring to the report's finding, Gambia temporarily banned all foreign operations in its EEZ, and Namibia signed the FAO's Agreement on Port State Measures to Prevent, Deter and

[308] Port State Measures (PSM) are requirements established or interventions undertaken by port states which a foreign fishing vessel must comply with or is subjected to as a condition for use of ports within the port state. National PSM would typically include requirements related to prior notification of port entry, use of designated ports, restrictions on port entry and landing or transshipment of fish, restrictions on supplies and services, documentation requirements and port inspections, as well as related measures, such as IUU vessel listing, trade-related measures and sanctions.

Eliminate IUU Fishing, as the campaign was demanding. The *prima facie* evidence put forward by the visualisations was powerful enough to start changing things.[309]

Tomnod (in Mongolian 'big eye') is an initiative owned by the satellite company DigitalGlobe employing crowdsourcing to identify objects and places in satellite images. It utilises CARTO interactive cartography, among other options such as Google Earth, in order to 'solve real-world problems' (Tomnod 2016). Tomnod defines itself as 'a team of volunteers' who work together to 'identify important objects and interesting places in satellite images' (ibid.). It was originally a research project of the University of California, San Diego in 2010; three years later, Tomnod was acquired by DigitalGlobe. Tomnod uses online map interfaces that engage people to view and tag a small section of a large area on the planet. In 2011 Tomnod cooperated with the UN High Commissioner for Refugees (UNHCR), together with the Standby Task Force,[310] to locate refugee camps in Somalia. Users were asked to use satellite images to tally the shelters of refugees. Other Tomnod projects include mapping damage after Typhoon Haiyan in 2013 in Southeast Asia, and the earthquake in Nepal in 2015, which killed over 8,000 people and injured more than 21,000.

OpenStreetMap, known as OSM, is Borges's cartographers dream: a collaborative initiative to create a free editable map of the world with as much detail as its contributors can provide from the ground. Map data are collected by volunteers implementing methodical ground surveys using instruments such as a Global Positioning System (GPS) unit, a digital camera and

[309] The FAO's Agreement, for example, had been drafted in 2009, and up to 2016, there had not been enough signatories to enforce it.

[310] Standby Task Force is a 'global network of trained and experienced volunteers working together online... Standby Task Force is a not for profit organisation incorporated in the United States of America and have a number of specialist teams to organise our work... We were set up in 2010 and have been active in many natural disasters since then as well as supporting humanitarian agencies with election monitoring and other projects' (Standby Task Force 2016).

a voice recorder. The data are then entered into the OSM database. The advent of inexpensive portable satellite navigation devices, and the availability of aerial photography and other data from commercial and government sources make this possible. Created by Steve Coast in the UK in 2004, it was inspired by the success of Wikipedia, and in the face of the prevalence of proprietary map data. Since then, it has grown to over 2 million registered users. Rather than the bi-dimensional geographic map, the data generated by the OSM project are considered its primary output. OSM free map data has been used for humanitarian aid and economic development. A Humanitarian OpenStreetMap Team (which they call HOT) was created in the immediate aftermath of the Haiti earthquake, as a U.S. non-profit, which was then registered as a charitable organisation in 2013 (Open Street Map 2016).

Tomrod, Ushahidi, CARTO, OSM and Sahana are proof of the exiting exuberance in digital interactive, crowdsourced mapping that is applied to data projects focussed on development and humanitarian issues. There are many more examples outside of geoactivism catering for the needs of tourists, urban dwellers, public service seekers, head and job hunters, among others. However, they fall outside the scope of this dissertation.

6.4.2 Data activism by the origin of data

Another interesting angle that can be used to classify data activism endeavours is to look at where they get their data from, since most organisations tend to specialise in one or two data extraction routines. For example, ICIJ relies on public and whistle-blowers' data, while ODI resorts to public and datafied research. As said, big data are mostly generated and amassed by governments and big companies, such as Google and Amazon. However, some organisations get their hands on not easily available data or generate their own data (normally *small data*), and the origin of those data can be very varied. From this perspective, five groups of data organisations

and projects can be identified, from the ones that devote less effort to obtaining the data to the ones that require the most effort,[311] not considering how complex or arduous is to analyse, visualise or communicate the data afterwards: a) The organisations that become recipients of data via whistle-blowers (i.e. WikiLeaks or ICIJ's 'Panama Papers' investigative story –a case of data journalism with a touch of advocacy—, which is based on millions of documents transferred by a whistle-blower). b) The organisations that simply produce new analysis and obtain novel insights from available, but unrelated and unexplored, datasets, producing secondary research. Examples of this variety include the anti-fossil fuels subsidy campaign, or Kiln's 'Death on the roads' project, relying on WHO's ability to collect and analyse data from governments. c) The organisations that generate the means to crowdsource data, which are contributed by people; for example, the different deployments that use the Ushahidi platform. d) The ones that appropriate data (i.e. via MobileMinner). e) The organisations that conduct primary research whose findings can be datafied and analysed, such as surveys, questionnaires, interviews with individuals or groups, and focus groups, or deploy data-collecting sensors, drones and other devises (i.e. counter-mapping activists in Indonesia or the 'Vidas Contadas' project, which depends on disconnected public repositories, but also on new field research leading to mass graves being discovered thorough witnesses' accounts). The included at the end of this section shows the types of activism that emerge looking at where these organisations and projects get their data from. What follows is a brief review of some of these groupings. I am not including all the categories to avoid repetition or dwelling on already explored areas, such as how whistle-blower data have been utilised for enhanced journalism and policy change.

[311] This is of course an arbitrary judgement.

The ODI/OCI fossil fuels subsidies campaign is an example of datafied secondary research.[312] OCI has already been mentioned as an organisation practicing data activism, by means of a TAN, in the context of the fight against climate change and fossil fuels. It defines itself as 'a research, communication and advocacy organisation focused on exposing the true costs of fossil fuels and facilitating the coming transition towards clean energy'[313] (Oil Change International 2016a). This organisation depends totally on data and data analysis for its campaigns. The research in the examined campaign is secondary, though, as it is based on public or published reports, including data released by G20 countries and analysis done by groups such as the World Bank, OECD, Greenpeace, International Monetary Fund, Australian Petroleum Production & Exploration Association, Banco de Inversión y Comercio Exterior, British Petroleum, Climate Action Network (CAN)-International, Brazilian National Development Bank, China Development Bank, Compagnie Française d'Assurance pour le Commerce Extérieur, Governo Federal (Brazil), Climate Policy Initiative, Carbon Tracker Initiative, Deloitte, and media organisations such as *The Economist*, *BBC*, *Bloomberg News*, *Financial Times* and *Reuters*, among others. The accomplishment of this project is putting together, for the first time, all the available information, analysis and data on fossil fuel subsidies and communicating the findings in an impactful manner, using data visualisations. In other words, the great merit of projects such as the ODI-OCI subsidies initiative resides in surfacing data that

[312] ODI, with a more diverse portfolio, follows in OCI's steps, but from a different starting point. OCI is a campaigning organisation that bases it positions on data and research. ODI was founded in 1960 as an independent research centre specialised on development and humanitarian issues. Since then, it has been become a more *campaigny* think tank[312], lobbying with governments, issuing recommendations and stating positions on some key issues, and 'locking together high-quality applied research, practical policy advice and policy-focused dissemination and debate' (Overseas Development Institute 2016). ODI conducts primary research on development, environmental and humanitarian issues, but depends on public data as well for its analyses. In fact, both organisations, OCI and ODI, depend totally on data – whether obtained as a result of primary or secondary research, or public data — for their analysis and communication strategies[312]. That is, this type of organisations relies on several methods to mine data.

[313] See priceofoil.org/ [accessed on April 30, 2016].

may have been hidden, integrating unrelated datasets, asking novel questions and finding new relevant information buried in existing archives.

Relying on crowds to gather data is another method to generate data. The key is the concept of *crowds*. In fact, Meier calls the process of crowdsourcing 'human computing,' as opposed to machine computing or artificial intelligence (2013a; 2015c),[314] echoing an epoch 'when computers were women' (Light 1999),[315] as if computing were returning to its origins the same way journalism is. Aguilar, former head of the Coordination of Humanitarian Affairs in Colombia, interviewed for this dissertation, says that crowdsourcing is a way to exercise citizenship, a sort of grassroots democracy in action. Crowdsourcing can be an informational resource for development, crisis response and post-conflict recovery by allowing citizen participation in governance processes and enhancing democracy through openness and empowerment of citizens (Bott, Gigler and Young 2012, 47). There are three main advantages, compared to conventional relief methods, of leveraging crowdsourced data for disaster assistance (Gao, Barbier and Goolsby 2011, 11). The first one is the immediacy of collection of data; a second advantage comes from the macro perspective that emerges, which allows humanitarian agencies to envisage the whole picture; and third advantage comes from a micro perspective, since witness can include information sent from any platform and device, relief organisations can 'accurately locate specific requests for help' (ibid.). A case of a civil society organisation mixing two approaches to generate data is Bellingcat –a website that offers free video, maps and picture

[314] Meier does this when discussing the challenges of aerial imagery analysis and the virtues of utilising both strategies to deal with vast amounts of pictures obtained by a single drone flight (ibid.).
[315] Almost two hundred women, both civilian and military, worked as human 'computers' with Eckert and Mauchly to develop US's first electronic computer, ENIAC, to automate ballistics calculations during World War II. Six of them were designated to programme a machine that, paradoxically, would take their name and substitute them.

investigations to journalists and activists.[316] It generates primary information through the analysis of crowdsourced images. In 2014, this organisation became famous by suggesting that it was the Russian military who stroke the Malaysia Airlines Flight 17 down in Ukraine, analysing the pictures and videos published by people in social media (Bellingcat 2014). Bellingcat uses a different type of crowdsourcing, though. Based on a combination of expert eye and computerised analysis of pictures, maps, videos and other digitalised graphic objects uploaded (crowdsourced) in social media and websites, as well as Google Earth and geolocalisation services and software, Bellingcat has examined and provided evidence of the Syrian war and the 2014-2015 Russian military intervention in Ukraine, among many other issues.

Crowdsourced data have been classified here in a third place in terms of mining difficulty. The challenge in crowdsourcing data does not come only from its technical sophistication but from the convening power of the project in which is being used –that is, the capacity of the project to generate a motivated crowd. As it will be seen in the case study, Ushahidi's deployments have accumulated a number of failures due to the lack of a sound strategy and absence of crowdsourcers submitting reports, transforming maps into 'dead maps,' empty shells 'littering cyberspace' (Vota 2012).

As noted before, data and metadata exchanged on social networks are managed and seized by service providers, who may or may not share them with their originators. In some cases, they do share some data. For example, Twitter offers a lot of information that can be analysed on its API. 'Public user profile information and public tweets are immediately delivered via SMS and the Twitter APIs to our partners and other third parties, including search engines, developers, and

[316] Created by Eliot Higgings –also known as Brown Moses— in 2012, it adapts open source intelligence (OSINT) for investigative citizen journalism by analysing and tracking weapons appearing in pictures and videos shared via social media networks.

publishers that integrate Twitter content into their services, and institutions such as universities and public health agencies that analyse the information for trends and insights' (Twitter 2016). Likewise, anyone can use the twitter engine to search for a particular hashtag to find information on who talks about an issue, with whom and what do they talk about, and then employ open software (i.e. OpenRefine)[317] to clean the data and turn them into a usable table that can be analysed. But what happens with the data that companies do not what to share because it confers them a commercial advantage? Some social projects promote the appropriation of data using tools such as MobileMinner,[318] which helps harvesting one's own mobile data. Pybus, Cote and Blanke talk about the use of MobileMinner, which they created to gain access to one's own data, and highlight the 'data-making possibilities' that exist if users could either unearth or snatch 'what data controllers, such as Facebook, monetise' (Pybus, Cote and Blanke 2015, 1). These authors wonder about the 'profoundly asymmetrical, political economic dimensions of the production and circulations of data,' which reduce us to 'ever more traceable objects of surveillance' (ibid.), to conclude that 'there is tremendous untapped potential in the general intellect and technical practice… in the figure of the data generator who wants to be in control and, more importantly, seeks to understand the data they collectively generate' (ibid., 8). The ability to access to one's own data 'not only augments the agency of the individual but of the collective user' (ibid., 1). This benefit –that is, the McLuhanian agency augmentation of the individual and collective user via technological tools— does not happen only with the appropriation of data, but with any acquisition and employment of useful data.

[317] See openrefine.org/ [accessed on May 23, 2016].
[318] See kingsbsd.github.io/MobileMiner/ [accessed on May 8, 2016]. MobileMiner 'allows you to track the trail of data your Android device leaves behind you. It was written as part of a Big Social Data project at King's College London's Centre for e-Research (see big-social-data.net/)' (GitHub 2016).

The appropriation of data has generated a controversy and merits some attention here. In the context of medical data, Béranger warns against the risks of making 'c*ertain data'* universally available and using them without understanding their 'primary use context,' which may include distortions in the interpretation process (2016, 39). 'Data appropriation by the largest number of individuals quickly, indeed, collides with the difficult question of data culture' (ibid.). Nevertheless, what Pybus, Cote and Blanke defend stops at a much more basic level: these authors demand the appropriation of *harmless* data and metadata produced during our public social networking exchanges. This approach could be extended to one's own medical records, which are not always available to patients in some countries.[319] In this regard, Tennison differentiates between open data and data sharing (2014). Data sharing 'is providing restricted data to restricted organisations or individuals' (ibid.). Sometimes shared data are restricted because they provide a revenue stream and they are only available to people who will pay for them; or because they are sensitive in some way, either because they are personal or because of security issues (ibid.). Tennison defends that governments 'move away from restricted data sharing and towards open data' (ibid.). 'There is lots of data within the public sector that should be open. There is also lots of data that the public sector holds that should only be shared very carefully, or not at all. We must take care not to confuse the two' (ibid.). Obviously Pybus, Cote and Blanke do not talk about non- anonymised medical records or sensitive data when they defend the appropriation of data. Along the same lines, Kennedy, Poell, and van Dijck do argue for a two-way flow of data. 'Datafication should not only be understood as the process of collecting and analysing data about internet users (by companies), but also as feeding such data back to users, enabling them to orient themselves in the world' (2015, 1). Although these authors

[319] For example, in the US medical records can be requested, but there are stark differences among states in the period within which 'a state requires a medical record to be provided to a patient once requested' (Health Info Law 2013).

speak about companies ‘feeding’ user-generated data, and not about users appropriating their own data, it is worth noting that they contend that data are ‘also generated, collected and analysed by alternative actors, enhancing rather than undermining the agency of the public’ (ibid.). According to them, ‘the question of (citizen) agency should be central to our engagement with data,’ since ‘new power’ resides in the ‘alternative forms of datafication’ (ibid.).

What data should be private or public is an ongoing and contentious debate. While private and sensitive data should probably remain private, public data should be open by default. However, there is a growing sector dedicated to developing software and strategies to protect private and businesses data against cyber fraud and cyber theft, while big business and big government are still to embrace the open movement totally. Then again, I am less concerned about the improbable prospect of an isolated hacker appropriating my personal data from my computer or telephone, than about big business and big government having access to my data and metadata for secret purposes beyond my control.

I have not been able to idenfify any proactive data activist project that utilises data appropriation as a central method of obtaining data beyond theoretical musings. However, I daresay this can become an area of action in the future for proactive data activists, since it is already a part of avocacy campaigns within the open data movement and a space of action where reactive data activists –that is, hackers and crackers— have accrued a long trail of exploits.

The generation of data from primary research is possibly the area where the most exuberance can be found: from conventional surveys to aerial imagery produced by community drones, whose data is then analysed utilising methods that combine naked-eye, crowdsourced observation and verification with sophisticated artificial intelligence. The current fertility not

only in techniques to obtain data, but also in ways to analyse, communicate and visualise them, arouses hopes for much more diversity and ingenuity in the future of data activism. Let us briefly review some examples.

Big data society is, in itself, a source of interest for some organisations. Data & Society, for example, is a research institute focused on 'the social and cultural issues arising from data-centric technological development' (Data & Society 2016). It is focused on how 'the same innovative technologies and sociotechnical practices that enable novel modes of interaction, new opportunities for knowledge and disruptive business paradigms can also be abused to invade people's privacy, provide new tools of discrimination and harm individuals and communities' (ibid.). It does it by providing research 'that can ground public debates, and building a network of researchers and practitioners who can offer insight and direction' (ibid.). Data & Society fosters primary research as well.

At any rate, universities, think tanks and research centres have turned out to be real engines for 'socially responsible algorithms for data science' (Pew Research 2015), providing data from primary and secondary research. From Georgetown to Pew,[320] these organisations are behind some of the most interesting data projects. An example is Pew's massive survey on global attitudes towards climate change published on the eve of the Paris Agreement in December 2015[321] (Stokes, Wike and Carle 2015). However, only some of these initiatives can be considered data activism. The Pew survey on attitude trends towards climate change is not a data

[320] Pew Research Centre is 'a nonpartisan fact tank that informs the public about the issues, attitudes and trends shaping America (U.S.) and the world. We conduct public opinion polling, demographic research, content analysis and other data-driven social science research' (2015). They are careful to specify that, contrary to what ODI does, they do not 'take policy positions' (ibid.).

[321] In 2015 COP21, also known as the 2015 Paris Climate Conference aspiring to achieve a legally binding and universal agreement on climate, with the aim of keeping global warming below 2°C.

activist project; it is simply a data project, since it does not include any advocacy elements, in line with Pew's stated values of non-involvement.

One area that is currently generating great excitement in projects, ideas and debates is the use of digital content in the form of images and video files, which can be analysed and employed in data activism as well. The amount of projects that explore this possibility and the impact that this is having in activism warrant an in-depth exploration here. The datafication and analysis of images and video files entails dealing with a different set of challenges than tackling text, as Congosto (2016) –taking about pictures shared on Twitter— and Meier (2015c) –discussing drone-enabled aerial imagery— point out. Twitter contents, for example, are increasingly difficult to study as text disappears and more messages are shared in the form of pictures, charts and infographics, and videos, which require different exploration technologies (Congosto 2016). Talking about the difficulty of analysing images, Congosto notes that 'everything changes at an incredible speed, and when you think you master a technology, the reality changes and you need a different tool to examine it. Society goes several steps ahead of research' (ibid.). Meanwhile, 'aerial imagery captured during a single 20-minute UAV flight can take more than half-a-day to analyse,' so it has become 'a big data problem' (Meier 2015c). The way to overcome this difficulty is to use 'human computing (crowdsourcing),[322] machine computing (artificial intelligence) and computer vision to make sense of this new big data source' (ibid.), in a combination of bits and atoms techniques applied to data analysis and verification. An example is the MicroMappers project[323] in Namibia aimed at sorting through aerial photos looking for

[322] In this case, crowdsourcing is not used to gather data, but to verify, interpret and analyse them.
[323] See micromappers.wordpress.com/2014/09/28/early-results-of-our-digital-expedition-to-namibia/ [accessed May 26, 2016].

wildlife, specifically rhinos, which congregated 500 volunteering ‘digital rangers’ over a weekend (Sutter 2014).

> The goal was to survey more land more quickly than on-the-ground rangers could, and to consequently protect rhinos and the like from poaching or other threats. The online army of volunteers, which included people from dozens of countries, surveyed more than 25,000 aerial photos shot by drones from Friday to Sunday, according to the group’s website. They circled wildlife they found, and uploaded their findings. Volunteers could spend as much or as little time as they wanted (ibid.).

To put it briefly, the demand for data analysis in activism is expanding and getting more complex as content becomes more visual, and the organisations of the civil society and their partners are developing new and creative ways in which to tackle this matter.

An area where data is being generated by using drones is precisely alternative mapping. In order to counterbalance maps created by governments favouring the interests of businesses and local elites, unorthodox initiatives exist to generate geolocated data and alternative maps, as I have already mentioned earlier. ‘Counter mapping’ was described by Peluso in 1995 as ‘a uniquely late-20th century phenomenon, made possible in part by both technological developments and the last decade’s push toward participatory politics and management strategies’ (1995, 400). This reactive activity emerged in Indonesia ‘in response to two decades of intensive industrial timber exploitation and the Indonesian government’s superseding of customary forest rights through official planning and mapping efforts’ (ibid., 384). It took the form of ‘sketch maps to delineate and formalise claims to forest territories and resources their villages have traditionally managed’ (ibid.). In some cases, these efforts appropriated ‘the state’s

techniques and *manner of representation* to bolster the legitimacy of "customary" claims to resources,' and the result 'raises some questions about the control of power when NGOs and other local groups utilise high technology empowerment strategies' (ibid.).

From environmental research to wildlife conservation, drones are becoming more and more common for an extensive variety of applications beyond business and defence uses (Drone Code 2016). The utilisation of drones in humanitarian assistance, for example, is growing to the point there is a Humanitarian UAV Network that boasts 'well over 2,400 members in 80+ countries' (UAViators 2016). UAVs are employed in data collection, cargo delivery and communication services. However, not all UAV operations have been welcomed by people. For example, Charles has denounced that the UN Stabilisation Mission (MONUSCO) in the Democratic Republic of Congo 'had proposed to the humanitarian community that drone platforms could be shared with the military for information gathering, but that such an arrangement would compromise the core principles of relief organisations: neutrality, impartiality and operational independence' (Oliver 2014). Others paint a brighter picture of this operation in DR Congo. The UN drones 'are used not only for actual surveillance and monitoring tasks, but also to hover at low altitude in full visibility of hostile fighters, as a deterrent to remind them they are being actively observed' (Leetaru 2015b). Leetaru does not mention whether drones can also pre-empt abuse committed by government troops as well,[324] especially if the drone UN operation is implemented in collaboration with the army. This is a case of control: who controls the drones and the data analysis that they generate, and for what purpose, is the key question here. Whatever the case, if drone-enabled data are to be combined with crowdsourced analysis, legitimacy is a precondition to attract crowds providing the volunteering teamwork. On

[324] 'Reports of mass rapes, killings and other atrocities committed by rebels and government troops' continued in 2012 (BBC 2012, 3).

the other hand, machine computing for images analysis is also a booming area. Based on the huge volume of imagery that results from drone flights, 'significant advances are occurring in the algorithmic assessment and identification of aerial imagery' (ibid.). For example, as part of Kathmandu Flying Labs, drone imagery was used to reconstruct an interactive 3D model of earthquake-damaged areas of Nepal (Meier 2015d). 'What makes their emerging humanitarian deployment so exciting is that as the civilian sector has gained access to the same kinds of high-resolution taskable satellite and real-time drone imagery, the landscape of potential applications and technological solutions has exploded' (Leetaru 2015b).

The need to extract usable data from pictures has indeed generated digital image processing techniques, which allow, for example, the identification of a person from the physiognomy of her or his face using tools like edge detectors –which detect the sharp changes in image brightness on a picture, showing the contours of a face. Facial recognition systems are being applied to verify a person's identity from a digital image captured by cameras in security and surveillance schemes. However, there is a reaction to the blooming use of these infrastructures by the third sector, which at the same time has generated the development of tools to protect the identity of people participating, for example, in offline demonstrations and direct actions. Guardian Project's *CameraV*[325] is a software application to 'capture and share verifiable photos and video proof on a smartphone or tablet, all the while keeping it entirely secure and private' (Guardian Project 2016). In partnership with WITNESS and based on the InformaCam platform ,[326] CameraV has been developed 'for use by activists, journalists, advocates and others, working in very difficult and high-risk situations, to capture and gather visual evidence and proof of abuse and rights violations' (ibid.). This type of software allows organisations to gather

[325] See guardianproject.info/ [accessed on May 12, 2016].
[326] See witness.org/ [accessed on May 12, 2016].

primary data and metadata, which can be used as evidence. The mobile phone application InformaCam 'turns metadata from a surveillance risk into a method for the production of public proof' (van der Velden 2015, 1). InformaCam works by allowing the user to manage and remove metadata from image and video files, and to store the metadata 'in such a way that the documentary material becomes better accommodated to evidentiary settings, if needed' (ibid.). InformaCam recognises mobile phones as 'context aware,' and uses their sensors and structures metadata generated 'in order to facilitate data analysis after images are captured' (ibid.). By treating data in this way, 'surveillance resistance' goes beyond the mere seeking of awareness and 'becomes engaged with investigatory practices' (ibid.). Other projects and organisations are looking at ways to protect activists and witnesses. Mann, Nolan and Wellman note that, in order to counter government and corporative surveillance, citizens are using digital tools to monitor states and corporations –an activity that these authors call *sousveillance*, which denotes bringing the surveillance cameras or whatever means of observation down to human level, either physically (installing wearable cameras or computing devises on people, rather than on buildings) or hierarchically (when ordinary people do the watching, rather than higher authorities) (2003, 338-348). The panopticon is, then, inverted, and the inmates start watching the sentries. 'The role reversal between the surveilled individual and the act of surveillance allows for the exploration of the social interactions that are generated by these performances' (ibid., 347). The authors conclude that inverted surveillance practices have a strong emancipatory potential (ibid.). Both the drone-enabled images of government and corporations' activities in the Indonesian forest and the recording of police brutality during protests could be considered cases of *sousveillance*. Other examples include Copwatch, a network of US and Canadian volunteer organisations specialised in monitoring police's activities (2016).

As in other cases, the appropriation or generation of data and technology by the people for the people (Radjawali and Pye 2015) is a strongly empowering process increasingly unleashed by data activist projects. This excitement is also generating new alliances, networks and associations, which is the focus of the next section.

6.4.3 The associations

This section deals with the plethora of conferences, awards and networks that have emerged around data in the third sector, as it has been observed in the case data journalism earlier. Most of the projects I have mentioned so far are indeed the product of an alliance among organisations contributing their skills and knowledge (i.e. the fossil fuel subsidies campaign). Here, I review the networks of networks that are popping up lately, triggered by this data enthusiasm, at a more macro level. An example is the International Data Responsibility Group (IDRG), launched in 2015 as 'a global network of experts and organisations working on the principles and standards that are required for guiding the data revolution in the context of humanitarian action, sustainable development and peace and justice' (2016). Its objective is 'to build an authoritative knowledge platform that enables responsible experimentation on the release, processing and use of data and minimising risks' (ibid.). Some of these structures could be considered match-makers and skills transferers themselves, as their objectives include the transfer of skills and the exchange of lessons learned from one organisation to the next. In the case of IDRG, this organisation aims at being a knowledge platform that provides decision-makers who seek to release, process or use data for solving societal challenges 'with an enhanced understanding of how to leverage data in a responsible and ethical manner' (ibid.). IDRG –with its *motto* 'people first in a digital age'— has a coordinating secretariat in The Hague, and its research and affiliated partners will meet every year to host the Annual International Data

Responsibility Conference. It is supported by the University of Leiden, Data & Society, Data-Pop Alliance,[327] The Engine Room,[328] GovLab[329] and UN's Global Pulse.

The Digital Humanitarian Network (DHNetwork) is a 'network-of-networks' whose purpose is to leverage digital networks 'in support of 21st century humanitarian response,' by forming 'a consortium of Volunteer & Technical Communities (V&TCs) and to provide an interface between formal, professional humanitarian organisations and informal yet skilled-and-agile volunteer and technical networks'(2016). Members include some already mentioned organisations, including Relief 2.0,[330] DigitalGlobe (also linked to Tomnod and CARTO projects), CartONG,[331] Sahana Software Foundation (mentioned as an example of geoactivism in Sri Lanka), Statistics Without Borders,[332] Gnucoop,[333] Help Earth Foundation (HEF),[334] Connected Development (CODE),[335] Humanitarian OpenStreetMap Team (HOT),[336] Disaster Tech Lab,[337] Standby Task Force (SBTF),[338] CrisisCommons,[339] DataKind (a case of data skills transferer), Humanity Road,[340] Translators without Borders (TWB),[341] URISA's GISCorps,[342] PeaceGeeks,[343] Info4disasters[344] and Geeks Without Bounds (GWOB).[345]

---

[327] See datapopalliance.org/ [accessed on May 5, 2016].
[328] See theengineroom.org. This site may be hacked [accessed on May 5, 2016].
[329] See thegovlab.org/ [accessed on May 5, 2016].
[330] See facebook.com/relief20 [accessed on May 26, 2016].
[331] See cartong.org/ [accessed on May 26, 2016].
[332] See community.amstat.org/statisticswithoutborders/home [accessed on May 26, 2016].
[333] See gnucoop.com/ [accessed on May 5, 2016].
[334] See helpearthfoundation.org/ [accessed on May 5, 2016].
[335] See connecteddevelopment.org/ [accessed on May 5, 2016].
[336] See hotosm.org/ [accessed on May 5, 2016].
[337] See disastertechlab.org/ [accessed on May 5, 2016].
[338] See standbytaskforce.org/ [accessed on April 20, 2016].
[339] See crisiscommons.org/ [accessed on May 5, 2016].
[340] See humanityroad.org/ [accessed on May 5, 2016].
[341] See translatorswithoutborders.org/ [accessed on May 5, 2016].
[342] See giscorps.org/ [accessed on May 5, 2016].
[343] See peacegeeks.org/ [accessed on May 5, 2016].

There are many more examples. Data and research are the glue linking allied cities in the C40Cities project, 'a network of the world's megacities committed to addressing climate change' (2016). C40Cities engages in research and assessments around issues such as the co-benefits of urban climate action and how cities can overcome climate challenges around the world. Data Pop Alliance is 'a global coalition on big data and development,' which includes the Harvard Humanitarian Initiative,[346] MIT Media Lab,[347] Flowminder Foundation[348] and ODI, and 'brings together researchers, experts, practitioners and activists to promote a people-centred big data revolution through collaborative research, capacity building, and community engagement' (2016). Big data challenges and opportunities are proving too big for one single organisation to handle, and these alliances are destined to increase in number and intensity around issues such as development, climate change and the environment, and humanitarian assistance as big data infrastructures expand. Table 3 shows some of the abovementioned organisations, initiatives and projects, grouped by what they do and their objectives. The following sections provide a more comprehensive classification of their specific features and action repertoires.

Table 3: Clusters of projects classified by what they do

Doing data journalism	Skill transferers	Catalysts	Proactive data activists
'Vagabundos de la chatarra' 'Panama Papers' (ICIJ) 'El Indultómetro,' 'Dónde van mis impuestos,' 'El BOE nuestro de cada día,' 'Tu derecho a saber'(Civio)	**Data skills:** DataKind Bayes Impact School of Data Tactical Tech Code for Africa Data4Good **Social skills:**	Code for Africa Omidyar Network Open Society Foundation Knight Foundation Hewlett	**Whistle-blowers' data:** WikiLeaks 'Panama Papers' (ICIJ) **Public data:** The anti-FFS campaign (OCI/ODI) The IUU campaign (ODI/porCausa)

[344] See info4disasters.org/ [accessed on May 5, 2016].
[345] See gwob.org/ [accessed on May 5, 2016].
[346] See hhi.harvard.edu/ [accessed on May 26, 2016].
[347] See media.mit.edu/ [accessed on May 26, 2016].
[348] See flowminder.org/ [accessed on May 26, 2016].

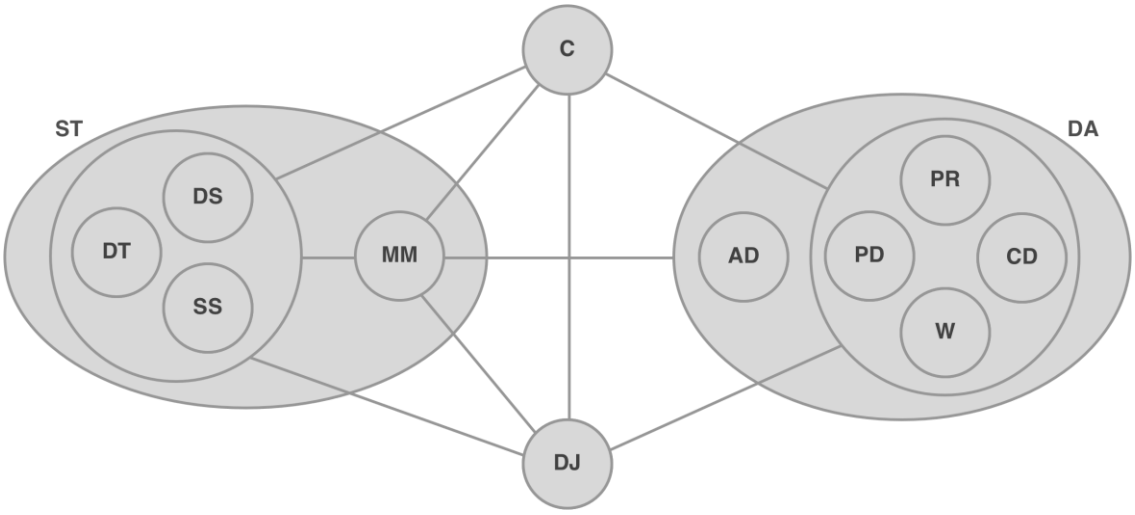
Morlan InfoAmazonia Environmental News Lab Oxpeckers Centre for Investigative Environmental Journalism	Data Science for Good eScience Outliers Collective	Foundation Shuttleworth Foundation Nominet Trust Rockefeller Foundation	'Death on the roads' (Kiln)
	Match-makers: Medialab-Prado Hacks/Hackers Hacks/Hackers Africa Code for Africa Data4Good Data Science for Social Good eScience Institute		**Crowdsourced data:** Ushahidi deployments Bellingcat
			Appropriated data: MobileMinner projects
	Data tools developers: SocialBrite Morlan CARTO Kiln Populate Vizzuality Civio Outliers Collective		**Datafied primary research:** Counter-mapping in Indonesia 'Vidas Contadas' Bellingcat Kathmandu Flying Labs CameraV Copwatch

Source: Elaboration by the author.

As shown, the organisations and data projects are assigned to one or two of the four categories: those that produce data journalism, skills transferers, catalysts and proactive data activists. Some of these categories include subcategories: skills transferers can be data skills transferers, social skills transferers, match-makers and data tool developers. At the same time, some organisations and projects appear in more than one category because they performed two tasks or have two objectives, such as Civio, which produces data journalism but also is dedicated to transferring skills. Figure 2 shows, instead, the relationships that exist among the different types of data organisations and initiatives that have been identified so far. C represents catalyst organisations; ST, skills transferers, and within these, DS are data skills transferers, SS social skills transferers, DT data tools developers and MM match-makers; DJ represents organisations producing data journalist content; DA, data activists, and within these, W depend on whistle-

blowers data, PD on public data, CD on crowdsourced data, AD on appropriated data, and finally PR on datafied primary research.

Figure 2: Associations between types of data organisations



Source: Elaboration by the author.

Figure 2 indicates, for example, that catalysts (C) facilitate and fund all sorts of data projects and organisations, and therefore are linked with data journalism projects (DJ) and all types of skills transferers (ST) and data activists (DA). Examples include Open Society Foundation, which funds the ICIJ (a journalistic organisation) and Tactical Tech, a skills transferer (ST); Omidyar Network, which funds Ushahidi (a data activist organization that gathers data through crowdsourcing) and Code for Africa (both a skills transferer and a catalyst); and Knight Foundation, which funds DataKind (a skills transferer) and Civio (the winner of the Investigation of the Year - Small Newsrooms, with the Medicalmentalia Project).[349]

The type of organisation that have less connections is the one that depends on appropriated data (AD), probably because there are still no cases that I know of yet. But the rest shows a high

[349] See community.globaleditorsnetwork.org/node/19936 [accessed August 23, 2016].

degree of collaboration and networking. In the middle of the visualisation rest catalysts (C) and organisations producing data journalism (DJ), with their central roles of facilitators and disseminators of content, and match-makers (MM), who are specialised in sitting everybody else at the same table.

6.5 The empowerment capacities of data

Pillars of Habermas's theory are the issues of labour and interaction —which can be equated with 'communication'—, as well as the possibility of emancipation (1971, 196). For Habermas, we really form ourselves as human selves, and can aspire to emancipation, in labour and in communicating with others.[350] One inherent human interest is communicating with one another, but not just interpreting and transmitting phonetic sounds, but real communication leading to undistorted understanding[351] (1984, 273-339). This notion transcends the realm of mere talking; it entertains the possibility of change through the power of the best argument and reason. Habermas includes labour in one of his earlier theories. He says that one of human beings' main interests is reproducing life through labour. 'The idea of self-constitution of the species through labour is to serve as the guide to appropriating the *Phenomenology* while demythologising it' he says referring to Hegel's *Phenomenology of Mind* (1971, 49). Labour is an instrumental effort driven by the imperatives of efficiency, and Habermas describes it as the endeavour surrounding production, as opposed to the endeavour surrounding communication (1984, 273-339). While for Habermas reason and understanding is possible, for other critical theorists, the scientific method basically led us to the atomic bomb. Adorno declares: 'no universal history leads from savagery to humanitarianism, but there is one leading from the

[350] But what about Actor–network theory's idea of machines learning and communicating with both humans and non-humans, devoted to instrumental efforts driven by efficiency? Could machines be as human as people? This an interesting question, but falls beyond the realm of this dissertation.

[351] In contrast with Hume, for whom you can become human just by interpreting and listening to phonetic sounds.

slingshot to the megaton bomb' (1973, 320) .[352] Habermas thinks that terror does not necessarily entrap us in chaos. For example, he typically calls for reasoning in order to avoid the worst consequences of science and technology: 'To the extent that the sciences are really taken into the service of political practice, scientists are objectively compelled to go beyond the technical recommendations that they produce and reflect upon their practical consequences' (Habermas 1989, 47). The same concerns reverberate today, for example, in Berendt's, Büchler's and Rockwell's reflexions on whether data scientists should cooperate in enhancing data infrastructures' faculties to spy on people (2015). Rationality, says Habermas, is the answer to these fears. So it is in labour and real communication that human beings can aspire to emancipation, and there is the possibility of understanding through participation and dialogue. But how does this relate to proactive data activism? It relates in every possible sense. At the beginning of this dissertation, I have asserted that big data infrastructures allow the tackling of the post-modern complexity from a democratic perspective. Throughout this study, I have also admitted that big data infrastructures also can be used to do just the opposite: amassed by governments –in some cases authoritarian— and big corporations, data can be used to control, manipulate, exclude and discriminate against people as well. The traditional methods of public control –that are propaganda, censorship and surveillance (apart from the monopoly of physical violence) — have been amplified and strengthened by big data. However, data and data

[352] Adorno is often read as a sombre theorist whose reflections on modern life lead to hopelessness, which some techno-pessimists echo today when discussing post-modern life. But Adorno is not a total pessimist; there is the possibility of progress, however small. 'For progress today really does mean simply the prevention and avoidance of total catastrophe' (Adorno and Tiedemann 2006, 146). In *Negative Dialectics*, he also says: 'After the catastrophes that have happened, and in view of the catastrophes to come, it would be cynical to say that a plan for a better world is manifested in history and unites it' (1973, 320). Today, a look at one of the scenarios that goes over two degrees centigrade of global warming projected by the Fifth Assessment Report (AR5) of the UN Intergovernmental Panel on Climate Change (IPCC) is enough to know that a nuclear war is not the only anthropogenic way to drive this planet into a catastrophe (IPCC 2014, 12).

infrastructures offer the possibility to empower people, who then can overcome exclusion, make decisions and act. As said, democracy is about real participation and equal opportunities, justice, voice and access for all. This access can be understood as power, and the process of getting access as empowerment. Data infrastructures employed in activism trigger such processes, through which individuals and groups, by taking active part, armed with the power of information, can make decisions and seize control over their circumstances, and that of others.

Examples are the community drones in Indonesia. Peluso has described the transforming capacity elicited by a technology-based empowerment process that can potentially alter authority dynamics, as discussed before. Peluso says 'the use of a new medium of expression, in this case maps, to express social relations has transformative power' (1995, 400). These data infrastructures have enhanced 'counter mapping' capacities, and community drones are now used to gather and consolidate data, and generate more accurate alternative maps faster in Indonesia[353] (Radjawali and Pye 2015). The drones were meant to be an additional tool to 'support research with high-quality and high-resolution spatial data in areas where access was restricted by company security and police' (ibid., 5). Radjawali and Pye describe their use as a technological revolution 'by the masses for the masses' (ibid., 13). Although they are not 'a magic wand,' combined with campaigning and political pressure, the use of community drones proliferated in Indonesia at a time when there was a political opportunity to reclaim millions of hectares of forest (ibid.). What can be inferred from this example is that the employment of drones to generate data was an empowering activity that allowed people to appropriate both data –which

[353] 'To make way for plantations, palm oil companies clear cut massive tracts of forest. In fact, palm oil production has become one of the world's leading causes of rainforest destruction. Since 1990, Indonesia has seen a 40% decline in lowland rainforests and a 600% increase in palm oil plantations' (The Goldman Environmental Prize 2015). Other areas affected by land-grabs are handled to mining companies (Radjawali and Pye 2015, 5).

were denied to them— as well as land rights in the form of symbolic maps: an appropriation process that happened in the realm of bits and atoms at the same time. Thus, the data and knowledge extracted from their analysis become, not only an advocacy component to be combined with other elements of campaigning, but also a form of real power.[354]

In this section, I address the issue of the empowering and emancipatory powers of data infrastructures applied to proactive data activism, in an attempt to describe the processes that they can unleash, described by Rodríguez as 'transformative processes that alter people's sense of self, their subjective positioning and therefore their access to power' (2001, 18).[355] I refer here to the cases already laid out, as well as new ones. Key concepts here are 'action' and 'agency,' 'control,' 'participation' and 'communication,' pertaining proactive data activism, and I start by examining them more in detail. Hereafter, agency in the context of activism is to be understood, not as simple acts (such as 'clicking on this button'), but as 'the longer processes of action based on reflection, giving an account of what one has done, even more basically, making sense of the world *so as* to act within it' (Couldry 2013, 13). This agency is not blind or detached, but cognisant and teleological, as a communicative action, in which the acting person or group is bestowed with cognitive abilities and volitional faculties. But in order to harness these abilities and faculties, this person or group must have access to information and means, and must be able to act. That is why 'agency is deeply connected to the distribution of knowledge and power' (Baack 2015, 1) as well. Many of the examples of data activism examined earlier are about granting people the access to information and enabling them so they can act.

[354] As well as an example of labour in a Habermasian sense, understood as an instrumental effort driven by the imperatives of efficiency, which produces outcomes.
[355] I am borrowing the notion of the transformations citizens' media can bring about from Rodríguez, and applying it to data infrastructures.

Exploring the 'location of agency' with regards to big data, Kennedy, Poell and van Dijck implicitly suggest that control and power lie in datafication too (2015). As seen in the case of the initiatives aimed at the appropriation and generation of data by crowds and communities, this is not only a question of accessing data, but also of producing them too, since data do not pre-exist or emerge *ex nihilo*, but arise from a gathering exercise framed in a concrete political, ideological, economic and social context. Some data activists wish to generate their own data on their own terms and within their own frameworks, exercising power by doing so. Data power, therefore, can include the ability to frame their extraction and engender alternative forms of datafication. 'It is important that debates about data power recognise that data are also generated, collected and analysed by alternative actors, enhancing, rather than undermining, the agency of the public' (ibid., 1). That is, people other than governments and big corporation can become alternative actors by generating their data on their own terms. The values behind this kind of agency are also interesting, and come from preceding non-governmental social movements engaged in human and political rights, the open source ethos, the tenet of the embryonic internet and hacker culture. Activists in the open data movement 're-articulate notions of democracy, participation and journalism by applying practices and values from open source culture to the creation and use of data' (Baack 2015, 1). Baack observes three transformations derived from these experimentation: first, 'by regarding data as a prerequisite for generating knowledge,' activists go from sharing and co-writing code to sharing data, breaking 'the interpretative monopoly of governments' and allowing people to 'make their own interpretation of data about public issues' (ibid). Second, activists connect, scale and extrapolate this idea 'to an open and flexible form of representative democracy by applying the open source model of participation to political participation' (ibid.). Third, acknowledging that 'intermediaries are necessary' to make

data accessible, activists also try to transform journalism 'to become an intermediary' (and regain its natural role in the process) and 'to act as intermediaries themselves and develop civic technologies to put their ideas into practice' (ibid.). The practices of open data activists are pertinent because they help to comprehend how data combined with a social cause enable the agency of people outside governments and big businesses.

Power can shift from brawn to brains, from north to south and west to east, from old corporate to agile start-ups, from dictators to people in town squares and cyberspace (Naim 2013, 14). But what are the principles under which data activism can unleash these processes? Although as far as I know nobody has yet articulated the principles of data activism, a look at internet rights could help develop them. The Dynamic Coalition on Internet Rights and Principles,[356] for example, have transferred universal rights from the physical world of atoms to the internet of bits, to include ten core rights and values: universal and equal dignity for all in the online environment; human rights and advancement of social justice; equal right to access and use a secure and open internet; free information without censorship; privacy online and data protection; respect for life, liberty and security; cultural and linguistic diversity; universal and open access to the internet's content, free from discriminatory prioritisation, filtering or traffic control on commercial, political, or other grounds; open standards for the internet's architecture, communication systems and formats; and human rights and social justice as the legal and normative foundations upon which the internet is governed (Internet Rights and Principles Dynamic Coalition 2015). These values underpin and are part of data activism's milieu as well, and are a prerequisite for emancipation in post-modernity. In contrast to other broadcast technologies, the internet offers expansive possibilities for horizontal communication among

[356] A group composed of activists and academics who congregate around the Internet Governance Forum (Internet Rights and Principles Coalition 2016).

citizens, while drastically lowering the barriers of organising and access to information (Shirky 2008). Data activists also draw on hackers' and hacktivists' culture when they take action to counterbalance the power of governments that amass data and restrain freedom of expression. Paraphrasing the Internet Rights and Principles Dynamic Coalition, proposed Data Rights and Principles could include: universal and equal dignity for all in both large datasets and data analysis; the primacy of human rights and advancement of social justice; equal right to access and use a secure and open data without censorship; data protection, especially for activists and humanitarian actors; respect for life, liberty and security; and cultural and linguistic diversity.

Emancipation can come also from the co-creation and transmission of content by means of 'liberated technologies' (Milan 2015), also known as 'liberation technologies' (Diamond 2010). In the case of data activism, this content will have to do with data analysis; however, the process is the same. Examining alternative media such as Radio Mutiny, the Independent Media Centre or Indymedia, Milan concludes that they are the 'outcome of emancipatory communication practices' (2015, 108):[357]

> "Practice" evokes the hands-on approach of grassroots groups in promoting reform-from-below of the communications system. "Emancipatory" denotes their promise to share and redistribute technical knowledge, in order to extend also to non-experts the possibility to control communicative actions bypassing commercial and state-owned platforms. Emancipation is strictly linked to the notion of empowerment, seen as "freedom to" communicate in one's own terms. In this sense, emancipatory communication projects should be seen as a space for people to enact their democratic agency (ibid.).

[357] Seen in this light, for example, Indymedia and its *motto* 'Don't hate the media, be the media' look more proactive than a reactive as well, although Milan and Hintz consider this organisation a beyonder (2013).

Milan notes that, although these emancipatory practices seen in the realm of media are nothing new, 'today's scale, autonomy, and self-sufficiency of the phenomenon is unprecedented' (ibid., 120). They signify a 'reconfiguration of power through prefigurative action' and 'expand unregulated spaces' (ibid.); thus, they generate alternative public spheres. The emancipatory practices' importance 'is dual: on the one hand, they offer options for participation in civic life, and secure, commercial-free spaces where people can communicate on their own terms, and on the other, their prefigurative realities are the harbinger of a more just and more secure mediascape to come. That is to say, a struggle for emancipation and democracy lies beneath the birth of data activism. What made the amalgamation of non-governmental social movements, the open source ethos and the hacker culture, with the world of data infrastructures possible? There was a first stage where the digital world and the physical world –bits and atoms— converged intensely. This period goes from 2005 to 2015, includes some of the struggles and social phenomena discussed earlier in this dissertation, and has been coined 'the era of transparency and social media activism' by Milan (2015, 113). Then 'global quest for democracy and the financial crisis that broke loose in 2008 encouraged people to eventually take back to the street again' (ibid.). And the wave of protests proliferated throughout the world, from the North African Arab Spring and the Spanish Indignad@s (or the 15M movement) to the Occupy movement, where people 'reclaimed their democratic agency' (ibid.). Concretely, 2012 was a big year for activism. A string of successes, or near successes, in the fight for privacy and digital freedom marked 2012, including the rejection of a new International Telecommunication Union (ITU) treaty on internet governance in Dubai; the defeat of the Hollywood-sponsored anti-piracy legislation, known as SOPA, in US's Congress; the rebuff in Europe of ACTA –an international treaty which, in seeking to enforce intellectual-property rights; and the opposition

in the Philippines against a cybercrime law the Supreme Court later put on hold (The Economist 2013).[358] In fact, Schmidt and Cohen find grounds for techno-optimism in 'the check that technology and connectivity bring against the abuses, suffering and destruction in our world. When exposure meets opportunity, the possibilities are endless' (2013). They advocate for the expansion of 'connectivity and technological opportunity' to improve the quality of life around the globe. Deibert and Rohozinski, however, point out that we should be careful not to ascribe emancipating capacities to digital technologies in all circumstances, and that authentic change demands 'the implementation of liberal-democratic constraints' to protect human rights, which would apply both to digital activists and to the corporations, criminal enterprises and governments their opaque activities (2010, 56). Meanwhile, Etling, Faris and Palfrey further argue that in authoritarian environments, the access to alternative online information is not enough to foster democratisation (2010, 37). Summarising, in spite of the setbacks, people have been securing gains in the fields civil liberties and others. While inequality or abuse of power cannot be eliminated, technological inclusion can help transfer power into the hands of individual people and organisations (Schmidt and Cohen 2013).

[358] This liveliness led *The Economist* to wonder whether we witnessing the rise of a new organic political power, like environmentalism in the 1960s and 1970s (ibid.). In its genesis, the environmental movement was an umbrella term for diverse groups: people concerned about nuclear energy, citizens interested in cleaning a particular river or land, anti-pesticide activists, and so on. Numerous interests can also be observed in connective activism too: from open source supporters to online privacy advocates and governmental transparency enthusiasts. The culture of the internet is a contemporary equivalent of the 1960s counter-culture in which much of the environmental movement grew up (Castells 2009). Gradually, such different components combined and formed a political platform with an overriding narrative and values that went on to wield legislative and executive power –the green parties in Europe and elsewhere (The Economist 2013). The similarities with the green movement are apparent. The former had 'intellectual leadership from within academia and so does the net movement.' Like environmental issues, 'the issues that this new movement cares about can be cast as economic ones; and when put that way they look somewhat similar' (ibid.). Environmental problems can be seen as negative externalities, and so can abuses of the internet and data as well.

Collaboration can be a source of empowerment as well. Mitchell points out how powerful it is to hear one's voice vibrate with hundreds of others.' Portes and Rumbaut talk about the group solidarity that is generated and the strength that individuals find in the collective as a reaction to a hostile environment and 'in opposition to external discrimination' (2001, 60). Although Portes and Rumbaut are talking about communities of immigrants living in the US, some of the cases examined earlier have these qualities as well (i.e. the Platform for People Affected by Mortgages). But how does it work? It works best when a shared social capital is strong and when there is a common sense of indignation. Social capital ensures that there is a diversity of shared resources (each individual, group or organisation bring in a set of abilities and networking capacities); a level of control (peer-review among the participating individuals and organisations that penalises transgressions); trust among its members that guaranties the action is transformative, and not mere a façade; dialogue among its members; cooperation in tasks that exceed the possibilities and resources of one or several individuals or organisations; mobilisation and management of capacities; generation of enthusiasm and feasible objectives; a legitimisation of leaders, when there are leaders; and a creation of internal communication systems that can be shared in equal terms, and of public spheres and structures for team work (Barandiaran 2015). Seen by this light, all the data activist cases observed in this dissertation are the product of a collective effort, motivation and social capital; data infrastructures are also the core of several global alliances sharing these elements. The 'glue' that keeps alliance members together includes an *identitarian* component that represent common attitudes, values and behaviours that generate social capital. But today the identitarian 'glue' that keeps communities together is being diluted in an ever more globalised and fragmented society. New identitarian distinctiveness emerge replacing nationality, race, language, religion and other identitarian features that have kept

communities together so far. Today the cohesion elements may not be material; they are intangible (ibid.). Data activists work mostly in alliances based on shared objectives and identities, and they rely very often on crowdsourced data and analysis (i.e. community drones). The strength of the numbers is the source of empowerment as well.

In conclusion, empowerment in data activism comes in different ways. By socialising the data infrastructures, it serves as an equaliser, equips individuals and groups that may have been excluded with the tools to participate as peers, and enables cognisant and teleological agency (Couldry 2013). Data infrastructures also allow large-scale autonomy and self-sufficiency (Milan 2015). They can also generate alternative digital public spheres that can frame dialogue and action. This agency can also be exercised in new alternative ways of datafication, allowing activists to generate data on their own terms (Kennedy, Poell and van Dijck 2015). Empowerment comes also from the individual and collective values embedded in data activism, which emanate from the social movements involved in human and political rights, the open source ethos and the hacker culture and highlight individual and collective freedoms. Empowerment can be originated in co-creation, collectiveness and common goals; that is, the strength that acting in association with others (Milan 2015) and sharing a strong social capital (Barandiaran 2015) can confer to individuals.

In this section, I have examined the different social uses of data infrastructures by governments and inter-governmental organisations, by the private sector and by journalism. I have paid special attention to the uses of data infrastructures within civil society and explored more in-depth the distinctions between reactive data activists and proactive data activists, the subject of this study. I have also expanded and reviewed a previous exercise examining proactive data activism from the perspectives of critical thought, and journalism, alternative media,

international relations and social movement studies. I have concluded that agency in data activism can be compared to a communicate action: it always includes a purpose, is framed by and refers to a normative system, and includes messages, which employ the dramaturgical, or subjective, language to connect with audiences. Proactive data activism can be compared to three varieties of journalism too: investigative journalism (because of its analytic nature), advocacy journalism (due to its political stand) and citizen journalism (because of its collaborative, rank and file character). Data activism and alternative and citizens' media have many points of contact as well. Similar to citizens' media, proactive data activism makes data part of the quotidian and places citizens at the centre of automatized data collection. As such, it brings back into the data collection machine the fundamental elements of agency and politics' (Milan and Gutiérrez 2015, 134). Citizens, whether individually or collectively, are at the forefront of production both in citizens' media and in data activism (ibid., 128), and that this fact constitutes a powerful transformative and empowering process (Rodríguez 2004). And similar to social movements, proactive data activists work towards long-term norm change, and are identified by the specific use they make of technology to form a collective identity, communicate, convey, mobilise, demonstrate and act.

I then resort to existing classifications of similar phenomena in order to confirm whether proactive data activists fall under prevailing taxonomies, to conclude that they do not conform to any. Therefore, I propose tailor-made categorisations. The first one, based on their objectives, takes me to divide proactive data activists in four main groups, with different subcategories: the ones that practice journalism, the skills transferers, the catalysts and the actual data activists. Within the data activists, I look into geoactivism, which is the field of data action where most cases are found. The second exercise consist of looking into proactive data activism cases from

the perspective of the origin of the data employed, which leads me to classify cases in six categories: the recipients of data via whistle-blowers; the ones that produce new analysis from available, but unrelated and unexplored, datasets, producing secondary research; the organisations that get data via crowdsourcing platforms; the ones that appropriate data; and the organisations that conduct primary research whose findings can be datafied and analysed. Finally, I inspect the empowering and emancipating processes unleashed by the use of data and connective technologies, finding commonalities between former social movements and causes, and reason for cautious optimism in the power proactive data activism. In the following section, I include the case study.

7. Case study: Ushahidi

> "We believe data should not be dehumanized. It should not be difficult to get, and it should not be difficult to understand," said Daudi Were, Ushahidi chief executive officer. "These are the problems our users tell us they have. So we design and build our platform to solve these challenges" (Ushahidi 2016c).

Ushahidi –'testimony' and 'witness' in Swahili— is a non-profit software company that develops software that collects and visualises data, information and action though interactive mapping to enable humanitarian efforts in cases of crisis, emergency, armed conflict or disaster. Ushahidi also trains people to use its platform and creates communities. In doing so, it generates social change. The deployments that use the Ushahidi platform usually work within situations of profound inequality (i.e. the Haiti earthquake of 2010), and provide both disadvantaged individuals and humanitarian organisations and volunteers with the tools to access usable information, make decisions and act upon their own circumstances, and that of others'. It does so by creating alternative digital public spheres where people can communicate and participate as equals; in this process, it becomes not only a channel for communication and coordination, but also an empowerment tool and a challenge to the *status quo*. Ushahidi is a proactive data activist organisation, because, taking advantage of the opportunities offered by big data infrastructures and the new knowledge that data analysis provide, equips people with the weapons of facts for civic engagement (Milan and Gutiérrez 2015, 127). Ushahidi is an instance of a class of phenomena (proactive data activism) that provides an analytical frame which the case illuminates. As said, this a descriptive exercise based on one case to show what proactive data activism looks like from proximity (Woodside 2010, 322).

Ushahidi has played an important role in developing what is known today as ‘digital humanitarianism’ (Meier 2015a), which is humanitarianism that uses spatial, communication and big data infrastructures in order to aid humanitarian efforts. Utilising crowdsourcing via different channels –from emails and text messages (SMS) and online social networking services—, and visualising the resulting data analysis by interactive mapping software, Ushahidi is among the pioneers of digital humanitarianism (Hesse 2010). The surplus of information generated during disasters and crises –called ‘big crisis data’— can be as disorienting as the lack of information. Making sense of big crisis data has always been a challenge for traditional humanitarian organisations, which is why they are relying more and more on digital humanitarians (Meier 2015a, 18-19). Since the Ushahidi deployment during the Haiti earthquake of 2010, crisis-mapping technology has emerged as a tool to help humanitarian organisations deliver assistance to victims of armed conflicts and disasters in different parts of the world. An independent evaluation found that the Haiti deployment addressed ‘key information gaps’ a) in the early period of the response before UN and other large organisations were operational, b) by providing situational awareness and critical early information with high degree of geographic precision, c) by offering situational information for small NGOs that did not have a field presence, d) by supporting small responses so they could adjust their target needs and, d) by facilitating private citizen actors (Morrow, Mock and Papendieck 2011, 4). Both theory and practice of humanitarianism has been transformed by this fact.

In this case study, I verify how Ushahidi fits in the description of a proactive data activism organisation that I have outlined in the first sections of this dissertation, and observe how Ushahidi uses data for social change. More precisely, I look into what their theory of change, tactics and strategy, challenges, and main successes and failures are in order to

understand the intersections between Ushahidi's humanitarian efforts and its long-term aspirations from the perspective of data use. The idea is to reinforce the understanding of the rules guiding data-enabled social change to generate an analytic paradigm that can be used in the future for further study; and to help identify various pressures that surround data-enabled social change in the context of disaster and develop a model for data activism.

7.1 Origins and how it works

Frustrated by a lack of mainstream media coverage and information of Kenya's postelection violence in 2007,[359] four tech-savvy activists and bloggers, called Ory Okolloh, Erik Hersman, Juliana Rotich and David Kobia,[360] set up a website to gather and visualise eyewitness reports of violence reported via email and text message on a Google Maps map[361] (Keim 2012; Ushahidi 2016a). They decided not to build a tool from scratch, but to combine the capacities of humble, widespread tools such as mobile phones, databases and online maps to develop the website quickly and cheaply (Keim 2012; Adewumi 2008). Others, like Daudi Were, soon joined in. His mission was to verify eyewitness testimonies with international and local media reports, NGOs, aid groups and government sources in order to avoid spreading rumours and false information, which could make things worse in a crisis (Smith 2008). Soon too, Patrick Meier became Ushahidi's director of crisis mapping. And with that core team was set up. By 2009, the Ushahidi platform had been used to observe elections in Kenya, India and Mexico, and to track medical supply shortages in Malawi and Zambia (Bernholz, Skloot and Varela 2010, 24).

---

[359] More than 850 people died in attacks and reprisals among different ethnic communities loyal to either President Mwai Kibaki or opposition leader Raila Odinga (Smith 2008).
[360] See their brief CVs in Annex III.
[361] Ushahidi provides the option of using OpenStreetMap maps in its user interface, but requires the Google Maps API for geocoding.

The first phase in the growth of any organisation, according to the Greiner's Growth Model, is growth through creativity (Exponential Training 2016). This is precisely how Ushahidi got on its feet. Keim describes how in these early days of Ushahidi solutions from a dedicated core group of bloggers and volunteering software programmers, using agile methodologies, would emerge on the move employing ingenuity and collective problem solving capacities (2012). Enthused by the examples of WordPress, Firefox and Linux, Ushahidi asked the programming community for help to develop a software packages built on open source code, and was able to engage 125 of them (ibid.). In 2010, Ushahidi gained world attention as an emergency facility in aftermath of the Haiti earthquake, and the proliferation of mobile technology played a major role in its success. A deployment was set up in the days following the earthquake.[362] A 'community of techno-humanitarian volunteers' coalesced around calls for help from *victims*, and, Ushahidi, among the volunteering organisations, put its mapping platform to the task (Hesse 2010). 'Chaos prevailed. Transportation was limited, if not impossible. Lines of communication were broken. A few radio stations continued to broadcast, but the disaster's scale was overwhelming. Only one form of mass communication remained relatively intact: cellular phones' (Keim 2012). Thousands of eyewitness accounts submitted via email, text or Twitter, were analysed and made visible through its digital platform, and the initiative was then used to coordinate disaster relief operations (Hesse 2010). 'Within days of the Haiti earthquake, the platform was customised and a text message hotline set up. Hundreds of volunteers from around the world processed texted reports of trapped people, medical emergencies, and requests for aid, feeding the reports into a map that rescue workers could use. A total of 1,500 reports were gathered and mapped in the first two weeks, and more than 3,500 created altogether' (ibid.). The

[362] Before the disaster, there were only 108,000 fixed land telephone lines in the country, compared with 3.5 million mobile phones, so after the earthquake the latter become almost the only means of communication (Hesse 2010).

Haiti deployment was a catalyst for change in the humanitarian community (Owen 2015, 124-125). Keim suggests that, in every step of the way, Ushahidi learned different valuable lessons (ibid.). In the Haiti disaster, Ushahidi managed to attract another group of volunteers –recruited largely through social media from the Haitian diaspora community— to customise the software to local needs. One of the important things they did was, for example, to translate to and from Creole locale-specific references (ibid.). This proved to be vital for the success of the deployment. However, in Haiti, Ushahidi had failed to engage with the community that would be the primary users of the platform: humanitarian workers (ibid.). This lesson was integrated in their guidelines. The Haiti experience was the starting point for a new figure in humanitarian emergencies, that of the 'digital humanitarian' (Meier 2015a), 'techno-humanitarian volunteer' (Keim 2012), 'humanitarian technologist' (Global Facility for Disaster Reduction and Recovery Labs 2010) or simply 'digital volunteer,' as Aguilar, interviewed for this dissertation,[363] calls them. From 2010 onwards, these experts —'who are most often technical professionals with deep expertise in geographic information systems, database management, social media and/or online campaigns'— started to apply their skills to deal with disasters (ibid.). Ushahidi's involvement in deployments can be none whatsoever, since anyone can launch a deployment without even communicating with the organisation (i.e. a 'Burger Map');[364] it can provide some support so the deployment can be customized (i.e. Ayuda Ecuador, in which Oduor, from

[363] Luis Hernando Aguilar is a knowledge management and digital information specialist, formerly head of OCHA Colombia, who has pioneered in testing Ushahidi for rescue operations since 2009, when Ushahidi liberated its code. He was instrumental in making OCHA the first UN agency to use the tool (in Libya), and participated in deployments to tackle the Colombian and Libya crises, the typhoon Haiyan and Ebola emergency responses in Philippines and West Africa. He was asked to go to Quito (with Ushahidi's funds) to set up Ayuda Ecuador in 2016.

[364] See flickr.com/photos/ushahidi/5703941504 [accessed on August 2, 2016]. The map is no longer active.

Ushahidi, was involved);[365] or it can lead the deployment (i.e. the deployments in Kenya in 2008, 2010 and 2013). In any of these cases, Ushahidi deployments work above all thanks to a big volunteer community unconnected with the organisation. So who are these volunteers? They can be divided into two main categories: expert humanitarians and non-expert witnesses.

To help ensure that the effort catalysed by Haiti's earthquake could be replicated elsewhere in the future, Ushahidi helped set up the Standby Task Force, an independent mapping community of trained volunteers who are permanently ready to participate in a crisis immediately after it is declared and before a larger community assembles (Standby Task Force 2016).[366] The idea was launched by Meier in October 2010, at the International Conference on Crisis Mapping.[367] The Standby Task Force was first tested a month later in November, 2010 in a drill exercise of an earthquake affecting Bogota, together with the local Fire Department, International Search and Rescue Advisory Group (INSARAG) and the UN Office for the Coordination of Humanitarian Affairs (OCHA) Colombia, prompted by Aguilar. The objectives were to test the ability of the crisis expert mapping volunteer network to participate and contribute to the efforts, assess and revise several protocols dealing with non-expert volunteers, and appraise the capacity of the tool to gather text messages on simple phones, visualise them on

[365] Although one of its main leaders, Aguilar, interviewed for this dissertation, says she could not contribute much since she does not speak any Spanish.

[366] In contrast, when UN agencies need to rely on volunteer networks, they summon the volunteers, train and deploy them on a case by case basis, says Aguilar, who has lobbied internally so the UN sets up a similar structure of volunteer experts, ready to step in before other assets are activated and deployed.

[367] The Crisis Mappers Network is the largest international community of experts, practitioners, policymakers, technologists, researchers, journalists, scholars, hackers and expert volunteers engaged at the intersection of humanitarian crises, new technology, crowdsourcing and crisis mapping (Digital Humanitarians 2016). The CMN was launched at the first International Conference on Crisis Mapping (ICCM) in 2009. It engages 6,000+ members in over 160 countries, who are affiliated with over 3,000 different institutions, including more than 400 universities, 50 UN agencies and projects, first responders operating in both the civilian and military space, dozens of technology companies, several volunteer and technical community networks and global, national, and local humanitarian and disaster response and recovery organisations (ibid.). Its co-founders are Patrick Meier and Jen Ziemke.

a map and a timeline, and categorise them, recalls Aguilar. In 2011, OCHA invited Ushahidi's collaboration as the Libyan civil war intensified in March to create the Libya Crisis Map 'to help provide better situational awareness of the unfolding situation on-the-ground,' because the UN did not have physical access to the country or the idle capacity to gather, verify and process the enormous amount of available online information (Verity 2011).[368] Verity says there are three core areas in this collaboration that have influenced OCHA's work: the speed of producing relevant information in the early phases of an emergency; the good coordination between the information management teams at OCHA with the self-organised task-team based volunteers; the culture change of using non-UN standard software (i.e. Skype) to be able to work with external volunteers (ibid.). In 2013, MIT nominated Ushahidi among the smartest companies generating breakthrough technologies. By 2013, the countries that have used Ushahidi the most were: US, Italy, UK, Canada, Russian Federation, India, Indonesia, Mexico, Australia, Brazil, Germany, Colombia, Egypt, Spain, Argentina, South Africa, Japan, Philippines, France and Kenya (Leson 2013). Ushahidi launched a new version in October 2015. And after a six month research period involving 100 users, in June, 2016, it introduced a design improvement 'to bring clarity to the data and storytelling,' and streamline workflow with more focussed interaction with the most popular features, that is Map, Timeline and Activity (Ushahidi Inc 2016a). The design includes new features such as support for uploading images to posts and exporting data to CSV, so users can analyse data outside the deployment if they wish (ibid.). On April, 2016, Ushahidi received a US$1.5 million grant from the Global Development Lab (from the US Agency for International Development) to work with partners on election strengthening platforms, anti-corruption initiatives, humanitarian crisis response and project that tackle systemic human rights abuses (Ushahidi Inc 2016b). By then, Ushahidi had raised about US$4 million with the vision to access

[368] OCHA Colombia was also involved.

fifty million impacted people and attain an 85% financial sustainability through earned revenues by the end of 2018 (ibid.).[369] This is an important threshold for the organisation, since initially the same 'informal, networked structure' that had made 'an effective platform for the advancement of social good' possible, constituted 'a stumbling block to raising foundation funds' because it did not conform to the organisational model funders were comfortable with (Bernholz, Skloot and Varela 2010, 30). Since its take-off, Ushahidi has been used in every major emergency or conflict around the world, expanding to developed countries and non-violent conflicts, for example, in the 2010 storm clean-up after Washington's *Snowmaggedon* (Bernholz, Skloot and Varela 2010, 24), in the wake of the evictions crisis in 2011 in Spain, by 'Stop Evictions' (Tetuan 2011) or in anti-corruption deployments in Brazil, Morocco, Indonesia and Kosovo, among others (Were 2015). Ushahidi has widely spread outside crises and developing countries (Bernholz, Skloot and Varela 2010, 24).[370] It currently boasts 90,000 deployments, 6.5 million testimonies or reports and a 20 million people reach (Ushahidi 2016a). Ushahidi works in partnership with iHUB,[371] m:lab (East Africa),[372] Gearbox,[373] BRCK,[374] Standby Task Force, AkirachixiHUB.[375] It also collaborates with organisations such the Humanitarian OpenStreetMap

[369] By December, 2015, Ushahidi declared it had grown its revenues by an average of about 60% year over year, resulting in roughly 30% of its overhead being covered by paid client services, primarily through consulting projects (Manning 2015).

[370] The Ushahidi platform was launched for the first time in January 2008, during the post-electoral violence in Kenya. Other key deployments, in chronological order, include: the Haiti earthquake (January, 2010); the Chile earthquake (February, 2010); the Kenyan referendum (August, 2010); the Libya crisis (starting in early 2011); the Syrian crisis (starting in April, 2011); the Kenyan general elections (March, 2013); typhoon Haiyan in the Philippines (November, 2013); the Ebola crisis in West Africa (starting in December, 2013); the Nepal earthquake (April, 2015); and the Ecuador earthquake (starting in April, 2016).

[371] See ihub.co.ke/ [accessed on August 6, 2016]. iHUB is an innovation hub, co-working and hacker space for the technology community that was started 2010 by Hersman in Kenya.

[372] See mlab.co.ke/ [accessed on August 6, 2016].

[373][373] See gearbox.co.ke/ [accessed on August 6, 2016].

[374] See brck.com/ [accessed on August 6, 2016].

[375] See akirachix.com/ [accessed on August 6, 2016].

Team (HOT), the Harvard Humanitarian Initiative, Translators without Borders and The Assessment Capacities (Were 2016).

How does it work? Witnesses submit reports by text message, email or web postings, these reports are verified, and the software aggregates and organises the data into a map and a timeline. The managers of the deployment can categorise and verify the data so it is more helpful to users. In addition to its crisis-mapping software, 'the company has also launched a product called Swift River that uses machine-learning algorithms to extract and organise accurate information from the flood of emails, text messages, blog posts and tweets that can seem overwhelming in the first days of a crisis' (MIT Technology Review 2013). With time, Ushahidi evolved into a more sophisticated organisation too. Aguilar, interviewed for this dissertation, says that it stopped being a citizen's initiative to become a commercial structure a long time ago. Today, Ushahidi offers a variety of free and commercial services, including setting up a website, servers and hosting to visualise information using multiple SMS integrations (Twilio, Nexmo or SMSSync); adapted maps with customised features to either integrate features that are not on offer in the standard version or to integrate crowdsourcing in other platforms; initiation and maintenance of platforms; customisation of Ushahidi deployment to match a given organisation's branding; and tailor-made training and consulting, online or on-site. It also issues user manuals with guidelines on how to set up and configure a deployment, how to manage people and data, and how to visualise and analyse data on a deployment. Ushahidi's business model is based on selling consulting services, and gathering grants and donations from philanthropic organisations, including Humanity United (which granted a kick-off grant of US$200,000 grant) and the Omidyar Network (which in 2009 gave the non-profit US$1.4 million) (MIT Technology Review 2013). Other funding partners include Cisco, Ford Foundation, Hivos, Google,

MacArthur Foundation, Rockefeller Foundation, Knight Foundation and USAid (Ushahidi 2016a). The staff has expanded from its six member core team to a total of 27 employees at the time of writing, although from the core team only Were is still with the organisation as member of the executive group (he is the Executive Director), while Rotich, Kobia and Hersman are members of the board (ibid.).[376] A 20% of Ushahidi's budget comes from fee-based consulting projects, with customers including the ICT4Peace Foundation and the World Bank (Keim 2012). Any organisation or community willing to launch a deployment can have free access to basic features –including unlimited posts, one survey, one user, web and email data collection, and map and timeline data visualisation (Ushahidi 2016c). A US$99 monthly fee gives organisations the access to all that, plus unlimited surveys, five users, SMS and Twitter data collection, bar and line chart visualisations, and the possibility of importing and exporting data; a US$499 monthly fee adds twenty users and the possibility of customisation, among other features; more complex adaptions can cost more and are dealt with on a case by case basis, but NGOs and grass-roots projects can apply for a *pro bono* plan that is customised (ibid.).

An example is the deployment to deal with the Ecuador's earthquake on April 16, 2016. The platform was set up immediately after the first tremors. The idea was 'to generate collectively data relevant to the emergency, threats, logistic needs and response that the affected population was experiencing… and effectively channel the efforts by different institutions and agencies' (Ayuda Ecuador 2016). It was an OpenStreetMap map powered by Ushahidi. By then, though, they were not alone. There were other three platforms performing similar tasks: Terremoto Ecuador,[377] set by Linq on OpenStreetMap; Unidos Ecuador,[378] set by

[376] During this study, although not member of the original team, Mitchell left the organisation as well.
[377] See terremotoecuador.com/#zona/-2.0024,-77.3288/7/5/0 [accessed on August 6, 2016].
[378] See unidosecuador.org/ [accessed on August 6, 2016].

Cuestionarix.com on Google Maps; and Yoveoveo.[379] Ayuda Ecuador shows reports from April 16 to May 5, 2016, categorised as 'emergencies' (a total of 61), 'trustworthy reports' (53), 'threats' (62), 'logistics' (294), 'responses' (254) and 'news about people' (34). Each of these categories has subcategories indicating action or places, for example, 'responses' was divided in 'housing,' 'health services,' 'search and rescue,' 'refuge,' 'food distribution,' 'water purification and sanitation,' 'basic non-food products,' 'rubbish removal,' 'morgue,' 'water distribution,' 'donation point' and 'evaluation of humanitarian needs.' The map shows where those actions or places were located.

'Throughout our history, we have been trend setters and norm-breakers. We are a unique organisation that has changed the paradigm of expectation more than once' (Manning 2015). Among its achievements, Manning lists the fact that an African start-up can compete as a global technology company and scale up worldwide, and that Ushahidi is now part of projects managing multi-million projects, such as Making All Voices Count[380] and Resilience Network Initiative.[381]

7.2 Ushahidi seen from different perspectives

In this section I briefly look at Ushahidi from the perspectives of critical theory, and journalism, alternative media, international relation and social movement studies, as I did more in general about proactive data activism in the previous section.

Ushahidi includes teleological or purpose-oriented elements, norm-regulated elements, and elements of the dramaturgical action. By examining, analysing, verifying and interpreting data

[379] It is a phone application to report anomalies, only available in Ecuador at the time of writing.
[380] See makingallvoicescount.org/ [accessed on August 6, 2016]. A partnership between Hivos, Ushahidi and the Institute for Development Studies.
[381] See cityresilience.net/what-is-rni.html [accessed on August 6, 2016]. An Ushahidi initiative with Rockefeller Foundation's funding.

generated by people affected by a disaster, crisis or conflict, there is an analytical action involved. A hermeneutic process places the findings of such analysis in a political and social context of the crisis that the initiative is tackling. The simple act of placing the information on a map makes this a political endeavour, as any representation of the world include social and political constructions, as examined earlier. Indeed, Ushahidi's maps are not a bi-dimensional geographic representations of territory, but a knowledge and production tool signifying complex situations (i.e. a crisis), relationships and heterogeneous objects (i.e. a place experiencing lack of water), social, political or technological processes (i.e. the location where the evaluation of humanitarian needs takes place), events and places (i.e. a donation point), and mutable interactions and networks (i.e. the location where distribution of water takes place). Because of what they do, Ushahidi's maps are 'action' instead of just representation (de Soto 2014, 361). They are not static pictures of a geographic territory, but knowledge and production tools, representing multifaceted situations, relationships and objects, social, political or technological processes, events and places, and changing connections and networks. Finally, the dramaturgical element is present in the communications between the team of professionals and volunteers with the witnesses, *victims* and other humanitarian workers and teams, as well as other participants (i.e. authorities, journalists). That is, Ushahidi initiatives involve communicative actions, since they always include a purpose, are framed and refer to a normative system, and include messages, which employ the dramaturgical, or subjective, language to connect with people.

Ushahidi generates alternative public spheres where people can exchange information in equal terms. The emancipation seen from this perspective comes when the *victims* stop being helpless, and become equal participants and proactive witnesses by joining in the creation and use of this alternative public sphere. Armed with self-generated, useful information and

knowledge, they not only act upon their situation but also upon others,' helping them by generating and sharing relevant information and prompting action. Ordinary citizens, empowered by technology, can participate in humanitarian efforts on equal terms. 'What we have done, and this is quite radical, is broadening the definition of who is an *expert*. Local knowledge is extremely important, and people without formal education can be experts on how thigs are to be done. The question is how you tap into that expertise' says Were, interviewed for this dissertation. The values of objectivity, fairness and accuracy of good journalism are embedded in Ushahidi's DNA as well (Milan and Gutiérrez 2015). Ushahidi started precisely as a reaction to the lack of coverage by mainstream papers of Kenya's postelection bloodshed in 2008, and its initial mission statement says that it intends 'to change the way information flows in the world' (HumTech 2015; Michael 2014) by giving voice to the voiceless and representing people previously ignored by the media. In this, it becomes close to a citizen medium.

Ushahidi does it by using the journalistic method of verification too. As seen, Were joined in originally to verify eyewitness testimonies with international and local media reports, NGOs, aid groups and government sources (Smith 2008). 'Relying on user-generated content – without adequate fact-checking – can mean that information is skewed or falsified to inflame passions… Organisers behind Ushahidi.com said they understood that false rumours could have major repercussions: more violence' (ibid.). So they established their own verification mechanisms. The platform offers users a Guide of Verification too, including advice, for example, to include only information coming from multiple sources.[382] Proactive data activism is comparable to three varieties of journalism: investigative journalism (because of its analytic nature), advocacy journalism (since both usually maintain a political view) and citizen

[382] See wiki.ushahidi.com/download/attachments/4260045/Ushahidi-Verification-Guide.pdf?version=1&modificationDate=1375995186000&api=v2 [accessed on August 7, 2016].

journalism (because of its collaborative character). Ushahidi not only complies with this characteristics; its platform has been employed as part of journalistic practices as well. Examples are the 2009 coverage of Gaza by *Al Jazeera*,[383] the *BBC's* coverage of the London tube strikes of 2010,[384] and *Houston Chronicle's* map showing the bad shape of the streets of Houston in 2011.[385] Sandoval-Martín and Espiritusanto say that tools like Ushahidi provide value in the communicative chain and generate a sphere of collaboration between 'professional' journalists and citizen journalists (2016, 458). They also point out that 'there is a massive externalisation of elements that were before distinctive of journalistic processes through the use of social innovation tools such as Ushahidi, which becomes integrated in the journalistic practices of both citizens and journalists' (ibid., 468).[386] And in these practices, 'journalism fulfils its social function of providing the information citizens need with most urgently' (ibid.).[387] This last remark echoes Sampedro's opinion about journalism recovering its original purpose through data practices (2014, 14). The alternative media perspective proposes that citizens' media create post-bourgeois 'public spheres.' The same way, although not properly a media organisation, Ushahidi creates the environment and the means –the alternative public sphere— for witnesses of crises and disasters to exchange information in their own digital, interactive, collectively-created space. As what happens with alternative media too, the specific use of the Ushahidi platform to represent armed conflicts could also be considered as manifestations of the *subaltern counterpublics* proposed by Fraser (1990, 58-61), as it is an expression of marginalised groups of

---

[383] See ushahidi.com/blog/2009/01/02/al-jazeera-labs-is-testing-ushahidi [accessed on August 7, 2016].
[384] See journalism.co.uk/news/bbc-using-ushahidi-039-s-crowdmap-to-plot-tube-strikes/s2/a540447/ [accessed on August 7, 2016].
[385] See badhoustonstreets.crowdmap.com/main [accessed on August 7, 2016].
[386] 'Se está produciendo una externalización masiva de elementos propios del proceso periodístico mediante la utilización de herramientas de innovación social como Ushahidi, que se integran en la práctica informativa tanto por la ciudadanía como por periodistas' (ibid.).
[387] '…el periodismo cumple su función social de informar sobre lo que los ciudadanos precisan con mayor urgencia' (ibid.).

victims and civilians forming their own post-bourgeois public spheres. The alternative discourse facilitated by Ushahidi around armed conflicts, for example, is different from the discourse emanating from the forces holding the weapons, and constitutes a challenge to hegemonic domination and views. An example is the very first Ushahidi deployment in Kenya in 2008. When Ushahidi was launched, Okolloh blogged that 'the number of deaths being reported by the government, police and media is grossly underreported,' while the true picture of what was going on came from family and friends in affected areas (2008). Her hope was to put names to the people who had disappeared and died in the post electoral violence in a way that they were not forgotten when the crisis came to an end, since to truly move forward, the truth needs to be told first. 'Ushahidi is our small way of contributing to that' (ibid.). Meier talks about Ushahidi as a 'liberation technology' (2012, 95), and considers its use during the Arab Spring as a threat to the Egyptian and Sudanese governments, 'because it challenged the *status quo*' (2011a, 3). But the result of these challenges is a matter of discussion. Meier acknowledges that 'evidence suggests that this challenge tipped the balance of power marginally in favour of civil society in Egypt, but not in the Sudan, and overall not significantly' (ibid.).

Ushahidi is not only a challenge of dominating groups monopolising the use of armed forces and mainstream information channels, it is generally acknowledged as a disruptive technology that has changed the *game* of humanitarian assistance, placing non-experts at the front of humanitarian efforts alongside with experts. Bernholz, Skloot and Varela use Ushahidi as an example of what they call 'disruptive philanthropy' (2010). 'Network-enabled volunteer groups like Ushahidi are radically different from incorporated enterprises with bylaws, mission statements, formal boards of directors and geographical limits' (ibid., 37). These groups operate outside the existing regulations applying to humanitarian agencies, are managed by individuals

who seek social solutions, not monetary gain, and rely on new models of peer-to-peer accountability (ibid., 42). Citing other examples, these authors proclaim that 'philanthropy in the US is entering a new phase,' during which an information infrastructure is being built connecting 'the long tail of donors to the long tail of doers,' with the potential to 'open up and systematise processes and decision-making practices that have heretofore occurred exclusively behind closed doors' (ibid.). Rice even suggests that Ushahidi deployments threatens aid organisations' funding model, and even their existence, as they exhibit a greater degree of transparency and are 'free' (2010). I do not totally agree with Rice, since Ushahidi is no replacement for humanitarian agencies, and they do not pretend to be, and they are not cost-free either. But he has a point in saying that 'the democratisation of information creates the ability to see where aid is flowing, where it should be going and just how affectively aid organisations are responding to a crisis' (ibid.). What Ushahidi means for traditional humanitarian agencies is a test that requires them to change, adapt and improve, as Verity suggests when talking about ways in which OCHA has been defied and transformed by its collaboration with Ushahidi (2011).

Another subtler way in which Ushahidi is challenging the *status quo* is with cross-pollinating, a typical practice in data-enabled activism. Is it a social movement? Some observers seem to think so (Grabowski 2012),[388] as Ushahidi shows many of the social movements' traits. Is it a humanitarian organisation dealing solely with emergencies? Or an advocacy organisation

[388] 'The Ushahidi deployments are often founded by one or two people, but rely on much more people to take action. From my interviews, I got the impression that many informants felt that they were a part of something bigger, as one informant put it: "*Political activism has grown because of the feeling of sharing".* An interactive online platform, as Ushahidi, appears to be an effective tool for a transnational audience, exactly because it allows cooperation on a massive scale. I place Ushahidi in the emergence of new social movements: loosely coupled egalitarian network that stretch beyond national borders and embrace diversity, decentralisation and informality to a larger degree than traditional social movements. I see Ushahidi as a movement of hundreds of small social movements –deployments with their own agenda connected in a large network' (Grabowski 2012). Others talk about Ushahidi being part of 'the start-up movement' (New Africa Business News 2016) or 'the digital democratisation movement' (Nelson 2014).

dealing with underlying causes? Does it practice citizen' journalism, since it relies on citizens' reports? It is probably a bit of all these options. As said, humanitarian action has evolved into a highly specialised endeavour, dealing with immediate emergencies on a short-term basis. Typically, in traditional international assistance, other institutions dealing with the underlying long-term socioeconomic factors replace humanitarian organisations once the crisis is over. But Ushahidi is being employed in both situations; that it, practices both 'crisis mapping' –to deal with crises— and 'activist mapping' –to gather evidence that can be used for lobbying and advocacy. The Kenyan 2008 elections deployment is an example of the first type of humanitarian undertaking: the website was launched to deal with the post-electoral violence of 2008, not to campaign against or for democratic development in general terms. However, other Ushahidi deployments,[389] such as 'I am Nirbhaya,' are focused on raising awareness about human rights issues, in this case about violence against women and children in India.[390] Also it is important to note that Ushahidi, on its own initiative, repeated the experience of supporting a peaceful, honest and free voting in the Kenyan constitutional referendum of 2010, creating the platform Uchaguzi, which was later employed in monitoring the general elections in 2010 in Tanzania, in 2011 Zambia and in 2013 again in Kenya. This long-term strategy is typical of campaigning organisations. One of the things that distinguish humanitarian agencies and advocacy organisations is that the latter develop theories of change beyond the immediate effect of assisting victims of disasters and violence. Does Ushahidi have one? The reason behind its electoral platform –Uchaguzi— is that 'elections are an important and integral aspect of

[389] A distinction has to be made between the deployments initiated by the Ushahidi team and the ones initiated by other organisations, which can count on support by Ushahidi (i.e. Ayuda Ecuador).

[390] However, the map highlighted at the Ushahidi website only shows 22 news (see /iamnirbhaya.ushahidi.io/views/map [accessed on August 8, 2016]), while the one shown at the campaign website, powered by Ushahidi too, shows thousands of reports disaggregated as 'abuse,' 'molestation,' 'murder,' 'abduction,' etc. (see iamnirbhaya.me/ [accessed on August 8, 2016]).

democracy, meant as a mechanism with which citizens hold their leaders to account for past actions and future promises,' says Mitchell. He even articulates a specific theory of change: 'When citizens are enabled to actively contribute to an environment in which a free and fair election can take place, it will become less attractive for politicians to meddle in elections and stimulate the emergence of truly citizen-oriented politics.' How does Ushahidi contribute to such objectives? By enhancing electoral monitoring, creating a faster reporting and alert system, and bringing in the voice of citizens as a new dimension, through crowdsourcing (ibid.). This ToC is taken to another level with another notion: 'When citizens see that their voices can make a difference, it will become more attractive to be politically engaged' (ibid.). So, this is not only about adding voices to the overall assessment of the process, but also about 'the promise that serious incidents would be relayed to those who could act upon and solve them' (ibid.). An evaluation of the Kenyan 2013 election says that 'Uchaguzi's strategy was to contribute to stability in Kenya by increasing transparency and accountability through active citizen participation in the electoral cycle' (Cio 2013). In the interview for this dissertation, Were leaves theories of change to the organisations Ushahidi partners with, although he acknowledges the fact that, with time, Ushahidi has got closer to them and their causes. In fact, specific ToCs are articulated in most deployments, including those initiated by Ushahidi. In the case of the Ushahidi Haiti initiative, the declared ToC is: 'Access to accurate and timely information from the ground during post-crisis response periods will enable humanitarian responders to act more efficiently' (Meier 2010). The intrinsic problem identified in this theory of change is that, in spite of being crucial, the immediate reactions to crises is not always efficient, in part because of an information overflow, and there is a need for creative solutions in order to increase usefulness and timeliness in rescue operations.

I believe that Ushahidi has a double nature, and has evolved from being just a record of information about brutality and abuse in Kenya 2008, to being a hybrid humanitarian/campaigning organisation, following in other organisations' steps[391]. Despite the obvious differences, and with all due caution, Oxfam and Ushahidi share some similarities. Oxfam was originally Oxford Committee for Famine Relief, an organisation founded by Quakers, social activists and Oxford academics in the middle of the Second World War. Soon, its mission became to persuade the British government to allow food relief beyond the Allied blockade so it reached the famished people of Nazi-occupied Greece. Lobbying is one of the tools that can be employed by campaigning organisations in order to change social practice or policy. Oxfam developed into a hybrid soon enough by creating humanitarian programmes to respond to emergencies on the ground. Ushahidi has stridden the same path, but from the other end. It was created initially as a response to an emergency because of a media shutdown, and became a hybrid organisation that relies in citizen-generated content.

From the perspective of international relations, Ushahidi deployments take the form of TANs dealing with local crises, which facilitate cross border coordinated action based on analysis, visualisations and interactive mapping that render crowdsourced data useful for decision-making during disasters and humanitarian crises. Ushahidi complies with the Waddell and Khagram model (2007, 263-265). That is: Ushahidi is active globally; it engages diversity (i.e. the platform has been used by organisations and campaigns of very different nature, from the Platform Against Evictions, in Spain, to the Haiti platform, in 2010, to tackle the earthquake

[391] At the time of writing, the organisation is experiencing changes in strategies and may take different paths, though.

aftermath);[392] it facilitates the creation of inter-organisational networks; it has become a systemic change agent, as it has transformed the rules of humanitarian assistance and pioneered digital humanitarianism; it has is learning from errors, sometimes from independent evaluations (Morrow, Mock and Papendieck 2011); it is a coalition of the willing; and ultimately it is a producer of public goods, generating information and transparency in an open access manner.

As noted, Ushahidi is not a social movement; however, it shows some of social movements' characteristics. Four elements are common in definitions of social movements: the use of unconventional means, a network structure, shared beliefs and solidarity, and the pursuit of some conflictual aims (della Porta and Diani 2006, 5, 7,159, 180). I examine these elements in relation to Ushahidi, interpreting della Porta and Diani and based on empirical observation and the in-depth interviews performed for this dissertation. Ushahidi employs disruptive technologies –unconventional means— to a point that has become a form of 'disruptive philanthropy' (Bernholz, Skloot and Varela 2010). It pioneered by utilising existing technologies (i.e. SMS, web postings) in an innovative way to connect real-time reports of incidents on an open and live map. The model was tested in Kenya in 2008, and once its value was proven, it expanded challenging hierarchical institutions with entrenched practices, interests and stakeholders, as it happens with other disruptive technologies (Owen 2015, 7), defying 'the old models, the old analogue models of control and command' (Conneally 2012). Paraphrasing della Porta and Diani, Ushahidi is disruptive in that it obstructs the normal course of events (2006, 174). One example is Ayuda Ecuador. One of the challenges was to feed information into traditional rescue

[392] In 2010, after the earthquake in Haiti, Ushahidi, The Fletcher School of Law & Diplomacy at Tufts University, UN OCHA/Colombia and the International Network of Crisis Mappers (CM*Net) collaborated to set up a Haiti Ushahidi-enabled platform to assist the humanitarian efforts. Nearly 40,000 to 60,000 independent reports were sent to the Ushahidi Haiti Project, of which nearly 4,000 distinct events were plotted (Morrow, Mock and Papendieck 2011).

operations, happening in a fixed and somewhat arthritic way. 'Civil servants are used to doing things in certain ways. They did not have the capacity to absorb 400 reports in two hours and incorporate them in their dynamics… Meanwhile in traditional humanitarian operations a situation map can take a day to be put together,' says Aguilar. Another example is how Ushahidi challenges the state monopoly of maps. Meier says that it enables a form of live-mapped *sousveillance*[393] for civil society. Gathering data and publishing maps –a process of coding and recoding of information, and its synchronisation— were once the exclusive privilege of the state. As Meier argues, the Ushahidi platform serves to democratise *dataveillance* by crowdsourcing and providing 'a participatory digital canvas' (2011a), or an alternative public sphere. It acts as an equaliser: anyone with a telephone or a computer and an internet connection can create and feed an Ushahidi map independently. The Ushahidi Haiti Project, although far from perfect, 'was significant in that it is one small part of a paradigm shift. Lives were saved – this is true and a point that cannot be overstated. And yet many of the significant results of this project are not seen in the immediate effects, but in the long-term influence that the initiative had in proving that open-source, participatory information gathering can work' (Martin 2011). The Global Facility for Disaster Reduction and Recovery Labs agrees. While volunteer technology communities, such as Ushahidi, 'will never replace the current institutional frameworks, crisis response and other elements of the disaster risk management cycle are approaching a state resembling a shift in paradigm' (2010). It has already been determined that Ushahidi does not engage in open protest, and therefore it is neither a reactive data activist organisation (or a *beyonder*) nor a social movement. However, this does not mean it does not embody a challenge to the *status quo* or that it has not created new disruptive forms of action (della Porta and Diani 2006, 29). Ushahidi is

[393] See Mann, Nolan and Wellman 2003, 338-348.

revolutionising pre-existing power structures and dynamics through technical innovation and social progress (i.e. spread of mobile technologies), new ideas and paradigm shifts (Kuhn 1970).

Another common characteristic of both social movements and Ushahidi is a network structure. As noted, Ushahidi counts on a group of dedicated people, but depends too on a network of volunteers that quickly and informally is fashioned right after the platform's launch. This is what Blackadar Nelson, Stamberger and Steckler call a 'hastily formed network' (HFN), which is a rapidly deployable *ad hoc* network that can be generated using a variety of different communication technologies (2011, 1). Boersma et al. talk about 'a novel, net-centric approach to disaster management,' which is challenging 'traditional forms of command and control, through technology-supported, self-directed networks of heterogeneous stakeholders including affected citizens' (2014, 125). As a result, disaster response and humanitarian organisations are gradually seeking to connect their response structures with these shared communication platforms and many-to-many information streams. This is made plain by the growing importance of platforms like Ushahidi. However, professional response agencies differ in their approaches compared to citizen-led, net-centric initiatives, as the example of OCHA shows. Boersma et al., for example, propose that the strongly formalised management structures of the traditional relief agencies relinquish control and command approaches to increase their adaptive capacity and capitalize on citizen-based information generated in net-centric structures (ibid.).

As della Porta and Diani note, it is often the case that a hybrid model of 'network organisation' combines elements of formality 'with those proper to a loose network structure' (2006, 159). Their model of organisation is based on the independence of single components of the structure, their horizontal integration, and the flexibility in goals and strategies, among others (ibid.). This type of networks are flexible and decentralised forms of organising, and can include

relatively bounded organisational forms such as the Climate Action Network, and whole movements, like the environmental movement, as well (ibid., 160). This model propably applies to Ushahidi too. For instance, evaluation documents and case studies show atttempts to integrate different elements of deployments, and formalise expectations and relations on the record while allowing a great degree of flexibility and horizontality.[394] Network organisational models are helpful to coordinate efforts around specific issues in which big numbers of volunteers and organisations collaborate, and they do so while being neither reliant on the organisations that initially launched them, nor capable of exercise control outside the limits of their sphere (ibid.). They allow 'some kind of mediation between the participatory ethos behind grassroots organising and the coordination guaranteed by formal structures' (ibid., 161). An example is the Kenyan elections in 2013 deployment, which was set up by Ushahidi members, with the collaboration of several institutions (Hivos, Canadian International Development Agency, USAid, Mercy Corps, NSC, Kenya Red Cross, CHF International, PACT, Constitution and Reform Education Consortium, Peace Net and Catholic University of Eastern Africa) and 'a large number of volunteers' (Cio 2013). Many network organisations are 'inherently temporary,' that is they do not outlive the specific mobilisation they are supposed to coordinate, although some of them can become independent entities with a distinct identity (della Porta and Diani 2006, 160). The deployment to tackle the Kenyan referendum of 2010 produced an ICT platform, called Uchaguzi –in which Ushahidi, Hivos, CRECO and CIDA, among others, collaborated—, to provide a permanent monitoring tool for citizens, civil society, election observers, law enforcers and humanitarian agencies. Some of these institutions were summoned and assembled again later in other deployments to monitor elections, including the 2013 Kenyan elections, but the exact composition of each partnership does not survive the specific

[394] See the section on controversies and criticism.

mobilisation they coordinate. Uchaguzi has been used in the monitoring of the elections in Tanzania (2010) and Zambia (2011) as well (Omenya 2013).

Ushahidi is a case of an organisation mobilising on a transnational scale with a network form too (ibid.). Its deployments usually are able to rally international networks of expert volunteers. Most of the issues they tackle are local in nature and therefore do not expand beyond national frontiers (i.e. Syrian Tracker, launched by Humanitarian Tracker in 2011). However, some of them expand outside national borders (i.e. Narrative Atlas,[395] which connects students and teachers around the globe, with MapWorks Learning and Teachers Without Borders).

Common beliefs 'on which to base foundations for collective solidarity' (della Porta and Diani 2006, 7) is another characteristic that Ushahidi shares with social movements. In Ushahidi deployments, a local cause triggers international solidarity that feeds support back to the local event (Grabowski 2013). A shared system of beliefs and a sense of belonging are behind some of the most successful Ushahidi deployments. An example is, again, the 2010 Haiti platform. Meier, a graduate of the Fletcher School of Law and Diplomacy, was able to mobilise dozens of volunteers at his *alma mater*, most of them from the Haitian diaspora, who were then crucial in translating SMS messages[396] and providing localised information (2011a). Actors with different identities, nationalities and orientations also come to elaborate interactions that bound them around other Ushahidi deployments. For example, Ayuda Ecuador gathered organisations and experts from Ecuador, Colombia, Kenya, South Africa, UK and US. However, when there are low levels of sense of belonging and buy-in is when deployments tend to fail.

[395] See narrativeatlas.ushahidi.io/views/map [accessed on August 11, 2016]. However, these initiatives seem to gather less interest and generate few reports.
[396] Machine translation engines for Haitian Creole were not available at the time.

Collective identities are also created around the technology employed at Ushahidi, and in fact this is one of the features that define proactive data activists. ITCs employed to mobilise and coordinate action, connecting and uniting diffuse actors with common interests (Shirky 2008, 45) are core features in all Ushahidi deployments. And its origins start with four 'tech-savvy activists' who put their skills at the service of social change (Milan 2014b). The networks of experts gathered around these deployments speak the same language of technology as well. In fact, what Ushahidi practices is a kind of 'tech activism' (ibid.) applied to humanitarianism.

Finally, the pursuit of conflictual aims at Ushahidi is to be understood as a challenge to the *status quo* and the use of disruptive means. As established, the nature of Ushahidi is not confrontational, even if Ushahidi deployments are often tossed in the middle of conflictual situations. Van Niekerk and Maharaj mention Ushahidi in the context of *information conflict* or both military and non-military applications of information warfare tactics, including strategic information security and influence operations (2013, 1163-1164). Digital technologies can generate political disturbances and advocacy 'to sway the perception of the general population' into action (ibid., 1164). Van Niekerk and Maharaj talk about shepherding support through digital technologies for physical protest action against a government, but in the case of Ushahidi this support is addressed at creating alternative public spheres of participation and collaboration. Table 4 summarises how Ushahidi appears seen from these perspectives.

Table 4: Ushahidi seen from different perspectives

Perspective	Description
Critical thinking	Ushahidi initiatives involve communicative actions and generate digital public spheres.
Journalism	The values of objectivity, fairness and accuracy of good journalism are embedded in Ushahidi's DNA. Ushahidi is born to end an information shutdown; uses the journalistic method of verification;

	and is employed regularly as an information enhancer in journalistic reports.
Alternative media	Ushahidi creates an alternative discourse around armed conflicts and can be considered as a manifestation of *subaltern counterpublics*.
International relations	Ushahidi deployments generates TANs that deal with local crises, coordinating action based on analysis, visualisations and interactive mapping that render crowdsourced data useful for decision-making.
Social movement	Ushahidi shows some of social movements' characteristics: the use of unconventional means (disruptive technologies), a network structure, and shared beliefs and solidarity, although it does not have a confrontational nature.

Source: Elaboration by the author.

7.3 Ushahidi's brand of proactive data activism

Ushahidi deployments chart catastrophes, build communities and trigger action. Using the categorisation emerged in this study, it is a geoactivist endeavour, although it shows some traits of the match-maker and catalyst, as it seeks to transfer skills through training and form alliances, as well as facilitate activism by providing the resources to do so, as it will be examined. Its attributes include that it always produces of proactive data activist content or projects, mainly creates data platforms and software for activism, facilitates data projects, works mostly in collaboration with others, and occasionally engages in advocacy exercises. Ushahidi works under a Lesser General Public License (LGPL),[397] which means anyone can download the source code and create their own implementation. For those who do not have the time or skills to set up their own customized platform, Ushahidi offers Crowdmap, a free online service that provides a pre-built version of the software. Open data are something else, and the crowdsourced data that Ushahidi generates and uses is contributed by people who are facing danger. By 2015, the question had not been settled. Terp, the then Director of Data Projects at Ushahidi, says that

[397] The open source code that powers the latest version of Ushahidi is available to download at GitHub. The LGPL is a free software license published by the Free Software Foundation (FSF), which allows other developers and businesses to use and integrate software released under it into their own (even proprietary) software without being compelled by the terms of a copyleft license to release the source code of their own components [see fsf.org/?set_language=es, accessed on August 3, 2016].

Ushahidi was still pondering ‘about what it means to balance the potential social good of wider dataset release with the potential risks that come with making any data public’ (2015). ‘Many of the datasets managed by Ushahidi users contain information that is personal, often gathered under extreme circumstances, and potentially dangerous to its subjects, collectors or managers,’ so when making them public one has to be sure, ‘to the best of your knowledge, skills and advice,’ that no harm will come to the people who are connected to the dataset (ibid.).

As said before, proactive data activists do not conform to the description of the beyonders. Although it employs disruptive technologies, this is clear by examining how Ushahidi behaves in relation with local authorities, as it lacks the confrontational, out-of-the-system nature that defines beyonders, who ‘seek to act regardless of the logics and rules of the game of the known social system’ (Milan and Hintz 2013, 20-21). Ushahidi is more concerned with enhancing and acting than with shielding or protesting. Ushahidi’s Director of Community Engagement Oduor, interviewed for this dissertation, says that her organisation ‘encourages more collaboration (with authorities). Our goal is not to be an antagonist.’ In this light, Ushahidi presents itself as an instance of an inclusive platform for empowerment, decision-making and action, which enhances existing mechanisms and creates new ones to enable people to participate in decision-making processes and action in times of crisis. Ushahidi was created because there was a gap, because nothing like that existed before; but not as a rejection of existing institutional venues and norms. A social challenge is embedded in this very gap-filling action, but it is not at the heart of Ushahidi’s vision. Oduor puts it like this: ‘Based on crowdsourced data, our mission is to ensure that people’s voices are being heard, and that decision-making is enabled for positive action in cases of disaster and crisis.’

Within geoactivists, Ushahidi deployments exemplify how data can be generated for geoactivism via *crowds*. Ushahidi offers a platform that enables local observers, witnesses, *victims* and humanitarian agencies to submit reports using their mobile phones or the internet, and access the information, while simultaneously creating a temporal and geospatial archive of events. The organisation uses the concept of crowdsourcing for social activism and public accountability, drawing on previous experiences of social activism, citizen journalism and geospatial developments.

Ushahidi deployments illuminate the three main advantages, compared to conventional relief methods, of leveraging crowdsourced data for disaster assistance identified by Gao, Barbier and Goolsby (2011, 11). The first one is the immediacy of data collection. An example, again, is the Ushahidi Haiti deployment, which 'was set up two hours after January 12 (2010) earthquake by volunteers from Tufts University in Medford, Massachusetts. Soon after, organisations were able to borrow a SMS short code phone number (Mission 4636) to send free SMS texts' (ibid.). Instantly, news of the free emergency number spread out through local and national radio stations, and by within days, the Haiti crisis map had more than 2,500 incident reports. 'The large amount of nearly real-time reports allows relief organisations to identify and respond to urgent cases in time' (ibid.). A second advantage comes from a macro perspective, that is, the use of big data infrastructures as *telescopes*. Crowdsourcing tools 'can collect data from emails, forms, tweets and other unstructured methods, and then do rudimentary analysis and summaries... These can help partition the data into *bins* (such as most-frequently requested resources) and requests into predetermined, most-urgent categories (such as medical help, food, shelter or people trapped)' (ibid.). Humanitarian agencies can then focus on the matters that seem most urgent from the visualisation of data intensity and concentration of geolocalised reports

(ibid.). A third advantage comes from a micro perspective, offered by big data infrastructures used as *microscopes*. Since providers can include information sent from any platform (i.e. Twitter) and device (i.e. smartphone), relief organisations can 'accurately locate specific requests for help' (ibid.).

The micro perspective is also changing the way conflict analysis is dealt with until now. Meier says that this type of study was limited by macro-level data framed by a country-year logic (2011b), but new ways to render data allow for subnational, sub-annual analysis of data, leading to new insights into conflicts. He refers to a study of the Gaza conflict that draws on hourly temporal data offered an *Al-Jazeera's* "War on Gaza" project powered by the Ushahidi platform (Zeitzoff 2009). Using a dataset of hourly conflict intensity scores from Twitter and other social media sources, Zeitzoff tested the theory's implications to find that over the course of the conflict (2008-2009) Israel responded more to Hamas provocations than vice versa (ibid., 28-30). With its formula, Ushahidi has pioneered in both 'crisis mapping' and 'activist mapping.' Cases of both can be found among deployments of the Ushahidi platform. The former is simply 'the act of receiving reports… and plotting them on a map,' as Mitchell defines it. Meanwhile, activist mapping is a longer term activity focused on gathering bodies of evidence that can be used for lobbying and advocacy. As Were admits in an interview for this dissertation, 'technology is just the 10% of the solution; the other 90% is in our partners' hands... We are increasingly becoming more and more involved in that 90% of the work.' That is why Ushahidi is working on anti-corruption mapping, supporting anti-corruption advocacy in different countries around the globe, engaging more and more in the policy and advocacy side of things, attending more conferences and taking part in regional processes.

Summarising, Ushahidi is a geoactivist endeavour which produces of proactive data activist content or projects, creates data platforms and software for activism, works mostly in collaboration with others, and occasionally engages in advocacy exercises. Within geoactivists, Ushahidi deployments generate data via *crowds*. Ushahidi has pioneered in short-term 'crisis mapping' and longer-term 'activist mapping,' and is increasingly involved in campaigning.

7.4 Repertoires of action

In order to power early warning systems for rapid response to humanitarian emergencies, crisis mapping applies different tools and methods, including participatory maps and crowdsourced data, geospatial platforms, integrated mobile applications, aerial and satellite imagery, visualisations, and computational and statistical models. Ushahidi was among the first to use these tools for humanitarian purposes. Instances of crisis mapping usually have the aim to process and produce quick information that is of value in times of crisis. Some of the repertoires of action of crisis mapping, also employed by Ushahidi, include: a) geolocalising news reports, providing the identification of the geographic location of an event (Birregah et al. 2012); b) categorising and translating text messages to facilitate communication and understanding between witnesses and assistance providers, or among the former (Meier 2012, 100); c) scraping social media sites for crisis-specific keywords (i.e. gathering and visualising tweets that have a specific designated hashtag so a comprehensive picture of an emergency situation emerges) (Rafoth 2015; Terp 2015); and d) locating refugee camps, buildings, roads, water distribution, evaluation points and logistics on aerial imagery (Parker 2015), as it will be seen in the case study of this dissertation.

Its repertoires of action are important too as a unifying element. Ushahidi gathers informal, horizontal networks that work across borders remotely. Collective identity can come from shared

objectives, but is also created on the basis of the data-based tools they employ to achieve those objectives. Ushahidi's platform is what clearly identifies the projects that use it, which are associated with it by using its open source software platform as an enabling vehicle for data analysis, communication and action. In digital humanitarianism, this is called interactive 'activist mapping' or 'critical cartography.' This is a form of 'tech activism,' or activism that 'concerns itself with creating and shaping digital technology infrastructure and tools to facilitate communication and networking for social change activists, and free likeminded individuals from the constraints and threats of commercial communications' (Milan 2014b).

7.5 The empowerment process at Ushahidi

The stated Ushahidi's mission was 'to change the way information flows in the world and empower people to make an impact with open source technologies, cross-sector partnerships, and ground-breaking ventures' (HumTech 2015; Michael 2014).[398] Were talks about 'civilian empowerment' (Were 2015). The empowering process in the use of the Ushahidi platform comes in different ways, and one of them is numbers and the community, what Sebastian Mitchell calls the 'power of a crowd… saying the same message.' According to him, one thing Ushahidi team members do in speaking events is to ask people in the audience to say things that come to their heads individually. Then, they ask the audience to say the same things all at once. 'That illustrates our point. If many, many people are saying the same things at the same time, the message is very, very loud and very clear, much louder that would ordinarily be. So we really think that ability to enable people to say the same messages at the same time and have their voices heard collectively and shared with the world is empowering to groups' (ibid.). When facing a disaster or reporting violence, in the middle of the chaos and destruction, the individual

[398] This statement is no longer exhibited on Ushahidi's website at the time of writing.

gets encouraged when knowing he or she is not alone. From being a mere map where an electoral crisis got georreferenced and describing itself in technological terms, Ushahidi has been transformed into a social network of information generared by a mass of people involved in the same event or concept and building community (Sandoval-Martín and Espiritusanto 2016, 469; Keim 2012). But how does this community get built? Indeed, Ushahidi deployments usually generate several communities; that is, the group of organisations setting up the deployment, the techies and experts supporting them from other locations, the humanitarian agencies resorting to the information released by the deployment and the people on the ground affected by the disaster or violence and reporting. In the case of Ayuda Ecuador, for example, the deployment was put together by volunteers from a bunch of organisations from different countries, including FLOK Society,[399] Soporte Libre,[400] Asociación para el Progreso de las Comunicaciones,[401] ThoughtWorks,[402] Ayni Consulting,[403] and supported by three experts –Aguilar, on knowledge management in humanitarian operations; Herrera, on communications and software development; and Navarrete, on mobile software development—, with help from Oduor, from Ushahidi.[404] The deployment generated 410 reports from April 16 to May 5, 2016, including people looking for other missing people[405] and lists of deaths.[406] And this happens on the basis of trust and legitimacy. As the mentioned case of the UN Stabilisation Mission (MONUSCO) in the Democratic Republic of Congo suggests, in order to utilise crowdsourcing to generate data,

[399] See floksociety.org/ [accessed on August 6, 2016]. Based in Ecuador.
[400] See soportelibre.com/ [accessed on August 6, 2016]. Based in Ecuador.
[401] See apc.org/es [accessed on August 6, 2016]. An international organisation based in South Africa.
[402] See thoughtworks.com/ [accessed on August 6, 2016]. Based in the US.
[403] See ayniconsulting.com/ [accessed on August 6, 2016]. Based in the UK.
[404] Escuela de Datos (School of Data), the Humanitarian OpenStreetMap Team (HOT) and OpenStreetMap Ecuador are also involved in this development, although they do not explicitly say so in their communications.
[405] See mapa.desastre.ec/reports/view/830 [accessed on August 14, 2016].
[406] See mapa.desastre.ec/reports/view/145 [accessed on August 14, 2016].

motivated and engaged crowds are needed; and to count on crowds, legitimacy and a track record of fulfilling expectations has to be in place first. At the end of the day, this is about individual human beings sharing vital information about their whereabouts and circumstances.

Crowds can be leveraged as knowledge generators, as sources, as analysts and as communities (Pawlak and Ricci 2014). Ushahidi has mastered the employment of crowds as knowledge generators, (i.e. allowing the population to geo-tag and map reports from Twitter and categorise them by types of available resources, such as evacuation shelters, operating food stores, petrol stations, and locations for charging mobile phones), sources (i.e. crowdsourcing services deployed to support the disaster response and reconstruction during the first moments of a crisis) and communities (i.e. the Standby Task Force). But as seen in the example of the Indonesian drones, crowds can be analysts and verifiers too.

The technology is also liberating. In a McLuhanian interpretation of how this works, Meier puts is like this: 'The Haiti story is, without doubt, just as much a human story as it is a story about new technologies. Come to think of it, these technologies actually make us more human… As digital humanitarians in Haiti ably demonstrated, these new connection technologies can extend and amplify our humanity, can translate our initial private emotions of sadness and powerlessness into public—indeed global—action to help others thousands of miles away' (2015a, 18). Through technology, Ushahidi creates the space and, equipped with self-generated data and knowledge, activists and disaster *victims* generate their own alternative public spheres for participation, collaboration, decision making and action in rescue and humanitarian operations. As observed earlier, big data infrastructures used in Ushahidi's deployments allow for emancipatory practices at an unprecedented scale, and with unparalleled levels of autonomy and self-sufficiency, which signify a reconfiguration of power (Milan 2015, 120). They show

that an alternative, safer space for communication and democratic participation is not only possible, but probably also desirable (ibid.).

A more complex way in which people are empowered using Ushahidi deployments is by transforming *victims* in proactive reporting witnesses. The asymmetrical, political economic dimensions of the production and distribution of data (Pybus, Cote and Blanke 2015, 1) are reversed here by victims generating their own data, and getting control and seeking to understand the data they collectively generate. This action, not only augments their agency as individuals but as community (ibid.), since agency is directly connected to the distribution of knowledge and power (Baack 2015, 1). Among other considerations, Burns considers that big data 'should be understood as a new set of practices' that is 'constitutive of a social relation in which both the formal humanitarian sector and victims of crises are in need of the services and labour that can be provided by digital humanitarians' (2014b, 1). Some of the social changes that Ushahidi has generated were not in their specific objectives or even in its vision, in the same fashion that some technological innovations have unanticipated impacts, and generate new 'irritations' and stresses (McLuhan 1964, 53-54).[407] Ushahidi deployments allow individuals and groups to access practical information about how to behave and what to do in a situation of emergency. Datafication should not only be understood as the process of collecting and analysing data by companies, but also the appropriation of one's own data (Kennedy, Poell and van Dijck 2015, 1), which is what happens with Ushahidi's deployments. Although these authors were speaking about data collected for example by internet and mobile phone service providers, the concept

[407] In the case of the car, for example, which replaced horse-drawn carriages, transportation patterns and logistics, the layout of entire cities, and the movement of people were deeply transformed because automobiles could go much farther and faster. The cycle is completed as these social preferences and institutional changes in turn moulded the design of next generation automobiles. 'A second source of unexpected impacts is scale. A single car doesn't pollute significantly; but millions of cars can transform the atmosphere' (Ito et al. 2008).

works to explain how Ushahidi allows alternative forms of datafication and cognisant and teleological agency that empower people in a way that lets them make sense of the world *so as* to act upon it (Couldry 2013, 13). Witnesses using the Ushahidi platform undergo this process: by participating actively and acting upon their own situations, they change communicative processes, generate alternative public spheres and experience an empowerment process that enables them and others to make decisions and act. This process is democratizing too. Although no commonly accepted definition exists, democracy is about access and equal opportunities. Using Ushahidi deployments, willing to take part in public debate, people exercise their responsibility, solidarity, duty, civic sense and awareness of community belonging, values that the makeup of democracies. 'Ushahidi democratises access to information in crisis situations: every post, text, photo and video is made visible on a map, empowering people to use this primary sources to create their own multiple narratives' (Multiple Journalism 2008).

7.6 Controversies and lessons learnt

Ushahidi' platform has evolved into a social network around citizen-generated vital information about an event or issue. But each deployment requires enthusiastic cooperation and buy-in, careful adaptation of the platform to a local situation and its environment, and verification of the information provided, as well as the protection of the individuals providing it. Overtime, these issues are the source of tensions and controversies, which are examined in this section. The impacts and evaluation of Ushahidi deployments are not always clear either.

A number of Ushahidi deployments have been set up and then abandoned right after (Slater 2014). 'At face value, this can be thought of as a criticism of the Ushahidi platform, but if you dig deeper, what you find is that lots of folks have quickly jumped to deploy Ushahidi because: a) it's gotten a lot of publicity, and b) it's easy to install and get going. But what a lot of folks

haven't thought through are questions about how mapping can support their own campaigns and missions' (ibid.). Leson points out that groups setting up deployments should start by asking themselves first: 'why a map?' and see whether that instigates a conversation about how the map fits into their efforts (2016). Before attempting any deployment, Chamales recommends asking key questions including 'what are you trying to accomplish with this deployment?' 'what information do you want to collect?' 'who will submit that information?' 'how will it be submitted?' 'who will monitor and approve reports?' 'how will you verify the reports?' 'who will use the data?' (2013).

A 2012 quantitative analysis that provides a baseline, indicates, for example, that, based on almost 13,000 deployments,[408] 93% had fewer than 10 reports while 61% were had not been customized or used at all (Crowdglobe 2012, 5). The study finds that about 16% of the users who launched the map felt that 'they were not able to generate the required public awareness vis-a-vis their map to make it as effective as they had hoped' (ibid., 6).[409] Commenting on this report Meier highlights that by 2013, there were more than 36,000 deployments (Meier 2013b), and Were reports 90,000 deployments at the time of writing (2016). Meier suggests that Ushahidi

[408] This was a study of the usage of Crowdmap [see crowdmap.com/welcome, accessed on August 23, 2016] a free online service that provides a pre-built version of the Ushahidi software, which was reintroduced in 2014.

[409] Other interesting findings include that 30% of the deployments had been originated in North America, 18% in Western Europe and 16% in Africa (ibid.). 'The most common themes that emerged from the 63 deployments in the Middle East and Northern Africa pertain to crime and public safety issues (43%), human rights abuses (40%), emergency-related infrastructural issues (30%), and political organization (25%). The distribution of themes in the 79 Western European deployments, on the other hand, paints a very different picture, with entertainment and leisure appearing in 32% of the deployments, followed by non-emergency infrastructural issues (25%), and media reports (23%). It is also not surprising that the 16 deployments from the Caribbean region, 12 of which hailed from Haiti specifically, heavily featured issues related to the occurrence and aftermath of a natural disaster (63% and 50%, respectively), emergency-related infrastructural issues (63%), health and medical-related issues (50%), and crime and public safety issues (38%)' (ibid., 5).

should embrace open data totally to dispel rumours and criticism, by updating the Internews analysis. The analysis, to my knowledge has not been repeated (2013b).

Like other organisations, Ushahidi combines data analysis and emotion –which need each other in effective activism. Empathy and emotion are required to generate reporting crowds, and expert and non-expert volunteers; thanks to the convening powers of collective emotions, data are gathered to paint the big picture of catastrophes and disasters.[410] Failure to motivate and generate communities of users for each deployment prompted a 'Dead Ushahidi' map,[411] a 2012 (crowdsourced) map showing maps that did not succeed, powered by Ushahidi. 'Dead Ushahidi' includes maps such as 'Honduras Health Mapping,[412]' 'Elections Cote d'Ivoire'[413] and 'Wildlife Tracker in Kenya,'[414] among 30 others. These maps contain few reports in comparison with relevant population or just go quiet. And emotions can be cajoled and strategized. 'Using crowdsourcing tech like Ushahidi maps without doing the strategic and programmatic ground work is likely not going to work or change much of anything. Trying to crowdsource a map without a goal or strategy is well, just a map, and pretty soon a dead map' (Vota 2012). Vota concludes that, to succeed in the art of crowd-powered geoactivism, it is vital to think how a crowdsourced map will advance goals, to get good data, to have a clear focus for report collection, so as to generate or tap into a community (ibid.).

---

[410] As noted before, data infrastructures are telescopes, microscopes and scalpels (Tufekci 2012).
[411] See deadushahidi.crowdmap.com/ [accessed on August 16, 2016].
[412] See deadushahidi.crowdmap.com/reports/view/26 [accessed on August 16, 2016]. The site includes this comment: 'Either Honduras is populated by citizens with mutant immune systems or the health map with no reports since July 2011 is dead' (Dead Ushahidi 2012). The map is no longer available, but it was tagged as verified in the Dead Ushahidi map.
[413] See deadushahidi.crowdmap.com/reports/view/21 [accessed on August 16, 2016]. 'With 5.3 million voters on the rolls and 80 reports on the map, this map was officially dead on arrival' (Dead Ushahidi 2012). The map is no longer available, but it was tagged as verified in the Dead Ushahidi map.
[414] See deadushahidi.crowdmap.com/reports/view/18 [accessed on August 16, 2016]. 'With 40 reports, and a dead link, this wildlife tracker was dead on arrival' (Dead Ushahidi 2012). The map is no longer available, but it was tagged as verified in the Dead Ushahidi map.

There are two sorts of crowds that are important: the expert volunteers and the non-experts. Kenya's Ushahidi endeavours count on the deep roots and connections with networks of techies, bloggers and activists the organisations has in Kenya, as well as the possibility to utilise the iHUB space,[415] for example. But this is not the case everywhere. Community building and co-working spaces, where policy makers, developers and civic hackers can gather together, chat and build together, such as the iHUB, are fundamental. The second crowd, the final non-expert users that will report the data to populate maps depend on media outreach, publicity and dissemination strategies typical of advocacy organisations. Ushahidi platform includes manuals and advice for developers on how to attract and share with a crowd, use social networking, phone messaging and blogging, build a community, reach out to the media, seek partners and produce calls to action.[416] Deployments, in fact, often include people specialised in these tasks, including a media and outreach coordinator and a volunteer and training coordinator (Ushahidi 2011).[417] But as seen, not all deployments count on a crowd, and therefore die of inactiveness.

Another important aspect is buy-in among humanitarian agencies working on the ground. This is another challenge. Barriers to use of the information provided by the Haiti deployment were the discrepancies between the dynamic event data aggregated and syndicated by the platform and 'the specific and often relatively rigid information requirements of traditional responding organisations' (Morrow, Mock and Papendieck 2011). Aguilar notes that, even if at the time OCHA's head, Verity, was in the On-Site Operations Coordination Centre *(*OSOCC*)* –

[415] By 2014, the lab has more than 11,000 members, and has incubated 48 start-ups to date (Granger 2014). The iHUB is an independent Ushahidi space for co-working, to which Ushahidi has free access.
[416] See ushahidi.com/support/outreach [accessed on August 17, 2016].
[417] The people who need Ushahidi typically fill also some of the following roles: Executive, making decisions; project manager, setting up the deployment; developer, installing and customizing the deployment; operator, managing data and tasks; gatherer, summiting posts as team member; responder, responding to posts in the field or virtually; reporter, submitting posts as a citizen; receiver, receiving aid; viewer, visiting the deployment (Rosage 2016).

where decisions are made and rescue teams are coordinated in every disaster or crisis—, he did not gathered immediately how Ushahidi could be utilised. This disconnection continues to be a challenge. During the earthquake in Ecuador, Aguilar says: 'it was very difficult' for local authorities to accept the reports sent to the platform and integrate them in their efforts. The adaptation of the platform to local realities, however, is proved vital for their success. Although other factors may have been at play,[418] the DR Congo 2011 elections deployment was a failure. Okolloh admitted that the lack of proper localisation of the deployment contributed to the fiasco (Ekine 2009).[419]

It is clear that democracy is not only about mobile phones, apps and maps. Wrong, among others, rejects 'the techno-election as democratic panacea,' which she labels 'a glitzy red herring' that 'cannot replace a society's generalised buy-in to the democratic process' (2013). But for platforms, such as Ushahidi, to work, an engaged blogging and activist community, a degree of freedom, and timely translations have to be in place beforehand. 'Ushahidi worked well in Kenya because of the maturity and community of Kenyan bloggers. In addition, many

[418] Ruffer, who led ad team deployment for the DR Congo elections in 2011, wrote that everything started to go wrong when the government shut down the SMS network due to concerns of violence, people were asked to stay home and foreigners evacuated Kinshasa (2012). Volunteers could not send any reports. She admits that Ushahidi was not widely known in the DRC and the website was not always available due to slow internet connections. People were also weary of sharing data with Ushahidi given security concerns. Human rights and humanitarian organisations –who needed reports in real time to respond to crisis situations— had different expectations than official monitoring efforts –who were interested in postponing the release of information until they could better assess and negotiate the political implications— (ibid.). She concludes that 'although more time to conduct networking on the ground would have produced more buy in from humanitarian, civil and human rights organisations, it seems that in the end we might have had the same problems given the SMS and need of the largest observers to withhold their reports until the best strategic moment' (ibid.).

[419] Okolloh's blog post is no longer available.

Kenyan bloggers possessed considerable technological skills — which meant they could see the potential of a site like Ushahidi' (Ekine 2009). However, the DRC was a different story.[420]

Another problem of emergency response is the lack of localised guidelines to customise the platform. In Ayuda Ecuador, 'we did not find manuals and recommendations in Spanish to share with the volunteers, or the materials were not easily replicable. We found some stuff in Spanish, but it was scattered' says Aguilar. Coordination between the expert volunteers who can be working remotely and local experts responding to the emergency on the ground is vital as well. 'The local rescue experts who respond to the emergency may ignore the role digital volunteers are playing in a crisis. When some of these guys, who may have been working in crises for twenty years, hear the word "volunteer", they imagine the typical "emergency tourist" who takes pictures, raises funds for his or her own cause and gets back home contributing nothing to the overall effort.' Addressing mistrust and misinformation that can be present when dealing with local realities and foreign support is, therefore, vital too.

Having access to correct information in a crisis leads to targeted actions for both grassroots support and emergency response professionals. If crisis maps are based on a collection of reports from a number of sources, there has to be a system for verification that can stand up to scrutiny, especially in situations of violence where false entries can trigger more violence. Ushahidi founders realised this was the case early on, and Were joined them with the mission to put in place a system to verify eyewitness testimonies (Smith 2008). Although anyone can start an Ushahidi map and send a report to the map, only administrators can approve and publish reports

[420] DR Congo is ranked 152nd in the World Press Freedom Index (Reporters Without Borders 2016). 'Journalists are exposed to threats, physical violence, arrest, prolonged detention and even murder. The main perpetrators –the army, police and security services– enjoy complete impunity' (ibid.). DRC is labelled 'not free' by Freedom House, while Kenya is deemed 'partly free' (Freedom House 2016).

or tag a report as verified. However, this system has not forestalled a number of glitches. Ford recalls how, during the Chile 2010 earthquake, two false reports received by the Ushahidi team prompted a rescue police operation before they were investigated and found false (2011). She suggests more flexibility to allow platform users avoiding the 'verified' and 'unverified' tagging functionality, since 'unverified' may suggest that the report is unverifiable or false, or that the corroboration process is incomplete; contextual information –containing the 'who, what, where, when, how and why of traditional journalism'— to allow users determine whether a report is trustworthy or not.

But what if thousands could be mobilised to cross-reference and authenticate content shared on social media during a disaster? Ford speculates that not only the collection of data could be crowdsourced, but the process of analysis and verification as well. That is, 'the crowd (i.e., people who are not known to the system) might assist in either providing more context for reports or verifying unverified reports' (ibid.). For example, Valuch says that, while mapping the floods in the Czech Republic in 2013, the responsibility of verification fell into crowdsourcing naturally, as people started correcting some entries and providing additional information and evidence (2013). How to crowdsource the verification process is an area where there has been some exploration, inside and outside Ushahidi. Meier writes about experimental research on how to crowdsource verification tasks combining best practices with gamification and reputation mechanisms 'to leverage the good will of (hopefully) thousands of digital Samaritans during disasters,' and the integration of metadata of images and videos taken by eyewitnesses (2013c). Meier talks too about how artificial intelligence —such as data mining and machine learning— and advanced computing is being used to make sense of vast volumes of text messages in response to crises or to deal with satellite and aerial imagery (Meier 2015e).

Together with the opportunities for digital humanitarianism and data activism, the use of technologies also generates new threats for both human rights defenders and witnesses, since these reports can reveal their whereabouts, activities and networks to direct surveillance and interception (Hankey and Clunaigh 2013, 536). In doing so, social media users' safety might be compromised. Starting with the Haiti deployment in 2010, Ushahidi was criticised for not protecting its reporters. Meier recalls how, before making public personal information for that deployment, they consulted with experts who suggested reporters were giving implied consent by reporting information, and how the team agreed that they would do more good if the data were kept open (2013d). Haiti was the starting point for many things, but also for a big debate on security, what procedures could protect it and several experimental approaches. Anonymity is not even enough to guarantee the safety of a reporter, since it is not a dependable means of protecting identity –especially if that information includes location data— because some identifiers are not connected to names (Slater 2014). Security is an issue that needs to be integrated across every step of the reporting process, including making the information available, to avoid putting people at risk. For the Syria Tracker deployment, reporters would be given instructions on how to mask their identity and location, which guaranteed that, by the end of 2012, it had compiled reports of 47,887 deaths culled from thousands of emails, tens of thousands of news reports and millions of tweets (with the hashtag #basharcrimes) (McConnell 2013). The new version of Ushahidi includes also new features to manage the security of sensitive data. Terp suggests that before opening data, a data risk analysis should be done anticipating what could happen and how bad the consequences would be for the people connected with them, including reporters, population in general, mappers, organisations and leads (ibid.).

Ownership of the datasets is also a question, which may depend of the type of data. Terp differentiates between 'direct reports,' or messages contributed by reporters via web or SMS (with standard fields such as title, description, category list); 'indirect reports' are messages scraped from other applications, such as Twitter or Facebook, either through APIs or by reporters adding them as direct reports, and geolocations (latitude and longitude for each location name in a list); 'category lists' that reports are tagged with, usually created by the administrator; and 'media,' such as images, video footage and audio files that can be added to a report by the reporter or the administrator (ibid.). Whoever owns the data, Terp notes privacy concerns (i.e. accidentally making a phone number public), safety concerns for reporters contributing information about violence and safety risks from revealing sensitive locations (i.e. addresses of places like rape crisis centres that need to be kept secret), among others. Ushahidi goes by this rule: 'Assessing the potential risks in sharing a dataset; selecting which data you should and should not share. Balancing the potential harms from sharing information from the deployment against the potential good. If you're not sure, don't share, but if you've checked, cleaned, double-checked and the risk is minimal… seriously consider it. If it's come from a personal source (SMS, email etc.), check it. At least twice' (ibid.). Trustworthy repositories, specialised in social datasets, include the Humanitarian Data Exchange[421] and like Datahub.[422] Ushahidi data can also be shared by making it public on an Ushahidi instance with an API (i.e. crowdmap.com) or CSV download button (this is an Ushahidi plugin), or by making data available to people on request. Terp advices to look for the following information in order to start evaluating risks: identification of reporter (phone numbers, email addresses, names, personal addresses); military

[421] See data.humdata.org/ [accessed on August 25, 2016], which specialises in disaster-related data.
[422] See datahub.io/ [accessed on August 23, 2016]. Datahub is a free data management platform from the Open Knowledge Foundation, based on the Comprehensive Knowledge Archive Network (CKAN) data management system.

information (actions, activities, equipment); uncorroborated crime reports (i.e. violence, corruption); inflammatory statements; and level of veracity (asking questions such as are these reports supported by external information?) (ibid.).

Other lessons learnt include the fact that similar catastrophic events can generate different needs. 'In Haiti the search and rescue needs had to do with food and water security, and health issues. In Chile, people wanted to know where they could find open shops to avoid ransacking businesses in search of food and water, and where they could contribute or find groceries and other supplies,' says Aguilar. Mitchell, from Ushahidi, admits that he is working on integrating tools, such as encryption and face recognition protection, in order to minimise the risks for witnesses reporting and providing information.

Finally, the impacts of Ushahidi deployments are not always clear at a granular level. An evaluation of the deployment in the Kenyan referendum in 2010 notes that, overall, the project was a success, that the collective action provided a communication channel for Kenyans to share information, and that it also enabled some organisations to take immediate action based upon the information received (Knight Foundation 2010). For example, of 1,500 reports, 149 resulted in 'actions taken' primarily through one of the partners, the Constitution & Reform Education Consortium (CRECO) network (ibid.). But what does it mean? If only less than 10% of the reports resulted in 'actions taken,' is this a normal outcome? What sorts of actions were taken? It is standard that the rest of the reports did not require any action? Why then register them? The fact that many deployment simply disappear from the internet galaxy for lack of maintenance makes is difficult too to study them. At the time of writing, Ushahidi does publish useful commentary by team members and deployment users, which gathers information about how they were put in place, and what the challenges and what the results were. But it does not seem a

standardised practice that serves as a monitoring and evaluation tool. It also gathers lessons learnt from previous experiences, but it is not clear whether there is any supervision to make sure that mistakes are not repeated. For example, Aguilar says that Ayuda Ecuador had initial difficulties to customise the platform at the beginning due to the lack of guidelines in Spanish.

Another evaluation of the Kenyan elections in 2013 deployment, carried out by the iHUB and looking at six critical issues,[423] found that not all previous recommendations emerging from the 2010 deployment had been put in place, including planning early (six to twelve months prior) ;[424] building effective partnerships so roles and expectations are agreed in advance; developing a strategy that takes in account the campaign, feedback to action and security matters to improve verification, strengthen feedback loops by building an urgent response team, and provide security for the project and its users; and performing simulation exercises (Cio 2013). In spite of having set up an ecosystem of pertinent partners –from funders to response and support providers— a year in advance, the evaluation found that, for example, there were neither memoranda of understanding signed between partners nor regular meetings during the deployment to coordinate things; and other peace monitoring initiatives in place during the elections were not engaged either (ibid.). The absence of an overall publicity and outreach strategy resulted in a 'fragmented mode of publicity and outreach,' leading to duplication or gaps, which made this the weakest aspect in the deployment (ibid.). As noted before, the publicity of a deployment is fundamental to motivate and gather a significant number of reports.

[423] That is, partnerships, publicity and outreach, personnel, technology, workflows and citizen engagement (ibid., 4).

[424] This can be done in the case of an election, but not in the case of many climate-related disasters, although emergency preparedness should be in place in hazard-prone locations.

To sum up, Ushahidi is an example of a class of phenomena (proactive data activism) that provides an analytical frame which the case illuminates. Its platform generates alternative digital public spheres and involves communicative actions, which allow normal citizens to act, especially in case of crisis. Ushahidi, born out of an information shutdown in Kenya during the post electoral violence of 2008, has integrated the values of objectivity, fairness and accuracy of journalism, and uses the method of verification. Since then, apart from being a tool in humanitarian operations, it is also employed regularly as an information enhancer in journalistic reports. Above all in the case of armed conflicts, Ushahidi creates an alternative discourse than the one being disseminated by the people in power, and as such can be considered as a manifestation of *subaltern counterpublics*. Ushahidi deployments generates *ad hoc* TANs that deal with local crises, coordinating action based on analysis, visualisations and interactive mapping that render crowdsourced data useful for decision-making. Although Ushahidi does not have a reactive nature, it shows some of social movements' features in that it makes unconventional means (disruptive technologies) available to normal people, and its deployments work in network structures, and based on shared beliefs and solidarity.

Among proactive activist initiatives, Ushahidi deployments are to be considered geoactivist endeavours, as they have been specialised in charting catastrophes, disasters and other events and issues on interactive, multi-layered maps that seek to build communities and trigger action. Within geoactivists, Ushahidi deployments always rely on data generated via *crowds*. Ushahidi also shares some of the features of the match-maker and catalyst proactive activist initiatives, as it seeks to transfer skills through training and form alliances, as well as facilitate activism by providing the resources to promote data activism. As a geoactivist, its attributes include that it always produces of proactive data activist content or projects based on maps, mainly creates data

platforms and software for activism, facilitates data projects, works mostly in collaboration with others, and occasionally engages in advocacy exercises.

Ushahidi's repertoires of action include crisis mapping applying different tools and methods, including participatory maps and crowdsourced data, geospatial platforms, integrated mobile applications, aerial and satellite imagery, visualisations, and computational and statistical models. As said, Ushahidi was among the first to employ these tools to enable humanitarian rescue operations.

As said at the beginning of the dissertation, big data infrastructures can be compasses that help us navigate in complex situations, and face problems in a democratic way. This is precisely what Ushahidi does through a process of empowerment, even in very unequal situations when catastrophe strikes. This empowerment process acts in different ways: the individual gets empowered in the knowledge of not acting alone, of being part of a networked community; he or she gets enhanced and liberated as individual by technology as well, since it creates safe alternative public spheres that allow connection and action; *victims* get empowered by being transformed into reporting witnesses when they take action upon their situations.

Although exemplary, Ushahidi is not a flawless implementation of what could be a theoretical proactive data activist model, though. A number of Ushahidi deployments have experienced problems in their ability to generate communities and to keep on already established initiatives; to set up processes of verification of the data and protection of reporters; to engage other key actors and adapt to local realities; and to introduce comprehensive and practical evaluation and monitoring systems. However, its team members seem open about criticism, ready to try and fail, and prepared to absorb and integrate lessons learnt.

8. Analysis of the data

Ultimately, pro-active data activism can be considered a new, advanced form of citizens" media, one that has a critical approach to big data at its core. Similar to citizens" media, pro-active data activism involves a politics of the quotidian, as it alters the everyday relationship between citizens and automatized data collection. As such, it brings back into the data collection machine the fundamental elements of agency and politics. Similar to other social movements, pro-active data activists work towards long-term norm change: they struggle for "the social control of the main cultural patterns (…) through which our relationships with the environment are normatively organised" (Touraine, 1985). Thus, in the long run, data activism is likely to change the way citizens approach computational politics and the informational state, as well as the way we see and practice social change (Milan and Gutiérrez 2015, 130).

In this section, I examine the qualities, activities and values that distinguish organisations practicing proactive data activism by looking at whether and how a set of assertions describing behaviour and actions applies to each of them, and to what extent. These assertions are listed below and are assigned a code. I visualise the resulting grid (see Table 5) and identify visually the organisations and initiatives that show these qualities, and then propose a model for effective data activism. The conceptualisation of the grid in Table 5 comes from section 6, which looks at the shapes proactive data activism is taking and what sort of action repertoires these organisations and initiatives engage with. As said, here I establish attributes that describe the actions typical of data activist organisations, to which values are assigned. In order to determine the level of intensity of the attributes, I have set up a range, from 1 to 3, in which 3 is the maximum degree of intensity of each attribute. The non-assignment of a value means there is an absence of an attribute, that an attribute is not relevant or that, in a few cases, I ignore whether

that attribute is present in a given organisation. The results of the exercise are arranged in the grid shown in Table 5. The attributes are:

The organisation in question…

- includes skills transfer activities (which I code with a ST);
- produces data journalism content (DJ);
- produces data technologies, platforms or software (DT);
- provides funds or resources and are catalysts (FR);
- produces proactive data activist content or projects (DAP);
- produces data visualisations (DV);
- works in alliances or collaborates often (AC);
- offers open data so others can 'play' with them (OD);
- shows reactive data activism leanings, i.e. offers protection to members or participants, or is involved in disruptive strategies or protest (RA);
- provides match-making opportunities between organisations with synergies to produce proactive data activist projects (MM);
- or engages in campaigning exercises, including lobbying, mobilisation, *clickivism* and media outreach (CP).

Table 5: A classification of proactive data initiatives

ORGS.	**ST**	**DJ**	**DT**	**FR**	**DAP**	**DV**	**AC**	**OD**	**RA**	**MM**	**CP**
DataKind	3			1	1		3			3	
Kiln	1	2	2		2	3	1			1	2
Civio	1	3	1	2	3	2	3	2	2	2	3

Greenpeace		1		1	1	1	2		3	1	3
ICIJ	1	3	1	2	2	2	3	1	1	2	1
Morlan	1	3					1				
InfoAmazonia	2	3	1		1	3	2	3	1	1	2
Medialab-Prado	3	1		1	1	1	2			3	
Environ. News Lab	3	3				3	1	3			
ICFJ	3	3					1			2	
Oxpeckers	3	3				3	1				
Bayers Impact	3		3	1	1	1	2			2	
School of Data	3	1	2	1	1	1	1			1	
Tactical Tech	3		1	1	1		2			2	3
Code for Africa	3	1	1	3				1		2	
Data4Good	2				1					2	
Data Science for Social Good	3			1	1	1	2			2	
eScience Institute	3			1	1	1	2			2	
Hacks/Hackers	3	3	1		1	1		1		3	
ODI				1	2	2	2				3
Social Brite	3		1						1	1	
Gap Minder		2				3			1		1
Vidas Contadas		1			3	3	1	2	1		2
España en llamas		1			2	3	1	2			3
PAH		1			2	3	2	1	3		3
Death on the Road		1	1		2	3	2	2			3

Take back the Tech						3	1		3		3
Ushahidi			3	1	3	3	3	2			1
NY BigApps				3							
Google Flu Trends						3					
Comprehensive Knowledge Archive Network								3			
OpenSpending								3			
Knight Foundation	2			3							
WikiLeaks		3			3				3		2
ViveloHoy		3							1		1
Internews	1	3		2		1	2				
Community drones in Indonesia	1				3	3	1		3		3
Outliers Collective					2	3	1				
Vizzuality					2	3	1				
CARTO		1			2	3	1				

Source: Elaboration by the author.

The objective is of this exercise to test the coherence and robustness of the classification of proactive data activist organisations and initiatives in four groups, summarised in Table 1: that is, the ones practicing data journalism, the skills transferers, the catalysts and the proactive data activists (and geoactivists). For example, ICIJ is an organisation producing data journalism and visualisations. As such it always generates data journalistic products, sometimes provides funds or resources for data projects, occasionally produces proactive data activist content, often publishes visualisations and maps, frequently works in alliance with others, when possible uses

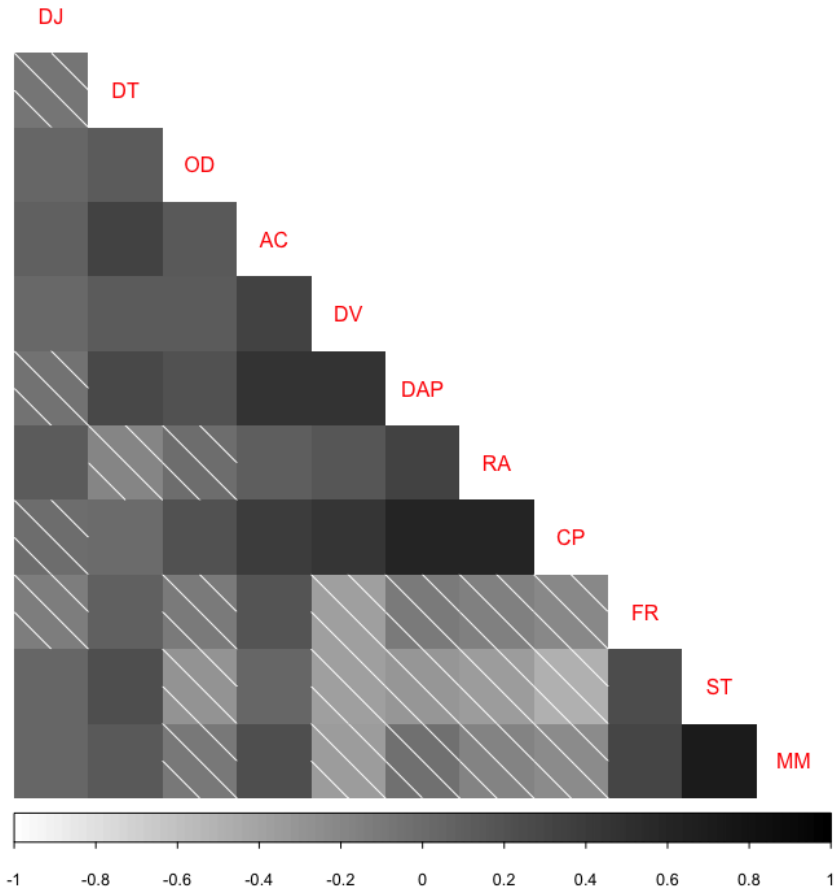
and offers open data, from time to time acts as a match-maker, and sporadically engages in advocacy activities, such as media outreach. This is the case of an organisation dedicated to do investigative journalism that has produced the big data-based 'Panama Papers' international investigation and its ensuing media outreach campaign. The other objective is to identify the most frequent attributes, and how they are associated, in order to base a model of effective proactive data activism.

The attributes that have more frequency include doing data visualisations (DV), transferring skills (ST), working in alliances (AC), producing data journalist outputs (DJ), doing proactive data activism (DAP) and resorting to campaigning and advocacy (CP). The less common attributes include fund-raising (FR), resorting to reactive activities (RA) and creating data platforms and tools (DT). Although there is an element of arbitrariness in the number and selection of organisations and initiatives, as well as the allocation of the level of intensity of each attribute, there is some logic in it. Data visualisations, including above all interactive maps, are a widespread analysis and communication tool, as explained in this dissertation. In this initial phase of proactive data activism, transferring skills seems to be a necessary activity as organisations learn to utilise data infrastructures to their maximum potential. Data activist organisations' missions include tackling immense challenges (i.e. disasters), so they tend to work in alliances and partnerships. As noted, in many places traditional media organisations are not producing the journalism that is needed to sustain democratic societies, and NGOs are stepping in to fill the gap, producing data journalism. Producing data activist outputs is naturally one of the most common endeavours among data activist organisations. And resorting to campaigning and advocacy work is also a common characteristic. This assessment could serve as a baseline

for further study, since these trends will evolve as the sphere of action of data activism matures and journeys towards making full employment of the data infrastructures.

Figure 3 below shows the correlations that can be established within attributes. The idea is to learn whether there are certain attributes that are more strongly associated than others. It shows, for example, that there is a strong correlation between skills transferring (ST) and match-making (MM) activities; that is, organisations that focus on transferring skills so others can develop data projects are often committed to making connections among organisations, with the expectation that data projects will emerge from such connections. In fact, the organisations that I have classified as skills transferers include those organisations dedicated to transfer data and social skills, as well as those dedicated to generate tools and generate match-making opportunities. An example could be Medialab-Prado, an organisation dedicated above all to both transferring data skills and generating the necessary synergies and contacts among other organisations so data projects are generated and facilitated in order to spread a culture of data. Code for Africa, Data4Good, Data Science for Social Good, eScience Institute and Hacks/Hackers could be included in this hybrid group showing both attributes. This is coherent with Tables 3 and 4, which show the clusters of organisations that gathered by their activities and a classification of data organisations by their attributes, where the organisations showing match-making attributes appear also within the groups of skills transferers. This is also coherent with the initial assumption that data activist organisations are natural hybrids.

Figure 3: Correlations among proactive data activism attributes (activities and features)



Source: Elaboration by the author.

There are strong correlations between the production of data activist projects (DAP), working in alliance (AC) and campaigning and advocacy activities (CP). That is to say, organisations that generate proactive data outputs normally work in alliance with others, and logically engage in campaigning or advocacy activities. Examples of these organisations with such attributes include, for example, Civio and its data projects (i.e. 'España en llamas'), Greenpeace, ODI, initiatives such as 'Vidas contadas,' and to a lesser extent because of its less *campaigny* nature, Ushahidi. Again, this is coherent with Tables 3 and 4.

Other logic correlations exist having reactive traits (RA) and a *campaigny* nature (CP). Clear examples include Greenpeace, Plataforma de Afectados por la Hipoteca (PAH), Take Back

the Tech, WikiLeaks and the drones in Indonesia project, all of them resorting to protest and advocacy to generate social change. And finally, producing data activist outputs (DAP) also shows a correlation with generating visualisations (DV), since most data activist organisations resort to visualising their data in one way or the other. Examples include Ushahidi, Kiln, InfoAmazonia, Oxpeckers, Gap Minder, and ‘Vidas contadas,’ among others.

9. Conclusions and areas for further research

Citizen-generated data and data analysis can offer alternative ways of thinking about the often-monolithic discourses around environmental data and big data to develop an understanding of the concrete ways in which data practices unfold in relation to environmental problems. They potentially help to generate a set of techniques for engaging with and querying evidence and evidence making; while providing an opening into what Gray et al. (2016) refer to as a challenge to the usual ways of collecting and interpreting data. Such challenges could be staged by looking at data gaps, data biases and the data status quo. In turn, by attempting to do data differently, collective environmental problems could be more creatively and expansively addressed and mitigated. This is the other aspect of 'just' good enough data –that citizen data could provide ways of realising environmental and social justice (Gabrys 2016, 12).

In this study, I have illustrated how the two premises that have been used in this dissertation as entry points to the field can be considered valid. The first is that, as information systems, big data infrastructures are on level pegging with the kind of diversity, complexity and disarray of the post-modern network society (Castells 2010b). In all ages, people have struggled come to terms with their lifeworld and have tried to map it (Jarvie 2015, 165). Big data and their infrastructures' own complexity are both a manifestation of a society composed of hyperconnected individuals and machines, as well as a suitable tool to explore and navigate this post-modern society. However, it is not a perfect mapping system, since today's unmanageably large and changing society allows only for imperfect mapping systems (ibid.), and big data sets are never raw (Gitelman 2013). They always emerge incomplete, biased and framed by social, economic and ideological constructs. But even considering their flaws, inadequacies and gaps, an information system based on big data infrastructures –along with ICTs— can help us understand and navigate this post-modern complexity.

Big data infrastructures can propagate inequality and be employed in discriminatory and hegemonic practices (Gangadharan 2012; Tufekci 2012; Braman 2006; Bauman and Lyon 2013; Lyon 2014; Brevini, Hintz and McCurdy 2013; Berry 2011; Deibert 2010). In spite of it, the second premise is that big data can also enable tackling social problems and complexity in a democratic manner, as data infrastructures and other technical artefacts, in the hands of people, can empower them and act as social equalizers. Data infrastructures can help individuals and groups critically approach data agency, independently produce their datasets on their own terms, shape practices and issues, generate alternative epistemologies in a datafied environment (Milan and van der Velden 2016), and ultimately change things. In fact, using data and communications infrastructures, with a focus on social good, individuals and organisations are responding to some of the most pressing challenges of today (i.e. action in humanitarian crises and disasters). They are enhancing democratic participation as well, by allowing people to create their own alternative digital public spheres and empowering them to take action (i.e. the data analysis that is both at the heart of the political messaging and the mobilisation of the Occupy movement).

Proactive data activism –a communicative action— assimilates the values of journalism, acquiring the objectivity ethos that rules good journalism and research, and then takes a subjective turn of the dramaturgical nature, and becomes non-objective when it uses data to connect with and mobilise people. In that can be compared to alternative media –media steered by a political aim, therefore non-objective—, as its values are a challenge that has both ‘a normative’ and ‘an epistemological’ aspect (Downing 2011, 18). As alternative journalism, data activism presents alternative versions of reality from those of the mass media. An example is the first Ushahidi deployment, during the slaughter that followed the elections in Kenya in 2008, which was being concealed by mainstream media.

This section summarises the main conclusions of the study.

Big data infrastructures, alone, are not going to solve humankind's problems. They are not a panacea (Innerarity 2016a, 2).; big data can hide gaps and biases (Gitelman 2013, 147-167; Kitchin 2014); and data infrastructures are also being used to create inequality and discriminate against certain groups (Gangadharan 2012), and to apply manipulating and censoring practices, such as the post-modern *panspectron* (Braman 2006; Tufekci 2014). Contrary to what Rossetto says, technology does exclude people. But the paradox is that it also can serve to redress exclusion and promote the Habermasian ideal of equal participation, and transforming *data poor* in *data rich* (Innerarity 2016a). Data activist projects can, in fact, rescue people from structural exclusion and place them at the decision-making table (i.e. some of Ushahidi projects). This could be look upon as a *retribalisation* process (McLuhan1964, 380), as data and communications infrastructures connect us and return us to a tribal village by making public interaction possible.

Governments and businesses were among the first social agents to utilise big data infrastructures in an useful manner: the former, in order to come up with more and better public services, but also to spy on, manipulate and control citizens in a more efficient and comprehensive manner than ever before (Berendt, Büchler, Rockwell, 2015); the latter, in order to generate more personalised marketing, products and services, and access more consumers in an individualised way, but also to influence them and generate more profits (Hawley 2012). Besides, big data infrastructures have also bred a negative form of public-private cooperation: some businesses cooperate in governmental massive surveillance as well, as the Snowden case illustrates (Castells 2015, 1; Sampedro 2014, 193). But in pioneering these technologies, they also have innovated and made big data's value chain increasingly more efficient, inexpensive

and accessible, so people in journalism, in the private sector and in the civil society have started to use and appropriate them for other social uses.

Using big data infrastructures, for example, to analyse and render useful data provided by whistle-blowers in the WikiLeaks cases, civic hackers have become examples of how to put in practice a prototype of critical and free press, restoring journalism to its roots in muck-racking investigative reporting (Sampedro 2014, 14). By employing and demanding data in a vocal way, many of data journalists also have stepped into the realm of activism (i.e. Clark and Cabra).

Likewise, combined with other elements like solidarity, collaboration and ingenuity, as well as with ICTs, big data infrastructures are being exploited to generate alternative digital public spheres in which individuals and groups can participate as equals, exchange vital information, and respond to humanitarian crises quicker and more efficiently than ever before (i.e. Ushahidi). As a result, data activists are apprising and renovating some of the old Enlightened ideals of freedom, equality and fraternity, and affording the possibility of equal participation as Habermas conceived it (1984, 147) and approaching a communicative ideal: a space uninhibited and free of restrictions, in which participants treat each other as equals in an effort to arrive to a mutual understanding (1991, 33-34). That is, proactive data activist projects create alternative public spheres, spaces where real communication and collective mobilization can be articulated. In doing so, data activism is facilitating ‘a re-energised, activist, engaged citizenry working together to create new small-scale communicative associative institutions that over time either merge into larger ones or at least join forces’ (Froomkin 2003, 753).

Among these ground-breaking agents are civic hackers and data activists who generate their own close-doors spheres and networks, sheltered by firewalls and encryption, to protect themselves against governmental and corporate intrusion (Milan and Hintz 2013, 22). We have

called them reactive data activists (Milan and Gutiérrez 2015, 127). Civic hackers and reactive data activists, working 'in the fringes of society,' led the way in data activism, quickly advancing from a marginal to a dispersed phenomenon (ibid.). This process happened as data infrastructures and practices got more accessible and prevalent, and they were transferred from skilled hackers to the ordinary citizen and civil society groups (ibid.). Proactive data activism –a new technology-enabled social practice rooted in data infrastructures— emerges from this transformation. It combines, politically and proactively, communicative practices, technology and information at its outermost complexity (that is, big data), was well as collective organising, to foster social change (ibid., 125-133).

Proactive data activism can be examined from different perspectives. It is a communicate action, with empirical-analytical, hermeneutic and critical elements (Habermas 1984). It is also similar to investigative journalism for its analytic nature, to advocacy journalism for its political stand, and to citizens' media for its collaborative, rank and file character. It places citizens at the forefront of production (Rodríguez 2004). It is local and transnational at the same time, as it is based in collaboration across borders (i.e. Ushahidi). Finally, it works towards long-term norm change and gathers people around the specific use they make of technology to form a collective identity, communicate, mobilise and act (Melucci 1996a; 1996b).

Based on empirical observation, proactive data activist organisations and initiatives can be classified in four different types: the activist organisations that produce data journalism when media outlets are unwilling or unable to do so; the skills transferers, focused on transferring data or social skills, building opportunities for alliances and creating digital platforms for data activism; the catalysts who fund and enable data projects; and the actual data activists, most of them geoactivists, because they utilise geolocalised data. Proactive data activists mostly engage

largely in creating alternative mapping systems and geoactivism –the most extended form of proactive data activism.

Proactive data activists can be also divided in several subgroups, depending on how they gather and generate data: they can a) rely on whistle-blowers for data; b) resort to open-source, publicly available datasets; c) use crowdsourcing tools to generate data from communities; d) turn to appropriating data; and e) get data from primary research that can be datafied and analysed, or generate their own data (i.e. via sensors).

Proactive data activism sets in motion a formidable empowering process. The domino effect of, for example, WikiLeaks or the 'Panama Papers,' mediated by journalists doing the analysis and visualisation, enthused new energy in declining journalism across the western world. In the case of organisations or projects depending on crowdsourced data, this process is unleashed when data and communications infrastructures allow *victims* to become reporting witnesses, who regain their agency by generating data in their own terms and taking action upon their situations and that of others. Because, in fact, control and power lie in datafication as well (Kennedy, Poell and van Dijck 2015), and when people engage in datafication, they become powerful. Data activism –data power— becomes a redisribution of sway by granting access to knowledge and action to people who had no such access before (Baack 2015, 1).

Table 6 summarises the characteristics, objectives, repertoires of action, and methods of data generation and empowering effects that are typical of proactive data activist initiatives extracted by empirical observations of relevant cases, semi-structured and in-depth interviews, and the case study.[425] This table could be considered a model of proactive data activism for future endeavours.

[425] As the general purpose of the case study is exploratory, exploratory conclusions are necessarily drawn with extreme caution (Yin 2002).

Table 6: Proposal for a model of proactive data activism

Features	Description
Characteristics	**Hybridisation**: Proactive data activists are unapologetic and natural hybridisers: business models, contents, repertoires of action, organisational structures, activities and objectives can be mixed and reworked as needed. Their proactivity makes buy-in and collaboration easy.
	Similar to social movements: Although never totally confrontational, proactive data organisations that generate TANs, such as Ushahidi, can look very much like social movements: they employ unconventional means, develop in ample network structures, are based on shared beliefs and solidarity, and pursue conflictual aims (della Porta and Diani 2006, 5, 7,159, 180).
	What they do: According to what they specialise in doing, proactive data activist initiatives can be classifyed as skill transferers (of data or social skills, or they can be match-makers and tools producers), producers of data journalism, catalysts and data activists (i.e. geoactivists).
Objectives	**Social change**: Although they may have immediate objectives (i.e. support rescue operations), as social movements, proactive data activists work towards eventual norm change and are identified by the specific use they make of technology to form a collective identity, communicate, convey, mobilise, demonstrate and act.
Repertoires of action	**Tech & tactical**: From a tactical point of view, geoactivist humanitarians, usually resort to: a) geolocalising information and witnesses' reports, providing the identification of the geographic location of an event (Birregah et al. 2012); b) categorising and translating text messages to facilitate communication and understanding between witnesses (reporters) and assistance providers, or among the former (Meier 2016; Adewumi 2008; Meier and Munro 2010); c) scraping social media sites for crisis-specific keywords and information (Terp 2016; Rafoth 2015); and d) locating key logistic infrastructures on aerial imagery (Parker 2015).
	Attributes: The attributes most common in proactive data activism include the following repertoires of action and elements: they generate visualisations; transfer skills to enable data projects; work mostly in alliances and collaborate; often produce data journalist outputs when media organisations are not doing it; generate data activist projects and outputs; and resort to campaigning and advocacy activities for social change. These attributes are present in concrete organisations with different levels of intensity.
How they generate data	**Strategy**: The manner in which data are generated for proactive activism is vital and can be an effective way to categorise proactive data initiatives into: a) organisations that become recipients of data via whistle-blowers; b) organisations that obtain novel insights from available, but unrelated and unexplored, datasets, producing secondary research; c) organisations that generate the means to crowdsource data, which are contributed by communities; d) organisations that appropriate data; e) organisations that conduct primary research whose findings can be datafied and analysed, or generate their own

	data. Having a concrete strategy in place is vital.
	Volunteers and authorities: Geoactivists tend to rely on collective participation to generate data contributed by users (reporters), and these data are then transformed into layers of information in maps which are employed for communication, coordination and action. For a geoactivist endeveaur based on crowdsourced data to succeed, two types of communities are needed: a) a critical mass of bloggers, techies and activists –that is, a community of digital humanitarians— who are willing to become expert volunteers working in a sufficiently free environment, and can contribute to the overall effort online; b) a robust community of non-expert reporters and users, who contribute their data and use them to act upon their situation on the ground; and the humanitarian organisations who lead the aid operations on the ground, based on the information generated by the latter and analysed by the former.
	How to crowdsource in a crisis: Each successful Ushahidi deployment requires: a) enthusiastic cooperation and buy-in both to form a *crowd* and to work with relevant services on the ground, b) careful adaptation of the platform to the local situation and its environment, as well as to the crisis at hand, c) verification of the information provided by users on the ground, d) the protection of the individuals reporting information, and e) collaboration and coordination with rescue services, both local and international. Given the usual need for immediate results in the middle of a crisis, this mechanism has to be prepared months in advance.
Empowering effects	**As a community**: What makes proactive data activist endeveaurs like Ushahidi special is that they rely on crodsourced data, and not on secondary data, for example (i.e. IUU fishing project). The empowering intensity of such projects is bigger, indepedently of the outcome. However, managing expectations is important. Crowds would not flock an Ushahidi deployment to provide reports if they do not expect that action will take place as a consequence of their reporting.
	As an individual: In the case of Usahidi, so called *victims* are empowered by being transformed in reporting witnesses, generating awareness, aiding humanitarian efforts, and changing humanitarian practices on the ground as well as the way people theorise about them.

Source: Elaboration by the author.

However wide-ranging, this study leaves many questions behind. This section compiles some of the most interesting ideas emerging from the research.

The increasing socialisation of big data infrastructures –that is, information at its most complex form– is one of the trends identified in this study. Big data are ‘going mainstream’ (Schroeder 2015)**.** Aided by both connective and data infrastructures, decision making is moving

‘from the elite few to the empowered many’ (IBM 2015). However, I have explored too how data infrastructures are not the new *punk*; that is, in spite of living in big data society, individual citizens have not embraced the culture of data, and even some organisations that are in a better position to make the most out of data infrastructures are yet to adopt them. Examining how this trend evolves in the context of activism and what rare the mechanisms set in motion as it does could be an interesting line of work. In fact, how action repertoires spread from one community to another is an aspect that has been neglected in this dissertation for lack of space and focus, but this is nevertheless a thought-provoking question in the context of data activism.

Couldry argues that media, as object of study, can be understood, not only as texts or structures of production, but as practice (2004). Couldry’s proposal is to move beyond previous debates about media effects, the technical properties of media, and the importance of political economy and audience interpretation, to focus instead on whether media might have a privileged role in organising other practices in media-saturated cultures. Data can also be seen as media (Couldry and Povell 2014, 3). So, if media can be understood as practice, why not data? Studying how data can be understood as practice could be another stimulating line of work.

Exploring the transborder nature of information and data flows is also a field that deserves more attention. What elements of proactive data activism are common across cultures, languages and frontiers? And what is intrinsically characteristic of the proactive data activism practiced in certain geographies and communities? On the other hand, paraphrasing Couldry, an examination of how data-related practices are embedded in countries with diverse cultural, social, economic norms could be another interesting endeavour (Couldry and Hobart 2010, 7).

The issue of data governance has not been tackled in this dissertation, although it deserves intense examination. Who is to govern data and data infrastructures generation and

access? Is it a citizen's right? What happens when people are denied their data rights? What threats can emerge out of poor data governance? Should a 'data justice' exist, like Dencik, Hintz and Cable suggest (2016, 1)?

According to Actor-Network Theory (ANT), actors in social networks include both humans and non-humans (i.e., objects), and there is no important distinction between them as both embody interests (Latour 1988, 109-117). ANT is an approach to social, scientific, technological and organisational structures and processes that consists in charting relationships that include concurrently human and non-human, which construct a network of links that can be represented and defined (Law 2011, 389-390). ANT strives to map associations that are at the same time material (between things) and semiotic (between concepts) (Law 2007, 8). Together these form a single network. The blurring of the notion of agency explains how different elements, including code and data, can play a role in determining the outcomes of social action. ANT describes processes of knowledge creation, by explaining the relational ties within the 'heterogeneous networks' (Law 2011, 385). How does data activism look like from an ANT perspective? This is another question deliberately left out of this dissertation, but nonetheless interesting.

Data activism is not only a question of technological entry barriers going down, but also a question of data being put to work for objectives that are framed by common sets of values, shared collective identities and social causes. Exploring more in-depth the new ways in which activists are creating collective identities when they generate data and data analysis on their own terms seems another interesting venue.

The results of advocacy and campaigning efforts are not always easy to ascertain and quantify. Campaigns are long endeavours, and sometimes their impacts take years to materialise.

Evaluating the effectiveness of data activism in terms of impact, vis-à-vis other types of activism, and setting success indicators seems to be another stimulating enterprise.

Social change can be explained from different perspectives, and its roots can be tracked back to different factors, dynamics and triggers: power structures and dynamics, evolution, technical innovation, new ideas and paradigm shifts, individual and social progress, social development and expert change networks, markets dynamics, cycles, conflict and struggle, and chaos. Is one of them more useful than the others in the case of data activism? Does change happen abruptly or gradually in data activism? What about the broad social changes brought about by big data infrastructures and those that result from data-enabled data activism? In history, we find examples of all types of changes, abrupt, gradual and cyclical, as well as mixed varieties best represented by a stepladder or a leapfrogging.[426] Marxists present a dialectic and materialistic concept of history, and change as a result of a fundamental struggle between social classes. But Kuhn explains changes as the demise of an unworkable paradigm, which is then replaced by a more effective and commonly accepted one (Kuhn 1970). Some technological changes with profound social impacts are represented as leapfrogging (i.e. the quick introduction of the mobile phone in rural China, when no land lines were available). Each of these basic theoretical change models has its defenders, its explanatory strengths and its limits. Most of them

[426] The concept of leapfrogging was initially employed in explaining economic growth in the context of competition among innovation or technology companies. It is based on Joseph Schumpeter's idea of 'creative destruction' (Schumpeter 2006, 81-87). Schumpeter suggests that firms controlling monopolies based on dominating technologies have less incentive to innovate than newcomers, so they sooner or later become complacent and lose their technological edge. When technological innovations are proposed by new firms which are willing to take the risks, they fail and the innovations become the new paradigm, producing the leap frog. Of late the concept of leapfrogging is being used as a theory of development. Leapfrogging may quicken development by avoiding earlier, less efficient, more expensive or more polluting technologies, and moving directly on to more advanced ones. This explains the quick mobile phone integration in isolated regions where no telephone grids were available, and he same goes for solar or wind energy filling a gap in energy grids that do not reach remote regions. This theory proposes that, leapfrogging developing countries can circumvent environmentally destructive stages of development and do not need to follow the polluting development trajectory of industrialised countries.

are not mutually exclusive, as some of the explanatory factors behind them (like conflict and power) are two sides of the same coin, while different models can fit different situations most accurately than others. Indeed, social change can be explained from different perspectives, and its roots can be tracked back to different factors, dynamics and triggers: power structures and dynamics, evolution, technical innovation, new ideas and paradigm shifts, individual and social progress, social development and expert change networks, markets dynamics, cycles, conflict and struggle, and chaos. Exploring which model applies to which data-based social change is another interesting area.

It was hard to detect examples for some of the categorisations identified in this dissertation (i.e. cases of data activist projects that resort to appropriating data), but common sense dictates that there could be more cases in the future. A more encompassing empirical study and classification of cases, including other languages and geographies, using the groupings proposed in this dissertation would populate them with more examples and detail richness.

Finally, some aspects present in Ushahidi, which resemble social movements, could be explored more in-depth, including how individual commitment is generated in Ushahidi deployments; how the platform spreads from one country to another, from one community to another; what is their level of democratic representation, participation (rather than delegation) and consensus-building (rather than majority vote) (della Porta and Diani 2006, 243); whether technology determines how Ushahidi generates change or not; and what type of social exclusion Ushahidi deployments tend to tackle.

References

Ackoff, Russell L. *Re-Creating the Corporation: A Design of Organizations for the 21st Century*. New York: Oxford University Press, USA, 1999.

Adewumi, David. "Kenyan tech bloggers launch crisis-report site." *Venture Beat*, January 15, 2008.

Adorno, Theodor W. *Negative Dialectics.* London and New York: Routledge, 2004.

Adorno, Theodor W., and Tiedemann, Rolf. *History and Freedom: Lectures 1964-1965.* Cambridge, UK: Polity, 2006 [2001].

Aguado, Txetxu, and Martin, Annabel. "Crisis, change and the humanities." In *The role of the humanities in times of crisis*, edited by Pilar Rodríguez, 147-156. Madrid: Dykinson, S.L., 2016.

Akella, Mamata. "Free Your Maps from Web Mercator!" *CartoDB*, October 30, 2015. http://blog.cartodb.com/free-your-maps-web-mercator/ (accessed April 25, 2016).

Alemán, Eduardo, and Yeaji Kim. "The democratizing effect of education." *Research and Politics* 2, no. 2 (October-December, 2015): 1-7.

Alger, Dean. *Megamedia: How Giant Corporations Dominate Mass Media, Distort Competition, and Endanger Democracy.* Lanham, MD: Rowman & Littlefield Publishers, 1998.

Allan, Stuart. *Citizen witnessing: revisioning journalism in times of crisis.* Cambridge: Polity Press, 2013.

Amnesty International. "Exposed: Child labour behind smart phone and electric car batteries." *Amnesty.org*, January 19, 2016b. https://www.amnesty.org/en/latest/news/2016/01/child-labour-behind-smart-phone-and-electric-car-batteries/ (accessed January 20, 2016).

—. "This is what we die for." Amnesty.org, 2016a. https://www.amnesty.org/es/documents/afr62/3183/2016/en/ (accessed December 20, 2016).

Andersen, Kurt. 2011. 'The Protester'. http://content.time.com/time/specials/packages/article/0,28804,2101745_2102132,00.html (accessed December 20, 2016).

Anderson, C.W. "Data, diffusion, impact: Five big questions the Wikileaks story raises about the future of journalism." NiemanLab, 2010. http://www.niemanlab.org/2010/07/data-diffusion-impact-five-big-questions-the-wikileaks-story-raises-about-the-future-of-journalism/ (accessed December 20, 2016).

Anderson, Chris. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired*. June 23, 2008. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed April 30, 2015).

Andrade, Pedro, Jess Hemerly, Gabriel Recalde and Patrick Ryan. "From Big Data to Big Social and Economic Opportunities: Which Policies Will Lead to Leveraging Data-Driven Innovation's Potential?" Public Policy Division of the World Economic Forum and Google, Inc., 2014. http://www3.weforum.org/docs/GITR/2014/GITR_Chapter1.8_2014.pdf (accessed April 30, 2016).

Andrejevic, Mark. "The Big Data Divide." *International Journal of Communication* 8 (2014): 1673–1689. http://ijoc.org/index.php/ijoc/article/download/2161/1163 (accessed April 30, 2016).

Andronie, Mihai. "Airline Applications of Business Intelligence Systems." *Incas Bulletin* 7, no. 3 (July, 2005): 153 – 160. http://bulletin.incas.ro/files/andronie_vol_7_iss_3.pdf (accessed April 30, 2016).

Annan, Kofi. "Grain, Fish, Money." Africa Progress Panel, 2015. http://www.africaprogresspanel.org/grain-fish-money/ (accessed April 30, 2016).

Anti-Evictionss Mapping Project.. "Documenting the Dispossession and Resistance of SF Bay Area Residents," 2016. http://www.antievictionmap.com/ (accessed April 20, 2016).

Ávila, Ana Maria. "Periodismo de Datos: Historia y Momento Actual." Manual de Periodismo de Datos, 2015. http://manual.periodismodedatos.org/ana-maria-Ávila.php (accessed February 22, 2016).

Ayuda Ecuador. "Mapa Desastre - Powered by Ushahidi." *Ushahidi Blog*, 2016. https://mapa.desastre.ec/ (accessed August 5, 2016).

Baack, Stefan. "Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism." *Big Data & Society*, (July-December 2015): 1-11. http://www.rug.nl/research/portal/files/26209006/_Baack2015Datafication.pdf (accessed August 5, 2016).

Badhwar, Neera K., and Long, Roderick T. "Ayn Rand." *The Stanford Encyclopedia of Philosophy*, Spring 2016 edition. https://plato.stanford.edu/entries/ayn-rand/ (accessed December 5, 2016).

Bahja, Fatima. "Citizen Journalism and the Mainstream Media: An Analysis of 85 New York Times Articles on the Syrian Civil War." *Gnovis* (December 11), 2013. http://www.gnovisjournal.org/2013/12/11/bahja-citizen-journalism/ (accessed December 5, 2016).

Ball, James. "WikiLeaks publishes full cache of unredacted cables." *The Guardian.* September 2, 2011. https://www.theguardian.com/media/2011/sep/02/wikileaks-publishes-cache-unredacted-cables (accessed December 5, 2016).

Ball, Kirstie, Haggerty, K. D., and Lyon, D., eds. *Routledge Handbook of Surveillance Studies.* London and New York: Routledge, 2012.

Balsiger, Philip, and Lambelet, Alexandre. "Participant Observation." In *Methodological Practices in Social Movement Research*, edited by Donatella della Porta, 144-172. Oxford: Oxford University Press, 2014. https://www.academia.edu/10312409/Participant_Observation (accessed December 5, 2016).

Bandy, Joe et al. *Coalitions across Borders: Transnational Protest and the Neoliberal Order (People, Passions, and Power: Social Movements, Interest Organizations, and the P).* Oxford: Rowman & Littlefield Publishers, 2004.

Barandiaran, Xabier. "Lectures on social innovation." Bilbao: University of Deusto (April 2015).

Barker, Robert L. *The Social Work Dictionary.* Washington: NASW Press, 1995.

Barnes, Russell. *Hackers in Wonderland.* London: Channel Four, 2000.

Bauman, Zygmunt. *Liquid Times: Living in an Age of Uncertainty.* Cambridge: Polity Press, 2007.

Bauman, Zygmunt, and Lyon, David. *Liquid Surveillance: A Conversation.* Cambridge: Polity Press, 2013.

Bayes Impact, Web site. "About us," 2015. http://www.bayesimpact.org/about (accessed May 11, 2015).

Becker, Howard S. "The Epistemology of Qualitative Research." In *Ethnography and human development: Context and meaning in social inquiry,* edited by. In R. Jessor, A. Colby and R. A. Shweder, 53-71. Chicago: University of Chicago Press, 1996. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.394.7240&rep=rep1&type=pdf (accessed May 11, 2015).

Becker, Sascha O., and Wößmann, Ludger. "Was Weber Wrong? A Human Capital Theory of Protestant Economic History." *University of Munich, Ifo Institute, CESifo and IZA* 124, no. 2 (2007): 531-596.

Bellingcat. "Investigathon – War and Pieces – Social Media Investigations." November 25, 2014. https://www.bellingcat.com/resources/case-studies/2014/11/25/investigathon-war-and-pieces-social-media-investigations/ (accessed May 24, 2016).

Bellinger, Gene, Castro, Durval, and Mills, Anthony. "Data, Information, Knowledge, and Wisdom." *Systems Thinking* (2004). http://www.systems-thinking.org/dikw/dikw.htm (accessed December 24, 2016).

Benkler, Yochai. *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (2006): 490-515. http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf (accessed December 24, 2006).

Bennett, Lance, and Segerberg, Alexandra. "The logic of connective action: Digital media and the personalization of contentious politics." *Information, Communication & Society* (Routledge) (2012): 739-768. http://ccce.com.washington.edu/about/assets/2012iCS-LCA-Bennett&Segerberg-LogicofConnectiveAction.pdf (accessed December 24, 2006).

Béranger, Jérôme. *Big Data and Ethics: The Medical Datasphere.* London: ISTE Press, 2016.

Berendt, Bettina, Büchler, Marco, and Rockwell, Geoffrey. "Is it Research or is it Spying? Thinking-Through Ethics in Big Data AI and Other Knowledge Sciences." *Künstliche Intelligenz* 29, no. 2 (March 14, 2015): 223.

Berger, John. *Hold Everything Dear.* New York: ihavebook.org/, 2008.

Bernard, Russell H. *Research Methods in Anthropology: Qualitative and Quantitative Approaches.* Lanham, New York, Toronto and Oxford: AltaMira Press - Rowman & Littlefield Publishers, Inc., 2006.

Bernays, Edward L. "The Engineering of Consent." *Annals of the American Academy of Political and Social Science* (1947): 113-120.

http://classes.dma.ucla.edu/Fall07/28/Engineering_of_consent.pdf (accessed December 24, 2006).

Bernholz, Lucy, Skloot Edward, and Varela, Barry. *Disrupting Philanthropy: Technology and the Future of the Social Sector*. Durham: Duke University, 2010.

Berry, David M. "The computational turn: Thinking about the digital humanities." *Culture machine* 12 (2011): 1-22. https://www.culturemachine.net/index.php/cm/article/download/440/470 (accessed December 24, 2006).

Bezerra, Julio et al. *The Mobile Revolution: How Mobile Technologies Drive a Trillion-Dollar Impact*. Boston: The Boston Consulting Group January 15, 2015. https://eliasgagas.files.wordpress.com/2015/03/the_mobile_revolution_jan_2015_tcm80-180510.pdf (accessed December 24, 2006).

Bijker, Wiebe E. *Of Bicycles, Bakelites, and Bulbs. Toward a Theory of Sociotechnical Change.* Cambridge: The MIT Press, 1995.

Black, Edwin. *IBM and the Holocaust : The Strategic Alliance Between Nazi Germany and America's Most Powerful Corporation.* Washington: Dialog Press, 2008.

Bliss, Jeanne. "88% of Lost Wallets With Baby Photos Get Returned: How To Use That Fact to Boost Business." *Customer Bliss* (April 6, 2010). http://www.customerbliss.com/88-of-lost-wallets-containing-baby-photos-get-returned-how-to-use-that-fact-to-boost-business/ (accessed January 28 2015).

Bloom, Paul. "Agaisnt Empathy." *Boston Review* (September 10, 2014). http://bostonreview.net/forum/paul-bloom-against-empathy (accessed January 28 2015).

Bogomolov, Andrey, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi and Alex Pentland. "Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data." Proceedings of the 16th International Conference on Multimodal Interaction, Istambul, Turkey, November 12-16, 2014: 427-434. https://arxiv.org/pdf/1409.2983.pdf (accessed January 28 2015).

Bollen, Johan, Mao, Huina and Zeng, Xiao-Jun. "Twitter mood predicts the stock market." *Journal of Computational Science* 2, no. 1 (March 14, 2011): 1-8. https://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiztq6i8ZvRAhWPzRoKHUsTCwcQFgghMAA&url=https%3A%2F%2Farxiv.org%2Fpdf%2F1010.3003&usg=AFQjCNHzFNLl3rPRyG0PqO2MJWNON20tdA (accessed January 28 2015).

Bollier, David (Rap.). "The promise and peril of big data." *Communications and Society Program – The Aspen Institute* (2010). https://www.emc.com/collateral/analyst-reports/10334-ar-promise-peril-of-big-data.pdf (accessed December 28 2015).

Borges, Jorge Luis. "Del rigor en la ciencia." In *Historia universal de la infamia*, 2007 [1935]. http://www.palabravirtual.com/index.php?ir=ver_voz1.php&wid=726&p=Jorge+Luis+Borges&t=Del+rigor+en+la+ciencia&o=Jorge+Luis+Borges (accessed December 28 2015).

Borja, Jordi, and Castells, Manuel. *Local and Global: The Management of Cities in the Information Age.* London and New York: Earthscan, 1997.

Bott, Maja, Gigler, Björn-Sören, and Young, Gregor. "The Role of Crowdsourcing for Better Governance in International Development*" International Bank for Reconstruction and Development and The World Bank* (2012): 49. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.474.2448&rep=rep1&type=pdf (accessed December 28 2016).

Boulanger, David. "Big Data Definitions." *Learning Analytics* (January 13, 2015). http://learninganalytics.ca/research/?page_id=1018 (accessed January 28 2015).

Bourdieu, Pierre. *Campo de Poder, Campo Intelectual.* Translated by Jorge Dotti. Buenos Aires: Montressor - Colección: Jungla Simbólica, 2004 [1971]. http://ceiphistorica.com/wp-content/uploads/2016/01/bourdieu-campo-de-poder-campo-intelectual.pdf (accessed January 1, 2016).

Boykoff, Maxwell T., and Roberts, Timmons J. "Media Coverage of Climate Change: Current Trends, Strengths, Weaknesses." *UNDP: Human Development Report 2007/2008* (March 2007): 1-54. http://site.iugaza.edu.ps/tissa/files/2010/02/Media_Coverage_of_Climate_Change.pdf (accessed January 1, 2016).

Bradshaw, Paul. "What is data journalism?" In *Data Journalism Handbook*, edited by *Jonathan Gray*, Lucy Chambers, and *Liliana* Bounegru. Newton, Massachusetts: O'Reilly Media, 2011. http://datajournalismhandbook.org/1.0/en/introduction.html (accessed January 1, 2016).

Braman, Sandra. *Change of state: Information, policy, and power.* Cambridge: The MIT Press, 2006.

—. "The politics of big data." *Contentious Data Conference.* Amsterdam: Datactive (September15-16, 2016).

Brevini, Benedetta, Hintz, Arne, and McCurdy, Patrick. *Beyond WikiLeaks: implications for the future of communications, journalism and society.* Basingstoke: Palgrave Macmillan, 2013.

Briggs, Asa, and Burke, Peter. *De Gutenberg a Internet: una historia social de los medios de comunicación.* Madrid: Taurus, 2002.

Brown, Eric D. "What's the difference between Business Intelligence and Big Data?" *Data Ops*, October 5, 2015. http://dataops.co/whats-the-difference-between-business-intelligence-and-big-data/ (accessed August 16, 2016).

Brown, Phil, Stephen Zavestoski, Sabrina McCormick, Brian Mayer, Rachel Morello-Frosch, and Rebecca Gasior Altman. "Embodied health movements: new approaches to social movements in health." *Sociology of Health & Illness* 1 (2004): 50-80.

Bullock, Kevin. "Satellite maps have never been better." *Digital Globe*, June 4, 2015. http://blog.digitalglobe.com/2015/06/04/the-state-of-the-map-its-never-been-better/ (accessed August 16, 2016).

Burbach, Roger. *Globalization and Postmodern Politics: From Zapatistas to High-Tech Robber Barons.* London and Sterling: Pluto Press, 2001.

Burns, Ryan. "Moments of closure in the knowledge politics of digital humanitarianism." *Elsevier - Geoforum* 53 (2014a): 51–62. https://geog.ucalgary.ca/manageprofile/sites/geog.ucalgary.ca.manageprofile/files/unitis/publications/1-7471988/burns_-_moments_of_closure_in_the_knowledge_politics_of_digital_humanitarianism.pdf (accessed August 16, 2016).

—. "Rethinking Big Data in Digital Humanitarianism: Practices, Epistemologies, Social Relations." Editado por Daniel Sui. *GeoJournal – Springer* 80, no. 4 (2014b): 1-14. https://www.researchgate.net/publication/273764067_Rethinking_big_data_in_digital_humanitarianism_practices_epistemologies_and_social_relations (accessed August 16, 2016).

C40Cities, Web site. "About C40," 2016. http://www.c40.org/about (accessed May 26, 2016).

Cabra, Mar. "Map: The Offshore Leaks Revelations." *The International Consortium of Investigative Journalists*, January 23, 2014. http://www.icij.org/offshore/map-offshore-leaks-revelations (accessed February 19, 2016).

Cabrero, Gregorio Rodríguez. "Spain: In-work poverty and labour market segmentation." *Peer Review in Social Protection and Social Inclusion and Assessment in Social Inclusion*,

2010. http://ec.europa.eu/social/BlobServlet?docId=9016&langId=en (accessed December 19, 2016).

Calhoun, Craig. *Habermas and the Public Sphere*. Cambridge: MIT Press, 1992.

Cambridge Dictionary. "Policy." *Cambridge Advanced Learner's Dictionary & Thesaurus*, 2015. http://dictionary.cambridge.org/dictionary/english/policy (accessed December 12, 2015).

Campbell, Niall. "Tax Policy and Administration in an Era of Big Data." *Tax planning: International indirect taxes - The Bureau of National Affairs Inc*. 12, no. 12 (December, 2014): 1-6.

Careless, Sue. "Advocacy journalism." *Religion, Society & Culture- The Interim* (May 15, 2000). http://www.theinterim.com/issues/society-culture/advocacy-journalism/print/ (accessed March 6, 2016).

Carreras, Ignasi. "Lectures on leadership and social innovation." Madrid: Instituto de Innovación Social, IIS – ESADE (2012).

Carrión, Jorge, interviewed by Ser. "Vagabundos de la chatarra." *Ser* (April 3, 2016).

Carrión, Jorge and Sagar. *Los vagabundos de la chatarra.* Barcelona: Norma Editorial, 2016.

Castells, Manuel. "Social Movements and the Informational City." *Hitotsubashi journal of social studies* 21 (August, 1989): 197-206.

—. *The Informational City: Information Technology, Economic Restructuring, and the Urban Regional Process.* Oxford: Wiley-Blackwell, 1992.

—. *The Internet galaxy: reflections on the Internet, business, and society.* Oxford: Oxford University Press, 2001.

—. *Informationalism, networks and the network: Society, a theoretical blueprint.* Northampton: University of Southern California, 2004.

—. *Communication Power.* Oxford: Oxford University Press, 2009.

—. *The Information Age: Economy, Society, and Culture, Volume III. End of Millennium.* Malden: Blackwell Publishers, 2010a.

—. *The Information Age: Economy, Society, and Culture, Voulme I.The Rise of the Network Society.* Oxford: Wiley-Blackwell, 2010b.

—. "Vivir en estado de vigilancia permanente." *Revista Ñ* - Clarin, March 9, 2015. http://www.clarin.com/rn/ideas/Vivir-vigilancia-permanente_0_ByImHmcv7g.html (accessed March 6, 2016).

Castells, Manuel, interviewed by Harry Kreisler. "Conversations with History." *Institute of International Studies* (May 9, 2001).

Castells, Manuel, and Cardoso, Gustavo, eds. *The Network Society: From Knowledge to Policy.* Washington, DC: Johns Hopkins Center for Transatlantic Relations , 2005. https://www.umass.edu/digitalcenter/research/pdfs/JF_NetworkSociety.pdf (accessed March 6, 2016).

Center for Public Integrity. *The Panama Papers.* April 3, 2016. https://www.publicintegrity.org/2016/04/04/19518/center-unit-icij-releases-panama-papers (accessed April 12, 2016).

Chandler, Ian. "Advocacy and campaigning." *The Pressure Group - Advocacy & Campaign Planning Mini-Guide* 1 (January, 2012): 1-4. http://thepressuregroup.org/wp/wp-content/uploads/2014/11/Mini-Guide-1-Advocacy-Campaigning-overview.pdf (accessed April 12, 2016).

Choi, Hyunyoung, and Hal, Varian. "Predicting Initial Claims for Unemployment Benefits." *Research at Google* (July 5, 2009): 1-5. https://static.googleusercontent.com/media/research.google.com/es//archive/papers/initialclaimsUS.pdf (accessed April 12, 2016).

Chomsky, Noam. "What Makes Mainstream Media Mainstream." *ZMagazine* (October 1997): 7. https://chomsky.info/199710__/ (accessed April 12, 2016).

Chopard, Jean-Luc, and Lusser, Vincent. "For whom do humanitarian organizations speak? A few thoughts about dissemination." *International Review of the Red Cross* 319 (August 31, 1997). https://www.icrc.org/eng/resources/documents/article/other/57jnrf.htm (accessed April 12, 2016).

Christiansen, Jonathan. "Four Stages of Social Movements." *EBSCO Publishing Inc.* (2009): 1-7. https://www.ebscohost.com/uploads/imported/thisTopic-dbTopic-1248.pdf (accessed April 12, 2016).

Civio, Web site. "Open Data For Everyone. Transparency For A Stronger Democracy," 2016. http://www.civio.es/en/ (accessed March 30, 2016).

Cleveland, Harlan. "Information as a resource." *The Futurist* 3-4, no. 6 (December, 1982): 34-39. https://www.researchgate.net/publication/234578952_Information_As_a_Resource (accessed December 30, 2016).

Centre on Law and Globalisation. *How Do Transnational Advocacy Networks Work?* 2015. https://clg.portalxm.com/library/keytext.cfm?keytext_id=114 (accessed December 30, 2016).

Clicktivist.org, Web site. "What is clicktivism," 2016. http://www.clicktivist.org/what-is-clicktivism/ (accessed April 18, 2016).

Code for Africa, Web site. "Open Data & Open Government Are Revolutionary," 2016. https://codeforafrica.org/ (accessed April 4, 2016).

Colaborabora, Web site. "Sobre el procomun," 2016. https://www.colaborabora.org/colaborabora/sobre-el-procomun/ (accessed April 4, 2016).

Committee to Protect Journalists. *Congo-Kinshasa: DRC Halts Internet Access and Cellphone Services Amid Protests*, January 22, 2015. https://cpj.org/2015/01/drc-halts-internet-access-and-cellphone-services-a.php (accessed April 4, 2016).

Congosto, Mariluz. *Entrevista a... Mari Luz Congosto (Universidad Carlos III)* datos.gob.es/acerca-de. Madrid. December 27, 2012. http://datos.gob.es/es/comunidad-risp/entrevista-mari-luz-congosto-universidad-carlos-iii (accessed April 4, 2016).

—. Interviewed by Miren Gutiérrez. "Miren Gutiérrez entrevista a Mª Luz Congosto." *University of Deusto Blog* (January 19, 2016). https://www.youtube.com/watch?v=MuX6nqmvW_U (accessed April 4, 2016).

Conneally, Paul. "How mobile phones power disaster relief." *TED*, February 2012. https://www.ted.com/talks/paul_conneally_digital_humanitarianism (January 19, 2016).

Conover, Michael D., Clayton Davis, Emilio Ferrara, Karissa McKelvey, Filippo Menczer, and Alessandro Flammini. "The Geospatial Characteristics of a Social Movement." *PLoS ONE - Center for Complex Networks and Systems Research* 8, no. 3 (2013): 1-8. http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0055957&type=printable (January 19, 2016).

Copwatch. *Copwatch Database*, 2016. http://www.copwatch.org/ (accessed May 29, 2016).

Couldry, Nick. "A Necessary Disenchantment: Myth, Agency And Injustice In A Digital World." *London School of Economics.* London: LSE, November 21, 2013. 16. http://www.lse.ac.uk/media@lse/documents/MPP/Nick-Couldrys-LSE-INAUGURAL-SCRIPT.pdf (January 19, 2016).

Couldry, Nick, and Hobart, Mark. "Media as Practice: a Brief Exchange." *Criticalia*, 2010: 1-15. http://www.criticalia.org/Sources/Media%20as%20practice%20-%20the%20Couldry-Hobart%20exchange.pdf (accessed January 26, 2015).

Couldry, Nick, and Powell, Alison. "Big Data from the bottom up." *Big data & Society* 1, no. 2 (2014): 1-5. http://bds.sagepub.com/content/1/2/2053951714539277.full.pdf+html (accessed December 2, 2016).

Cover, Matt. "True Cost of Fannie, Freddie Bailouts: $317 Billion, CBO Says." *CNSNews.com*, June 6, 2012. http://www.cnsnews.com/news/article/true-cost-fannie-freddie-bailouts-317-billion-cbo-says (accessed January 26, 2016).

Cowen, Tyler. "Public Goods." *The Concise Encyclopedia of Economics*, 2008. http://www.econlib.org/library/Enc/PublicGoods.html (accessed January 26, 2016).

Cox, Geoff, and McLean, Alex. *Speaking Code: Coding as Aesthetic and Political Expression*. Cambridge and London: The MIT Press - Software Studies, 2013.

Crotty, James. "Structural Causes of the Global Financial Crisis: A Critical Assessment of the 'New Financial Architecture. '" *Economics Department Working Paper Series* 16 (2008). http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1017&context=econ_workingpaper (accessed January 26, 2016).

Cruickshank, Justin, and Chis, Cerasella Ioana. "Big Data, TTIP and the Hubris of Techno-Capitalism." *Social Epistemiology Review and Reply Collective* 4, no. 3 (February, 2015): 45-55. https://socialepistemologydotcom.files.wordpress.com/2015/02/cruickshank_big_data1.pdf (accessed January 26, 2016).

Cukier, Kenneth. "Big data is better data." *TED Salon*, June, 2014. https://www.ted.com/talks/kenneth_cukier_big_data_is_better_data?language=es (accessed May 12, 2015).

CU-Citizen Access, Web site. "About us," 2015. http://cu-citizenaccess.org/ (accessed May 12, 2015).

Cukier, Kenneth, and Mayer-Schoenberger, Viktor. "The Rise of Big Data: How it's Changing the Way We Think about the World." *Foreign Affairs* 92, no. 3 (May/June, 2013): 28-40.

Cummins, Linda K., Byers, Katharine V., and Pedrick, Laura V. *Policy Practice for Social Workers: New Strategies for a New Era.* New York: Pearson Higher Education, 2011.

Daniels, Alfonso, Miren Gutiérrez, Gonzalo Fanjul, Arantxa Guereña, Ishbel Matheson and Kevin Watkins. *Western Africa's missing fish.* London: Overseas Development Institute, 2016. https://www.odi.org/sites/odi.org.uk/files/resource-documents/10665.pdf (accessed December 6, 2016).

Data & Society, Web site. "About*,"* 2016. http://www.datasociety.net/ (accessed May 5, 2016).

Data Science for Social Good, Web site. "About," 2016. http://dssg.uchicago.edu/ (accessed April 6, 2016).

Data4Good, Web site. "About," 2016. http://www.data4good.co/about.html (accessed April 6, 2016).

DataKind, Web site. "About us," 2015. http://www.datakind.org/aboutus/ (accessed May 11, 2015).

Data-Pop Alliance, Web site. "About," 2016. http://datapopalliance.org/about/vision-and-members/ (accessed May 26, 2016).

Davenport, Thomas H., and Prusak, Laurence. *Working Knowledge: How Organizations Manage What They Know*. Brighton, MA: Harvard Business Press, 2000.

Davies, Harry. "Ted Cruz erased Trump's Iowa lead by spending millions on voter targeting." *The Guardian*, February 1, 2016. https://www.theguardian.com/us-news/2016/feb/01/ted-cruz-trump-iowa-caucus-voter-targeting (accessed July 31, 2016).

Davis, Claire, and Wilcock, Elizabeth. *Teaching Materials Using Case Studies.* London: The UK Centre for Materials Education, 2003.

de la Cruz, Cristina. "Democratic education and university: Resistances and challenges." In *The role of the humanities in times of crisis*, edited by Pilar Rodríguez, 147. Madrid: Dykinson S.L., 2016.

de Simone, Sergio. "SMS Uprising: Mobile Activism in Africa." *InfoQ*, July 7, 2015. https://www.infoq.com/articles/sms-uprising-africa (accessed February 15, 2016).

de Soto, Pablo. "Los mapas del #15M: el arte de la cartografía multitud conectada ." *Universitat Oberta de Catalunya* (May 2014): 357-381. http://tecnopolitica.net/sites/default/files/pablodesoto.pdf (accessed May 26, 2016).

de Tocqueville, Alexis. *Democracy in America.* State College: The Pennsylvania State University, 2002 [1835].

de Vicente, José Luis, Gemma Galdón, Philippe M. Frowd, and José María Zavala. *Anonymise Yourself: Electronic Self-Defence Handbook.* Barcelona: Eticas Research & Consulting and ZZINC, 2014.

Dean, Jodi. "Why the Net is not a Public Sphere." *Constellations*, 10, no. 1 (March, 2003): 95-112. http://onlinelibrary.wiley.com/doi/10.1111/1467-8675.00315/abstract (accessed May 26, 2016).

Deibert, Ronald. "After WikiLeaks, a New Era." *The New York Times*, December 11, 2010. http://www.nytimes.com/roomfordebate/2010/12/09/what-has-wikileaks-started/after-wikileaks-a-new-era (accessed May 26, 2016).

Deibert, Ronald, John G. Palfrey, Rafal Rohozinski, Jonathan Zittrain, and Miklos Haraszti. *Access controlled: the shaping of power, rights, and rule in cyberspace.* Cambridge: The MIT Press, 2010.

Deibert, Ronald, and Rohozinski, Rafal. "Liberation vs. control: the future of cyberspace." *Journal of Democracy* 21, no.4 (2010): 43-57. http://www.journalofdemocracy.org/articles-files/gratis/Rohozinski-21-4.pdf (accessed May 26, 2016).

Deleuze, Gilles, and Guattari, Felix. *A Thousand Plateaus: Capitalism and Schizophrenia.* Minneapolis: University of Minnesota Press, 1987.

della Porta, Donatella. *Social Movements, Political Violence, and the State: A Comparative Analysis of Italy and Germany*. Cambridge: Cambridge University Press, 2006.

—. *Can democracy be saved?* Cambridge and Malden: Polity Press, 2013.

della Porta, Donatella, and Diani, Mario, eds. *Social movements: an introduction.* Malden, Oxford, Victoria: Blackwell Pub, 2006.

della Porta, Donatella, and Tarrow, Sidney, eds. *Transnational Protest and Global Activism (People, Passions, and Power).* Oxford: Rowman & Littlefield Publishers, Inc., 2004.

Dencik, Lina, Hintz, Arne, and Cable, Jonathan. "Towards data justice? The ambiguity of anti-surveillance resistance in political activism." *Big Data & Society* 1, no. 12, (July-December, 2016): 1-12. http://orca.cf.ac.uk/96123/7/Towards%20Data%20Justice%20-%20The%20ambiguity%20of%20anti-surveillance%20resistance%20in%20political%20activism.docx.pdf (accessed May 26, 2016).

Diamond, Larry. "Liberation Technology." *Journal of Democracy - The Johns Hopkins University Press* 21, no. 3 (2010): 69-83. http://isites.harvard.edu/fs/docs/icb.topic980025.files/Wk%2011_Nov%2011th/Diamond_2010_Liberation%20Technologies.pdf (accessed May 26, 2016).

Diario Jurídico. "Los desahucios en España se disparan hasta los 517 diarios." *ABC*, August 29, 2012. http://www.abc.es/20120723/economia/abci-desahucios-espaa-201207231723.html (accessed May 26, 2016).

Dickinson, Elizabeth. "The first Wikileaks revolution?" *Foreign Policy*, January 13, 2011. http://foreignpolicy.com/2011/01/13/the-first-wikileaks-revolution/ (accessed May 26, 2016).

Digital Humanitarian Network, Web site. "About*,"* 2016. http://digitalhumanitarians.com/about (accessed May 26, 2016).

Dijkzeul, Dennis, and Moke, Markus. "Public communication strategies of international humanitarian organisations." *International Review of the Red Cross* 860, December 2005. https://www.icrc.org/en/download/file/20196/irrc_860_6.pdf (accessed May 26, 2016).

Downing, John D. H. *Encyclopedia of Social Movement Media.* Thousand Oaks: Sage Publications, Inc, 2011.

—. *Radical media: rebellious communication and social movements.* Thousand Oaks, London, New Delhi: Publications, Inc, 2001.

—. Interviewed by Miren Gutiérrez. "Special Report-Media: Alternative and Influential?" *Inter Press Service* (October 19, 2004). http://www.ipsnews.net/2004/10/special-report-media-alternative-and-influential/ (accessed May 24, 2016).

Doyle, Alister. "G20 pledges lift Green Climate Fund towards $10 billion U.N. goal." *Reuters*, November 16, 2014. http://www.reuters.com/article/us-g20-summit-climatefund-japan-idUSKCN0J00UL20141116 (accessed May 24, 2016).

Drone Code, Web site. "Why drones?" Linux Foundation Collaborative Project, 2016. https://www.dronecode.org/about/why-drones (accessed May 24, 2016).

Duffy, Matt J. "Smartphone in the Arab Spring: A revolution in gathering, reporting the news." *International Press Institute* (2011): 53-56. https://www.academia.edu/1911044/Smartphones_in_the_Arab_Spring (accessed May 11, 2015).

Dull, Tamara. "Big data and the Internet of Things: Two sides of the same coin?" *DZone – IoT Zone*, May 16, 2015. http://www.sas.com/en_ca/insights/articles/big-data/big-data-and-iot-two-sides-of-the-same-coin.html (accessed January 16, 2016).

Durkheim, Émile. *The Rules of Sociological Method.* New York: The Free Press, 1982 [1895].

Earl, Jennifer, and Kimport¸ Katrina. *Digitally Enabled Social Change: Activism in the Internet Age.* Cambridge: The MIT Press, 2011.

Einarsson, Torbjörn. "Membership and Organizational Governance." *Stockholm School of Economics*, 2012.

Ekine, Sokari. “Ushahidi Platform Used to Document Congo, Gaza Crises.” Media Swift, January 9, 2009. http://mediashift.org/2009/01/ushahidi-platform-used-to-document-congo-gaza-crises009/ (accessed August 15, 2016).

Elazari, Keren. “Hackers: the Internet's immune system.” *TED*, March, 2014. https://www.ted.com/talks/keren_elazari_hackers_the_internet_s_immune_system?language=es (accessed August 15, 2016).

Eliot, T.S. Chohruses from the Rock. In *Inside Work*, 2006 [1934]. http://courseweb.ischool.illinois.edu/~katewill/spring2011-502/502%20and%20other%20readings/eliot%20choruses_from_the_rock.pdf (accessed August 15, 2016).

Emerson, John. *Visualizing Information for Advocacy: An Introduction to Information Design*. India: Tactical Technology Collective, 2008.

Emiróglu, Sema. “The role of social media in crisis: Arab Spring as a test case.” *The 11th Metu Conference On International Relations.* Ankara, 2012. 10.

Environmental News Lab, Web site. “About us,” 2016. http://lab.oeco.org.br/about/ (accessed April 3, 2016).

Ericsson. *Mobility Report - On The Pulse Of The Networked Society.* Ericsson, Stockholm, 2012. https://www.ericsson.com/mobility-report (accessed April 3, 2016).

eScience Institute, Web site. “About us.” The University of Chicago, 2016. http://escience.washington.edu/get-involved/incubator-programs/data-science-for-social-good/ (accessed April 5, 2016).

Etling, Bruce, Faris, Robert, and Palfrey, John. “Political Change in the Digital Age: The Fragility and Promise of Online Organising.” *The Johns Hopkins University Press - SAIS Review* 30, no. 2 (December, 2010): 37-49 https://dash.harvard.edu/bitstream/handle/1/4609956/SAIS%20online%20organizing%20paper%20final.pdf?sequence=1 (accessed April 5, 2016).

Etzioni, Amitai. “Communitarianism.” *Encyclopaedia Britannica*, 2016. https://global.britannica.com/topic/communitarianism (accessed April 3, 2016).

European Union. *Big Data Public Private Forum (BIG).* European Union - BIG Consortium, 2014. http://www.big-project.eu/ (accessed April 3, 2016).

Exponential Training. *Greiner's Growth Model.* Exponential Training and Assessment, 2016. http://www.exponentialtraining.com/Downloads/Resources/Example%20Module%20-%20CBS/page_08.htm (accessed March 24, 2016).

Food and Agriculture Organisation, *The State of World Fisheries and Aquaculture 2010*. FAO Fisheries and Aquaculture Department, Rome, 2010. http://www.fao.org/3/a-i1820e.pdf (accessed April 6, 2016).

—. Web site. "Statistics at FAO," 2016. http://www.fao.org/statistics/en/ (accessed April 6, 2016).

Farrington, Julia. *Conference Report: Taking the offensive – Defending artistic freedom of expression in the UK.* Index on Censorship, London, 2013. https://www.indexoncensorship.org/takingtheoffensive/ (accessed April 6, 2016).

Feenberg, Andrew. *Transforming Technology: A Critical Theory Revisited* 2. Oxford: Oxford University Press, USA, 2002.

Fidelman, Mark. "These Smart, Social Apps Bring Big Data Down to Size." *Forbes*, October 2012. http://www.forbes.com/sites/markfidelman/2012/10/30/these-smart-social-apps-bring-big-data-down-to-size/#5765139f4985 (accessed April 6, 2016).

Fischer, Claude S. *America Calling: A Social History of the Telephone to 1940.* Oakland, CA: University of California Press. 1994.

—. *Lurching toward happiness in America.* Boston: The MIT Press, 2014.

Flyvbjerg, Bent. *Ideal Theory, Real Rationality: Habermas Versus Foucault and Nietzsche*. Paper for the Political Studies Association's 50th Annual Conference, The Challenges for Democracy in the 21st Century, April 10-13, 2000. London: London School of Economics and Political Science. http://flyvbjerg.plan.aau.dk/IdealTheory.pdf (accessed August 12, 2016).

—. "Five Misunderstandings About Case-Study Research." Inquiry 12, no. 2 (April, 2006): 219-245. https://arxiv.org/ftp/arxiv/papers/1304/1304.1186.pdf (accessed August 12, 2016).

Ford, Heather. "Can Ushahidi Rely on Crowdsourced Verifications?" *HBlog.org*, March 9, 2012. https://hblog.org/2012/03/09/can-ushahidi-rely-on-crowdsourced-verifications/ (accessed August 12, 2016).

Foucault, Michel. "What is Enlightenment?" In *The Foucault Reader*, edited by Paul Rabinow, 32-50. New York: Pantheon Books, 1984.

—. *Discipline and Punish. The Birth of the Prison.* New York: Vintage Books, 1995.

Fraser, Nancy. "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy." *Duke University Press* 25/26 (1990): 55-80. https://www.jstor.org/stable/466240 (accessed August 12, 2016).

—. *Transnationalising the Public Sphere.* Malden: Polity, 2014.

Freeden, Michael. *The New Liberalism: An Ideology of Social Reform.* Oxford: Clarendon Press, 1986.

Friedman, Thomas L. "Blowing a Whistle." *The New York Times*, June 11, 2013. http://www.nytimes.com/2013/06/12/opinion/friedman-blowing-a-whistle.html (accessed August 12, 2016).

Froomkin, Michael. "Habermas@Discourse.net: Toward a Critical Theory of Cyberspace." *Harvard Law Review* 116, no. 3 (January, 2003). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=363840 (accessed August 12, 2016).

Fuchs, Christian. "Alternative Media as Critical Media." *European Journal of Social Theory* 13, no. 2 (2010): 173–92. http://fuchs.uti.at/wp-content/uploads/altmedia.pdf (accessed August 12, 2016).

—. *Social Media: A Critical Introduction.* London: Sage, 2013.

Gabrys, Jennifer, Pritchard, Helen, and Barratt, Benjamin. "Just good enough data: Figuring data citizenships through air pollution sensing and data stories." *Big data & Society* 3, no. 2 (November 1, 2016). http://journals.sagepub.com/doi/abs/10.1177/2053951716679677 (accessed December 1, 2016).

Gallagher, Ian, Condron, Stephanie, and Watkins, Simon. "Exposed: Barclays account details for sale as 'gold mine' of up to 27,000 files is leaked in worst breach of bank data EVER." *Daily Mail*, February 8, 2014. http://www.dailymail.co.uk/news/article-2554875/Barclays-account-details-sale-gold-27-000-files-leaked.html (accessed August 12, 2016).

Gangadharan, Seeta Pena. "Digital inclusion and data profiling." *First Monday* 17 (May 2012). http://firstmonday.org/ojs/index.php/fm/article/viewArticle/3821/3199 (accessed October 16, 2015).

Gangahar, Anuj, and Jones Adam. "BNP Paribas investment funds hit by volatility." *Financial Times*, August 9, 2007. https://www.ft.com/content/9a4cabc4-464d-11dc-a3be-0000779fd2ac (accessed October 16, 2015).

Gao, Huiji, Barbier, Geoffrey, and Goolsby, Rebecca. "Harnessing the Crowdsourcing Power of Social Media for Disaster Relief." *IEEE Computer Society* 26, no. 3 (May/June, 2011): 10-14. http://isivast.org.vn:8888/bitstream/123456789/11336/1/Intelligent%20systems%20and%20their%20application.Vol.26.Iss.3.A.1.pdf (accessed October 16, 2015).

Garcia Rey, Marcos, and Garrido, Hugo. "Datos sepultados bajo la ceniza (Metodología)." España en Llamas, July 25, 2013. http://espanaenllamas.es/datos-sepultados-bajo-la-ceniza-metodologia/ (accessed April 20, 2016).

Gaus, Gerald, Courtland, Shane D., and Schmidtz, David. "Liberalism." *The Stanford Encyclopedia of Philosophy,* edited by Edward N. Zalta, Spring 2015. https://plato.stanford.edu/entries/liberalism/ (accessed April 20, 2016).

Gavilanes, M. Ángel. "249 alcaldes piden su reelección el 24M sin haber rendido cuentas en toda la legislatura." *Quien manda*, May 5, 2015. http://quienmanda.es/posts/alcaldes-opacos (accessed October 16, 2014).

Global Editors Network, Web site "Data Journalism Awards," 2014. http://www.globaleditorsnetwork.org/programmes/data-journalism-awards/ (accessed October 16, 2014).

Global Facility for Disaster Reduction and Recovery Labs, Web site. "Volunteer Technology Communities (VTC): Open Development." Global Facility for Disaster Reduction and Recovery Labs, 2010. https://www.gfdrr.org/sites/gfdrr.org/files/publication/Volunteer%20Technology%20Communities%20-%20Open%20Development.pdf (accessed October 16, 2014).

Gibbs, Mark. "Predicting crime with Big Data... Welcome to 'Minority Report'" for real." *Network World*, September 20, 2014. http://www.networkworld.com/article/2686051/big-data-business-intelligence/predicting-crime-with-big-data-welcome-to-minority-report-for-real.html (accessed February 5, 2015).

Gillmor, Dan. "Objectivity is a worthy ideal— But we are only human." *The Economist*, July 29, 2011. http://www.economist.com/ideasarena/news/by-invitation/guest-contributions/objectivity-worthy-ideal%E2%80%94-we-are-only-human (accessed March 6, 2016).

—. "Ferguson's citizen journalists revealed the value of an undeniable video." *The Guardian*, August 16, 2014. https://www.theguardian.com/commentisfree/2014/aug/16/fergusons-citizen-journalists-video (accessed March 6, 2016).

Gitelman, Lisa (ed.). '*Raw Data*' *Is an Oxymoron.* Cambridge: The MIT Press, 2012.

Global Pulse, Web site. "About*,"* 2016. http://www.unglobalpulse.org/about-new (accessed March 8, 2016).

Global Open Data Index, Web site. "Place overview," 2014. http://index.okfn.org/place/ (accessed June 23, 2016).

Global Reporting Initiative, Web site. "GRI at a Glance," 2016. https://www.globalreporting.org/Pages/default.aspx (accessed March 23, 2016).

Goldenberg, Suzanne. "Greenpeace exposes sceptics hired to cast doubt on climate science." *The Guardian*, December 8, 2015. https://www.theguardian.com/environment/2015/dec/08/greenpeace-exposes-sceptics-cast-doubt-climate-science (accessed March 23, 2016).

Goodwin, Jeff, and M. Jasper, James. *The Social Movements Reader: Cases and Concepts*. Malden, Oxford and West Sussex: Wiley-Blackwell, 2014.

Gorenstain-Massa, Felipe. "Insurgency on the internet: Organising the Anonymous online community." (PhD diss., Boston College, 2013). https://dlib.bc.edu/islandora/object/bc-ir:104073/datastream/PDF/view (accessed March 23, 2016).

Grabowski, Maria. "A Movement of Hundreds of Small Movements." *Ushahidi Blog*, May 31, 2012. https://www.ushahidi.com/blog/2012/05/31/a-movement-of-hundreds-of-small-movements (accessed August 16, 2016).

—. "Hopeful Action: An anthropological take on Ushahidi, Digital Activism and Hope." *Ushahidi Blog*, August 15, 2013. https://www.ushahidi.com/blog/2013/08/15/hopeful-action-an-anthropological-take-on-ushahidi-digital-activism-and-hope (accessed August 11, 2016).

Greenberg, Andy. "These Are the Emails Snowden Sent to First Introduce His Epic NSA Leaks." *Wired*, October 13, 2014. https://www.wired.com/2014/10/snowdens-first-emails-to-poitras/ (accessed August 11, 2016).

Greenwald, Glenn. "NSA collecting phone records of millions of Verizon customers daily." *The Guardian*, June 6, 2013. https://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order (accessed August 11, 2016).

Grossman, Lev. 2006. 'You — Yes, You — Are TIME's Person of the Year'. *Time*, December 25. http://content.time.com/time/magazine/article/0,9171,1570810,00.html (accessed August 11, 2016).

GSR, Web site. "Big data," *GSR Technologies, Inc.*, 2015. http://www.gsrti.com/technologies/big-data/ (accessed January 15, 2016).

The Guardian, Web site. "Keep it in the ground," 2016. http://www.theguardian.com/environment/series/keep-it-in-the-ground (accessed February 17, 2016).

Guardian Project, Web site. "Abous us," 2016. https://guardianproject.info/ (accessed May 12, 2016).

Gutiérrez, Miren. “Special Report: The ‘Prop-Agenda’ at War.” *Inter Press Service*, June 27, 2004. http://www.ipsnews.net/2004/06/special-report-the-prop-agenda-at-war/ (accessed May 12, 2016).

—. “De la acción al activismo.” In *Reflexiones sobre una comunicación que transforma*, edited by Raquel Martínez-Gómez and Mario Lubetkin, 225-230. Santander: Universidad Internacional Menéndez Pelayo (UIMP) - Inter Press Service (IPS), 2012. http://ipsnews.net/documents/reflexionessobreunacomunicacionquetransforma.pdf (accessed May 12, 2016).

—. “Feature: Adapting to flood and fire in Quito, Ecuador.” *Climate & Development Knowledge Network*, January 26, 2015. http://cdkn.org/2015/01/feature-much-water-little-adapting-flood-drought-quito/?loclang=en_gb (accessed May 12, 2016).

Gutiérrez, Miren, McFarland, Will, and Fonua, Lano. *Zero poverty... think again: the impact of climate change on development efforts.* London: Overseas Development Institute, 2014. https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/8863.pdf (accessed May 12, 2016).

Habermas, Jürgen. *Knowledge and Human Interests.* Boston: Beacon Press, 1968.

—. *Theory and Practice.* Cambridge: Beacon Press, 1973.

—. *The Theory of Communicative Action.* Vol. 1. Cambridge: Beacon Press, 1984.

—. *Toward a Rational Society: Student Protest, Science, and Politics.* Boston: Beacon Press (MA), 1989.

—. *The Structural Transformation of the Public Sphere: An Inquiry Into a Category of Bourgeois Society.* Cambridge: MIT Press, 1991.

—. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy.* Cambridge: MIT Press, 1996.

—. *The Postnational Constellation: Political Essays.* Cambridge: MIT Press, 2001.

—. “Political communication in media society: Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research.” *Communication Theory* 16 (November 2006): 411–426.

Hackett, Bob, and Carroll, Bill. *Remaking Media - The Struggle To Democratize Public Communication (Communication and Society).* New York and London: Routledge, 2006.

Hacks/Hackers Africa, Web site. "About," 2016. https://github.com/HacksHackersAfrica/wiki/wiki/About-Hacks-Hackers-Africa (accessed April 6, 2016).

Hacks/Hackers, Web site, 2016. http://hackshackers.com/ (accessed October 15, 2014).

Hagan, Jose. "Truth or Consequences." *Texas Monthly* (May, 2012). http://www.texasmonthly.com/politics/truth-or-consequences/ (accessed October 15, 2014).

Haimson, Oliver L., and Cartagena, Josh. *Information Occupation: Using Information Science to Explore Occupy Wall Street*. iConference, Fort Worth, TX, USA, February 12-15, 2013. https://www.ideals.illinois.edu/bitstream/handle/2142/42045/353.pdf?sequence=2 (accessed October 15, 2014).

Hamilton, Alexander. "The Consequences of Hostilities Between the States." *The Federalist Papers* 8 (November 20, 1787). http://avalon.law.yale.edu/18th_century/fed08.asp (accessed October 15, 2014).

Hammersley, Martyn. *The Dilemma of Qualitative Method: Herbert Blumer and the Chicago Tradition.* Oxford: Routledge, 1990.

Hammond, Jeremy. "I'm an Anonymous hacker in prison, and I am not a crook. I'm an activist." *The Guardian*, December 18, 2014. https://www.theguardian.com/commentisfree/2014/dec/18/anonymous-hacker-prison-jeremy-hammond-hacktivism (accessed October 15, 2015).

Hand, Martin. "From Cyberspace To The Dataverse: Trajectories In Social Digital Research." In *Big Data? Qualitative Approaches to Digital Research - Studies in Qualitative Methodology* 13 (November 6, 2014): 1-27 http://dx.doi.org/10.1108/S1042-319220140000013002 (accessed October 15, 2015).

Hankey, Stephanie, and Clunaigh, Danniel O. "Rethinking Risk and Security of Human Rights Defenders in the Digital Age." *Oxford Journals* (Tactical Tech), November 2013: 535-547. https://tacticaltech.org/rethinking-risk-and-security-human-rights-defenders-digital-age (accessed October 15, 2015).

Harford, Tim. "Big data: are we making a big mistake?" *Financial Times*, March 28, 2014. https://www.ft.com/content/21a6e7d8-b479-11e3-a09a-00144feabdc0 (accessed October 15, 2015).

Harkin, James. *Mobilisation: The growing public interest in mobile technology.* London: Demos, 2003.

Harris, Gardiner. "Borrowed Time on Disappearing Land." *The New York Times*, March 28, 2014. https://www.nytimes.com/2014/03/29/world/asia/facing-rising-seas-bangladesh-confronts-the-consequences-of-climate-change.html?_r=0 (accessed October 15, 2015).

Harvey, David. *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change.* Oxford: Wiley-Blackwell, 1991.

Hawley, Charles. "The World from Berlin: 'Every User Can Decide Alone What Facebook Knows.'" *Der Spiegel*, June 8, 2012. http://www.spiegel.de/international/germany/critique-of-german-credit-agency-plan-to-mine-facebook-for-data-a-837713.html (accessed October 15, 2015).

Hayles, Katherine N. "Print Is Flat, Code Is Deep: The Importante of of Media-Specific Analysis." *Poetics Today* 24, no. 1 (2004): 67-90. http://www.cws.illinois.edu/IPRHDigitalLiteracies/Hayles.pdf (accessed October 15, 2015).

Hegel, Georg Wilhelm Friedrich. *Phenomenology of Mind.* Blackmask Online, 2008 [1807].

Hellerstein, Joseph. "The Commoditization of Massive Data Analysis." *Radar*, November 19, 2008. http://radar.oreilly.com/2008/11/the-commoditization-of-massive.html (accessed October 15, 2015).

Hersman, Erik. "Official Bio." *White African*, 2016. http://whiteafrican.com/about/ (accessed October 23, 2016).

Heskett, Jim. "What Are The Limits Of Transparency?" *Forbes*, July 10, 2013. http://www.forbes.com/sites/hbsworkingknowledge/2013/07/10/what-are-the-limits-of-transparency/#165619836ad9 (accessed October 23, 2016).

Hess, David J. *Technology- and Product-Oriented Movements: Approximating Social Movement Studies and Science and Technology Studies.* Sage Journals 30, no. 4 (October 1, 2005): 515–535. http://journals.sagepub.com/doi/pdf/10.1177/0162243905276499 (accessed April 30, 2016).

Hess, David, Steve Breyman, Nancy Campbell and Brian Martin. "Science, technology, and social movements." In *The handbook of science and technology studies*, edited by Edward Hackett, Olga Amsterdamska, Michael Lynch and Judy Wajcman. Cambridge: The MIT Press, 2008.

Hesse, M. "Crisis mapping brings online tool to Haitian disaster relief effort." *The Washington Post*, January 16, 2010. http://www.washingtonpost.com/wp-dyn/content/article/2010/01/15/AR2010011502650.html (accessed April 30, 2016).

Hey, Tony, Tansley, Stewart, and Tolle, Kristin, eds. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond: Microsoft Research, 2009. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf (accessed April 30, 2015).

Hilbert, Martin, Web site. "The world's technological capacity to store, communicate, and compute information." *MartinHilbert.net*, 1 Abril 2011. http://www.martinhilbert.net/WorldInfoCapacity.html/ (accessed January 27, 2015).

Hin, Saskia. "Counting Romans." *Princeton/Stanford Working Papers in Classics* (November 2007): 1-37. https://www.princeton.edu/~pswpc/pdfs/hin/110703.pdf (accessed January 27, 2015).

Hine, Christine. "Internet Research and Unobtrusive Methods." *Social Research Update - University of Surrey* 61 (Spring 2011): 1-4. http://sru.soc.surrey.ac.uk/SRU61.pdf (accessed January 27, 2015).

Hivos. *theoryofchange.nl.* Hivos. 2015. http://www.theoryofchange.nl/ (accessed December 12, 2015).

Hobbes of Malmesbury, Thomas. *Leviathan or the Matter, Forme, & Power of a Common-wealth Ecclesiasticall and Civill.* London: Andrew Crooke, 1651.

Hogan, Mél, interviewed by Sarah T. Roberts. "Data Flows and Water Woes: An Interview With Media Scholar Mél Hogan." *Big Data & Society* (August 12, 2015). http://bigdatasoc.blogspot.com.es/2015/08/data-flows-and-water-woes-interview.html (accessed May 12, 2016).

Holmes, Harlo, Web site. "Introducing InfomaCam." Guardian Project, January 20, 2012. https://guardianproject.info/2012/01/20/introducing-informacam/ (accessed May 12, 2016).

Horkheimer, Max, Theodor Adorno, Gunzelin Noeri and Edmund Jephcott. *Dialectic of Enlightenment.* Stanford: Stanford University Press, 2002 [1944].

Howard, Alexander. "The Art and Science of Data-Driven Journalism." Tow Center for Digital Journalism, December 3, 2014. http://towcenter.org/research/the-art-and-science-of-data-driven-journalism/ (accessed February 17, 2016).

Howard, Philip N. *The Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam.* Oxford: Oxford Studies in Digital Politics, 2010.

Humanitarian Policy Group. "Saving lives through livelihoods: critical gaps in the response to the drought in the Greater Horn of Africa." HPG - Briefing papers, May, 2006.

https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/2041.pdf (accessed February 17, 2016).

Huesemann, Michael. "Why Technology Can't Save Us." IFG Teach-In: Techno-Utopianism & The Fate of the Earth, Great Hall of the Cooper Union, New York, October 25, 2014. https://ratical.org/ratville/AoS/MHuesemann102514.html (accessed February 17, 2016).

Huesemann, Michael, and Huesemann, Joyce. *Techno-Fix: Why Technology Won't Save Us or the Environment.* Gabriola Island: New Society Publishers, 2011.

HumTech, Web site. "Humanitarian Technology: Science, Systems and Global Impact 2015." http://www.humanitariantechnology.org/ (accessed August 6, 2016).

Husserl, Edmund, and Carr, David. *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*. Evanston: Northwestern University Press, 1970.

IBM. *Demystifying Big Data: Decoding The Big Data Commission Report*. *IBM presentation*, 2012. https://www-304.ibm.com/events/wwe/grp/grp004.nsf/vLookupPDFs/Tim%20Paydos'%20Presentation/$file/Tim%20Paydos'%20Presentation.pdf (accessed December 12, 2015).

IBM, Web site. "What is Big Data?" https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html (accessed December 12, 2015).

International Federation of Social Workers, Web site. "Global Definition of Social Work," July 2014. http://ifsw.org/get-involved/global-definition-of-social-work/ (accessed December 12, 2015).

Indymedia, Web site. "About*,"* 2016. https://www.indymedia.org/or/index.shtml (accessed April 3, 2016).

InfoAmazonia, Web site. "About*,"* 2016. http://infoamazonia.org/en (accessed April 3, 2016).

Innerarity, Daniel. "Ciudades culturalmente inteligentes." *Globernance.org*, May 2, 2016a. http://www.danielinnerarity.es/opinion-preblog/ciudades-culturalmente-inteligentes/ (accessed November 12, 2016).

—. "Ricos y pobres en datos." *Globernance.org*, February 22, 2016b. http://www.danielinnerarity.es/opinion-preblog/ricos-y-pobres-en-datos/ (accessed November 12, 2016).

Innis, Harold Adams. *The Bias of Communication.* Toronto: University of Toronto Press, 2008 [1951].

—. *Empire & communications.* Victoria, Toronto: Press Porcepic, 1986.

International Consortium of Investigative Journalists. "The Panama Papers," 2016. https://panamapapers.icij.org/ (accessed April 4, 2016).

Internet Live Stats, Web site. "Twitter Usage Statistics," 2015. http://www.internetlivestats.com/twitter-statistics/ (accessed February 8, 2015).

Internews Centre for Innovation & Learning. "Mapping the Maps: A Meta-Level Analysis of Ushahidi & Crowdmap." *Sage Journals* 22, No. 6 (2012). http://journals.sagepub.com/doi/abs/10.1177/1354856515584320 (accessed February 8, 2015).

Ismael, Chris. "List of 20 sentiment analysis apis." *MashApe,* April 24, 2013. http://blog.mashape.com/list-of-20-sentiment-analysis-apis/ (accessed February 8, 2015).

Israel, Jonathan I. *Enlightenment Contested: Philosophy, Modernity, and the Emancipation of Man 1670-1752.* Oxford: Oxford University Press, 2006.

International Telecommunication Union (ITU). "Measuring the Information Society Report." ITU, 2014. https://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2014/MIS2014_without_Annex_4.pdf (accessed February 8, 2015).

—. "ICT Development Index 2015." ITU, 2015. http://www.itu.int/net4/ITU-D/idi/2015/index.html (accessed August 25, 2016).

Izquierdo, José María. "En el futuro será factible comunicarse con el pensamiento." *El Pais*, December 7, 2015. http://elpais.com/elpais/2015/12/04/eps/1449250756_177381.html (accessed August 25, 2016).

Jarvie, Ian C. *Concepts and Society.* New York: Routledge, 2015.

Jeff, Goodwin, Jasper, James M., and Polletta, Francesca. *Passionate Politics: Emotions and Social Movements.* Chicago: University Of Chicago Press, 2001.

Jeffery, Simon. "Ushahidi Crowmap: Kenua Violence." *The Guardian,* April 11, 2011. https://www.theguardian.com/news/blog/2011/apr/07/ushahidi-crowdmap-kenya-violence-hague (accessed August 7, 2016).

Jordan, John. "Hedge Fund Will Track Twitter to Predict Stock Moves." *Bloomberg Business*, December 22, 2010. http://www.investmentweek.co.uk/investment-week/news/1934279/hedge-fund-track-twitter-predict-stock-moves (accessed August 7, 2016).

Jordan, Tim. *Activism!: Direct Action, Hacktivism and the Future of Society.* London: Reaktion Books Ltd, 2004.

Joyce, Mary, ed. *Digital Activism Decoded: The New Mechanics of Change.* New York and Amsterdam: International Debate Education Association, 2010.

Joyce, Mary. "Global Digital Activism Data Set 2.0: Inclusion Criterion and Search Patterns." *Digital Acitivsm Research Project*, May 13, 2013. http://digital-activism.org/2013/05/inclusion-criterion-and-search/

Kalathil, Shanthi, and Boas, Taylor. *Open Networks, Closed Regimes: The Impact of the Internet on Authoritarian Rule.* Washington: Carnegie Endowment for International Peace, 2003.

Kampmark, Binoy. "Her Majesty's Police State and Britain's Regulation of Investigatory Powers Act (RIPA): Keeping National Security Vague." *Global Research*, January 18, 2015. http://www.globalresearch.ca/her-majestys-police-state-and-britains-regulation-of-investigatory-powers-act-ripa-keeping-national-security-vague/5425389 (accessed July 18, 2016).

Kant, Immanuel. "Answering the Question: What is Enlightenment?" *Marxists.org*, 1996 [1784]. https://www.marxists.org/reference/subject/ethics/kant/enlightenment.htm (accessed July 18, 2016).

Kaplan, David E. "Why Open Data Isn't Enough." *Global Investigative Journalism Network*, April 2, 2013. http://gijn.org/2013/04/02/why-open-data-isnt-enough/ (accessed July 18, 2016).

Katrandjian, Olivia. "Hurricane Irene: Pop-Tarts Top List of Hurricane Purchases." *ABC News*, August 27, 2011. http://abcnews.go.com/US/hurricanes/hurricane-irene-pop-tarts-top-list-hurricane-purchases/story?id=14393602 (accessed January 29, 2015).

Kebo, David. "These Smart, Social Apps Bring Big Data Down to Size." *US Analytics*, October 2012. http://www.forbes.com/sites/markfidelman/2012/10/30/these-smart-social-apps-bring-big-data-down-to-size/#5c494a014985 (accessed January 29, 2015).

Keck, Margaret E., and Sikkink, Kathryn. *Activists Without Borders: Transnational advocacy networks in international and regional politics*. Ithaca and London: Cornell University Press, 1998.

Keim, Brandon. "Open Source for Humanitarian Action." *Standford Social Innovation Review* (Winter 2012). http://ssir.org/articles/entry/open_source_for_humanitarian_action (accessed August 5, 2016).

Kekic, Laza. "The Economist Intelligence Unit's index of democracy." *The Economist* (2007): 1-11. http://www.economist.com/media/pdf/DEMOCRACY_INDEX_2007_v3.pdf

Kemp, Simon. *Social, Digital & Mobile Worldwide report.* We Are Social report, Singapore, 2014. http://wearesocial.com/uk/special-reports/social-digital-mobile-worldwide-2014

Kennedy, Helen, Thomas, Poell, and van Dijck, Jose. "Data and agency." *Big Data & Society* (July-December 2015): 1-7. https://www.researchgate.net/publication/287107548_Data_and_agency (accessed December 30, 2016).

Khamis, Sahar, and Vaughn, Katherine. "Cyberactivism in the Egyptian Revolution: How Civic Engagement and Citizen Journalism Tilted the Balance." *Arab Media and Society* 13 (Summer 2011): 1-37. http://www.arabmediasociety.com/articles/downloads/20110603105609_Khamis.pdf (accessed December 19, 2015).

Kiln, Web site. "Carbon emissions: past, present and future – interactive," December 1, 2014. http://www.theguardian.com/environment/ng-interactive/2014/dec/01/carbon-emissions-past-present-and-future-interactive (accessed December 19, 2015).

Kiln, Web site. "In Flight." *The Guardian*, 2014. https://www.theguardian.com/world/ng-interactive/2014/aviation-100-years (accessed December 19, 2015).

Kirkpatrick, David D. "Tunisia Leader Flees and Prime Minister Claims Power." *The New York Times.* January 14, 2011. http://www.nytimes.com/2011/01/15/world/africa/15tunis.html?pagewanted=all&_r=0 (accessed December 19, 2015).

Kitchin, Rob. "Big Data, new epistemologies and paradigm shifts." *Big Data & Society* (April-June 2014): 1-12. http://journals.sagepub.com/doi/abs/10.1177/2053951714528481

Kitchin, Rob and Lauriault, Tracey P. "Small Data, Data Infrastructures and Big Data." *GeoJournal - The Programmable City Working Paper* 80, no. 4 (2014): 463-475. http://ssrn.com/abstract=2376148 (accessed August 8, 2016).

Knight Foundation. *Uchaguzi: A Case Study - Successes, Challenges, and New Ways Forward.* Report, 2010. http://www.knightfoundation.org/media/uploads/media_pdfs/uchaguzi-121024131001-phpapp02.pdf (accessed August 8, 2016).

Kobia, David, Web site. "LinkedIn*,"* 2016. https://www.linkedin.com/in/kobia (accessed August 5, 2016).

Kostelanetz, Richard. "Understanding McLuhan (In Part)." *The New York Times Magazine*, January 29, 1967. http://www.nytimes.com/books/97/11/02/home/mcluhan-magazine.html (accessed August 8, 2016).

Kovach, Bill, and Rosenstiel, Tom. *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect, Completely Updated and Revised.* New York: Three Rivers Press, 2007.

Kramer, Adam D. I., Guillory, Jamie E., and T. Hancockb, Jeffrey. "Experimental evidence of massive-scale emotional contagion through social networks." *CrossMark (PNAS)* 111, no. 24 (June, 2014): 8788–8790. http://www.pnas.org/content/111/24/8788.full (accessed August 8, 2016).

Krugman, Paul. "We Are the 99.9%." *The New York Times*, November 24, 2011. http://www.nytimes.com/2011/11/25/opinion/we-are-the-99-9.html (accessed August 8, 2016).

Kuhn, Thomas. *The Structure of Scientific Revolutions.* Chicago: The University of Chicago Press, 1970.

Kumar, Satish, interviewed by Miren Gutiérrez. "Todos tenemos un papel para cambiar la sociedad - Un viaje desde el interior de cada uno a la transformación del planeta." *Cuerpo* Mente (2015): 88-93.

Labarga, Alberto. "Vidas Contadas en el Seminario Visualizar'15 ." *Youtube*, May 11, 2015. https://www.youtube.com/watch?v=9JNhK69rzbU (accessed August 8, 2016).

Lancaster, Lewis. "From Text to Image to Analysis: Visualization of Chinese Buddhist Canon." *Digital Humanities - University of California* (2010): 1-3. http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/pdf/ab-670.pdf (accessed August 8, 2016).

Laney, Doug. "3D Data Management: Controlling Data Volume, Velocity and Variaty." *META Group* (February 6, 2001): 1-4. https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (accessed August 8, 2016).

Latour, Bruno. *Science in Action: How to Follow Scientists and Engineers Through Society* Harvard: Harvard University Press: 1988 [1987]

—. "Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern." *Critical Inquiry - The University of Chicago: Special issue on the Future of Critique* 30 (Winter 2004): 225-248. http://www.bruno-latour.fr/sites/default/files/89-CRITICAL-INQUIRY-GB.pdf (accessed August 8, 2016).

—. *Reassembling the Social: An Introduction to Actor-Network-Theory.* Oxford: Oxford University Press USA, 2005.

Latour, Bruno, Pablo Jensen, Tommaso Venturini, Sébastian Grauwin and Dominique Boullier. "The whole is always smaller than its parts: A digital test of Gabriel Tarde'ss monads." *The British Journal of Sociology* 63, no. 4 (2012): 590-615. http://www.bruno-latour.fr/sites/default/files/123-MONADS-BJSpdf.pdf (accessed August 8, 2016).

Law, John. "Actor Network Theory and Material Semiotics." Lancaster University (April 25, 2007): 1-21. https://pdfs.semanticscholar.org/b602/fd19b4b5cc4cad9d28752b65fe8d01dd7495.pdf (accessed August 8, 2016).

—. "Notes on the Theory of the Actor Network. Ordering, Strategy and Heterogeneity." Heterogeneities.net (2011): 379-93 [1992]. http://www.heterogeneities.net/publications/Law1992NotesOnTheTheoryOfTheActor-Network.pdf (accessed December 12, 2016).

Lawrence, Neil. "Why the Ownership of Personal Data in the Hands of a Few Should Worry Us." *The Guardian*, March 5, 2015. http://www.alternet.org/media/why-ownership-personal-data-hands-few-should-worry-us (accessed December 12, 2016).

Lazer, David et al. "Computational Social Science." Science 323, no.5915 (February 6, 2009): 721-723.

Lazer, David, and Kennedy, Ryan. "What We Can Learn From the Epic Failure of Google Flu Trends." *The Wire*, January 10, 2015. https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/ (accessed December 12, 2016).

Leetaru, Kalev. "Do Big Data Results Depend On What Data We Look At?" *Forbes.* October 5, 2015a. http://www.forbes.com/sites/kalevleetaru/2015/10/05/do-big-data-results-depend-on-what-data-we-look-at/#24d10ddf13c3 (accessed December 12, 2016).

—. "How Drones Are Changing Humanitarian Disaster Response." *Forbes*, November 9, 2015b. http://www.forbes.com/sites/kalevleetaru/2015/11/09/how-drones-are-changing-humanitarian-disaster-response/#424196396cee (accessed December 12, 2016).

Leonard, Andrew. "The Cybercommunist Manifesto: Are free-software hackers undermining capitalism and the free-market economy with their code giveaways?" *Salon*, September 10, 1999. http://www.salon.com/1999/09/10/cybercommunism/ (accessed December 12, 2016).

Leonelli, Sabina. "What difference does quantity make? On the epistemology of Big Data in biology." *Big Data & Society* (April-June 2014): 1-11. http://journals.sagepub.com/doi/abs/10.1177/2053951714534395 (accessed December 12, 2016).

Leson, Heather. “Ushahidi - Where in the World.” *Ushahidi Blog*, August 5, 2013. https://www.ushahidi.com/blog/2013/08/05/where-in-the-world (accessed August 6, 2016).

Lessig, Laurence. *Code 2.0.* New York: Basic Books, 2006.

Letouzé, Emmanuel, and Sangokoya, David. “Leveraging Algorithms for Positive Disruption: On data, democracy, society and statistics.” Working Paper, Data-Pop Alliance, 2015. http://datapopalliance.org/wp-content/uploads/2016/03/DataPopAlliance_LeveragingAlgorithms.pdf (accessed December 12, 2016).

Liew, Anthony. “Understanding Data, Information, Knowledge And Their Inter-Relationships.” *Journal of Knowledge Management Practice* 8, no. 2 (June, 2007): 1-10. http://www.tlainc.com/articl134.htm (accessed December 12, 2016).

Light, Jennifer S. “When computer were women.” *Technology and Culture* 40, no. 3 (July, 1999): 455-483. http://labweb.education.wisc.edu/elpa940/readings/Light.pdf (accessed December 12, 2016).

Lim, Merlyna. “Framing Bouazizi: ‘White lies,’ hybrid network, and collective/connective action in the 2010–11 Tunisian uprising.” *Journalism (Sage)* 14, no. 7 (2013): 921-941. http://journals.sagepub.com/doi/abs/10.1177/1464884913478359 (accessed December 12, 2016).

Linstone, Harold A., and Mitroff, Ian I.. *Challenge of the 21st Century, The: Managing Technology and Ourselves in a Shrinking World*. New York: State University of New York Press, 1994.

Lipovetsky, Gilles. “De la régulation collective à l'arbitrage personnel.” *Prodimarques - Revue des Marques* 64, October, 2008. http://www.prodimarques.com/documents/gratuit/64/de-la-regulation-collective-a-larbitrage.php (accessed December 12, 2016).

—. *De la ligereza Hacia una civilización de lo ligero*. Barcelona: Anagrama, 2016 [2015].

Lipovetsky, Gilles, and Serroy, Jean. *La Pantalla Global: Cultura mediática y cine en la era hipermoderna*. Barcelona: Editorial Anagrama, 2009 [2007].

Lippmann, Walter. *Public Opinion*. New Brunswick and London: Transaction Publishers, 2009 [1922].

Little, Daniel. “Mapping sociology.” *Understanding Society*, August 5, 2011. http://understandingsociety.blogspot.co.uk/2011/08/mapping-sociology.html (accessed December 29, 2015).

Lloyd, W. F. *Two lectures on the checks to population.* Oxford: The University of Oxford, 1833.

Lohr, Steve. "The Age of Big Data." *The New York Times*, February 2012. http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html (accessed December 29, 2015).

Lorenz, Mirko. "Why Journalists Should Use Data." In *Data Journalism Handbook*, 2012. http://datajournalismhandbook.org/1.0/en/introduction_1.html (accessed February 15, 2016).

Lowenstein, Jeff Kelly. "Crunch Time: Black people and jaywalking in Champaign." *ViveloHoy*, 2012a. http://www.vivelohoy.com/noticias/8123062/crunch-time-black-people-and-jaywalking-in-champaign (accessed Ausgust 17, 2016).

—. "Arrest disparities for black people in Champaign, Urbana." *ViveloHoy*, May 8, 2012b. http://www.vivelohoy.com/noticias/8121347/arrest-disparities-for-black-people-in-champaign-urbana (accessed Ausgust 17, 2016).

Luiselli, Valeria. "Mapas: Sospecho que no serán maputopías, sino cartografías "programadas a la medida" de la pesadilla que estamos alimentando." *El País*, February 20, 2017. http://elpais.com/elpais/2017/02/19/opinion/1487519835_490390.html (accessed February 20, 2017).

Lukes, Steven, and Urbinati Nadia. *Condorcet: Political Writings.* Cambridge : Cambridge University Press, 2012.

Lynch, Sarah. "Tunisia, Arab Spring's birthplace, takes on militants." *USA Today*, March 23, 2014. http://www.usatoday.com/story/news/world/2014/03/22/tunisia-revolution-arab-spring-radicals/5705461/ (accessed Ausgust 17, 2016).

Lynd, Robert Staughton. *Knowledge for What: The Place of Social Science in American Culture.* Princeton: Princeton University Press, 1967.

Lyon, David. "Surveillance, Snowden, and Big Data: Capacities, consequences, critique." *Big Data & Society* (July-December 2014): 1-13. http://journals.sagepub.com/doi/pdf/10.1177/2053951714541861 accessed February 15, 2016).

MacKinnon, Rebecca. *Consent of the Networked. The Worldwide Struggle for Internet.* New York: Basic Books, 2012.

Mann, Steve, Nolan, Jason, and Wellman, Barry. "Sousveillance: Inventing and Using Wearable. Computing Devices for Data Collection in Surveillance Environments." *Surveillance & Society* 1, no. 3 (2003): 331-335. http://www.surveillance-and-society.org/articles1(3)/sousveillance.pdf (accessed August 6, 2016).

Manning, Nat. "Changing Paradigms - A financially sustainble non-profit tech company." *Ushahidi Blog*, December 15, 2015. https://www.ushahidi.com/blog/2015/12/15/changing-paradigms-a-financially-sustainble-non-profit-tech-company (accessed August 6, 2016).

Manyika, James et al. *Big data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute, May 2011. http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation (accessed August 6, 2016).

Martin, Chrissy. "Ushahidi Haiti Project Evaluation Final Report." *Ushahidi Blog*, April 19, 2011. https://www.ushahidi.com/blog/2011/04/19/ushahidi-haiti-project-evaluation-final-report (accessed August 6, 2016).

Marr, Bernard. "Can Big Data Algorithms Tell Better Stories Than Humans?" *Forbes*, July 22, 2015. http://www.forbes.com/sites/bernardmarr/2015/07/22/can-big-data-algorithms-tell-better-stories-than-humans/#665813cc42ba (accessed August 6, 2016).

—. "Is Data Socialization The Next Big Thing In Data Analytics?" *Forbes*, February 228, 2017. https://www.forbes.com/sites/bernardmarr/2017/02/28/is-data-socialization-the-next-big-thing-in-data-analytics/#5b503a334a1d (accessed February 28, 2017).

Marx, Karl, and Engels, Friedrich. "A Critique of the German Ideology." *Marxists.org*, 2000 [1932]. https://www.marxists.org/archive/marx/works/download/Marx_The_German_Ideology.pdf (accessed August 6, 2016).

Mayer, Jonathan, Mutchler, Patrick, and Mitchell, John C.. "Evaluating the privacy properties of telephone metadata." *CrossMark (PNAS)* 113, no. 20 (March, 2016): 5536–5541. http://www.pnas.org/content/113/20/5536.full.pdf (accessed August 6, 2016).

Mayer-Schönberger, Viktor, and Cukier, Kenneth. *Big data: A revolution that will transform how we live, work, and think.* Boston: Houghton Mifflin Harcourt, 2013.

McAfee, Andrew, and Brynjolfsson, Erik. *Big Data.* Harvard: Harvard Business Review, 2009.

McConnell, Tristan. "Witness Projection." *Harper's Magazine*, March 2013. http://harpers.org/archive/2013/03/witness-projection/ (accessed August 4, 2016).

McGreevy, Madeleine. "How data journalism is making an impact around the world." *Blog IODC 2015 - Ottawa*, June 3, 2015. http://opendatacon.org/how-data-journalism-is-making-an-impact-around-the-world/ (accessed August 4, 2016).

McKeever, Darin. "Moving from Big Data to Big Wisdom." *Forbes*, March 28, 2013. http://www.forbes.com/sites/skollworldforum/2013/03/28/moving-from-big-data-to-big-wisdom/#16a9ba996a15 (accessed August 4, 2016).

McLuhan, Marshall. *Understanding Media: The extensions of man.* London and New York: The MIT Press, 1964.

MediaEthics. "Objectivity and Advocacy in Journalism." *Media Ethics* 25, no.1 (Fall 2013). https://www.mediaethicsmagazine.com/index.php/browse-back-issues/179-fall-2013-vol-25-no-1/3999003-objectivity-and-advocacy-in-journalism (accessed August 4, 2016).

Medialab-Prado, Web site. "About us," 2016. http://medialab-prado.es/article/que_es (accessed April 3, 2016).

Meier, Patrick. "Ushahidi Haiti Project Evaluation Terms of Reference." *Ushahidi Blog*, 2010. https://www.ushahidi.com/blog/2011/04/19/ushahidi-haiti-project-evaluation-final-report (accessed February 8, 2015).

—. "Do Liberation Technologies Change the Balance of Power Between Repressive States and Civil Society?" Dissertation - The Fletcher School of Law and Diplomacy, November 2011a. https://irevolution.files.wordpress.com/2011/11/meier-dissertation-final.pdf (accessed February 8, 2015).

—. "Using Ushahidi Data to Study the Micro-Dynamics of Violent Conflict." *iRevolutions*, November 19, 2011b. https://irevolutions.org/2011/11/19/micro-dynamics-of-conflict/ (accessed February 8, 2015).

—. "Ushahidi as a Liberation Technology." In *Liberation Technology: Social Media and the Struggle for Democracy*, edited by Larry Diamond and Marc F. Plattner, 95-109. Baltimore: Johns Hopkins University Press: 2012. http://cddrl.fsi.stanford.edu/sites/default/files/meier_ushahidi_as_a_liberation_technology.pdf

—. "Crisis Maps: Harnessing the Power of Big Data to Deliver Humanitarian Assistance." *Forbes*, February 5, 2013a. http://www.forbes.com/sites/skollworldforum/2013/05/02/crisis-maps-harnessing-the-power-of-big-data-to-deliver-humanitarian-assistance/#782514d31533 (accessed August 23, 2016).

—. "Why Ushahidi Should Embrace Open Data." *iRevolutions*, February 3, 2013b. https://irevolutions.org/2013/02/03/crowdglobe/ (accessed August 23, 2016).

—. "Verily: Crowdsourced Verification for Disaster Response." *iRevolutions*, February 19, 2013c. https://irevolutions.org/2013/02/19/verily-crowdsourcing-evidence/

—. "Haiti: Lies, Damned Lies and Crisis Mapping." *iRevolutions*, February 26, 2013d. https://irevolutions.org/2013/02/26/haiti-lies/ (accessed August 24, 2016).

—. *Digital humanitarians: how big data is changing the face of humanitarian response.* Boca Raton, London and New York: CRC Press/Taylor & Francis Group, 2015a.

—. "Counter-Mapping the State with UAVs." *iRevolutions*, August 4, 2015b. https://irevolutions.org/2015/08/04/counter-mapping-with-uavs/ (accessed August 24, 2016).

—. "Using Computer Vision to Analyze Aerial Big Data from UAVs During Disasters." *iRevolutions*, October 12, 2015c. https://irevolutions.org/2015/10/12/computer-vision-big-data-uavs/ (accessed August 24, 2016).

—. "Video: Crisis Mapping Nepal with Aerial Robotics." *iRevolutions*, November 4, 2015d. https://irevolutions.org/2015/11/04/crisis-mapping-nepal-aerial-robotics/ (accessed August 24, 2016).

—. "Verifying Crowdsourced Social Media Reports for Live Crisis Mapping: An Introduction to Information Forensics." *iRevolutions*, 2015e. https://irevolution.files.wordpress.com/2011/11/meier-verifying-crowdsourced-data-case-studies.pdf (accessed August 24, 2016).

—. "About me," 2016. https://irevolutions.org/bio/ (accessed August 5, 2016).

Melucci, Alberto. *Challenging Codes: Collective Action in the Information Age.* Cambridge: Cambridge Cultural Social Studies, 1996a.

—. *The Playing Self: Person and Meaning in the Planetary Society.* Cambridge: Cambridge Cambridge Cultural Social Studies, 1996b [1991].

Mentalacrobatics, Web site. "About Daudi Were," 2016. http://mentalacrobatics.com/about/ (accessed May 16, 2016).

Merriam Webster. "Algorithm," 2016. http://www.merriam-webster.com/dictionary/algorithm (accessed May 16, 2016).

Michael, Michalis. *Top 10 Non-Governmental Organisations Worldwide.* July 9, 2014. http://www.careeraddict.com/top-10-non-governmental-organisations-worldwide (accessed August 6, 2016).

Milan, Stefania. "When politics and technology speak the same language: Stewardship in cyberspace according to cyberactivists." *CyberDialogue 2012* (March, 2012): 1-9. http://www.cyberdialogue.citizenlab.org/wp-

content/uploads/2012/2012papers/CyberDialogue2012_Milan.pdf (accessed August 6, 2016).

—. *Social Movements and Their Technologies.* Hampshire and New York: Palgrave Macmillan, 2013.

—. "The ethics of social movement research." In *Methodological Practices in Social Movement Research*, edited by Donatella della Porta, 446-465. Oxford: Oxford University Press, 2014a.

—. "Stealing the fire: Empowerment and resistance from the margins of cyberspace." Roskilde, September 18, 2014b. https://stefaniamilan.net/sites/default/files/Orecomm_MILAN_keynote.pdf (accessed August 6, 2016).

—. "Liberated technology: Inside emancipatory communication activism." *Gordon-Civic Media* (November 2015): 107-124. https://www.academia.edu/20076132/Liberated_technology_Inside_emancipatory_communication_activism (accessed August 6, 2016).

Milan, Stefania, and Gutiérrez, Miren. "Citizens' media meets Big Data: The emergence of data activism." *Mediaciones* 14 (January-June 2015): 120-133. http://biblioteca.uniminuto.edu/ojs/index.php/med/article/download/1086/1027 (accessed August 6, 2016).

Milan, Stefania, and Hintz, Arne. "Networked Collective Action and the Institutionalized Policy Debate: Bringing Cyberactivism to the Policy Arena?" *Policy & Internet* 5, no. 1 (March, 2013): 7-26.

Milan, Stefania, and van der Velden, Lonneke. "The alternative epistemologies of data activism." *Digital Culture & Society, Special Issue: The politics of big data* (October 10, 2016). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2850470

Milberry, Kate. "Tech activism in the global justice movement: Geeks, wikis and IMC." *Indymedia*, 2005. http://studylib.net/doc/8963084/tech-activism-in-the-global-justice-movement (accessed August 6, 2016).

Mill, John Stuart. *A System of Logic Ratiocinative and Inductive, Part II.* Toronto: University of Toronto Press, 1973 [1843].

Miller, Harvey J. "The Data Avalanche is Here: Shouldn't We Be Digging?" Journal of Regional Science 50, no.1 (April 8, 2009): 181-201. http://www.fednewyork.org/research/conference/2009/jrs/Miller.pdf (accessed April 30, 2015).

Milnes, Paul. "Whatever Happened to the 'Twitter Fund'?" *HedgeThink*, April 24, 2014. http://www.hedgethink.com/whatever-happened-twitter-fund/ (accessed April 30, 2015).

Misiroglu, Gina, ed. *The Handy Politics Answer Book.* Armonk, NY: Visible Ink Press, 2002.

MIT Technology Review. "50 Disruptive Companies: Ushahidi," 2013. http://www2.technologyreview.com/tr50/ushahidi/ (accessed August 6, 2016).

Mitchell, Amy. "The State of the News Media 2014." *Pew Research Center*, 2014. http://www.journalism.org/2014/03/26/state-of-the-news-media-2014-overview/ (accessed February 17, 2016).

Monroe, Burt L. "The Five Vs of Big Data Political Science: Introduction to the Virtual Issue on Big Data in Political Science." *Political Analysis* 19 (2011): 1-9.

Monterde, Arnau, and Postill, John. "Mobile ensembles: The uses of mobile phones for social protest by Spain's indignados." In *Routledge Companion to Mobile Media*, edited by Gerard Goggin and Larissa Hjorth. Abingdon-on-Thames: Routledge, 2013.

Morozov, Evgeny. *The net delusion: the dark side of Internet freedom.* New York: Public Affairs, 2011.

Morrow, Nathan, Mock, Nancy, and Papendieck, Adam. *Independent Evaluation of the Ushahidi Haiti Project.* ALNAP, Port-au-Prince, 2011. http://www.alnap.org/pool/files/1282.pdf (accessed February 17, 2016).

MultipleJournalism, Web site. "Ushahidi," 2008. http://multiplejournalism.org/case/ushahidi (accessed August 8, 2016).

Naim, Moises. *The End of Power: From Boardrooms to Battlefields and Churches to States, Why Being In Charge Isn't What It Used to Be.* New York: Basic Books, 2013.

La Nación. "Declaraciones juradas." *La Nacion*, 2014. http://www.lanacion.com.ar/declaraciones-juradas-t48961 (accessed August 8, 2016).

Large Synoptic Survey Telescope, Web site. "About," 2016. https://www.lsst.org/ (accessed August 8, 2016).

Naidoo, Kumi, interview by Miren Gutiérrez. "We shouldn't have to do the job of journalists." *Index on Censorship* 42, no.1 (2012): 110-113. http://journals.sagepub.com/doi/pdf/10.1177/0306422013481705

Naughton, John. "'Undemocratic, unnecessary, intolerable'… The official verdict on Britain's state snoopers." *The Guardian*, June 13, 2015.

http://www.theguardian.com/commentisfree/2015/jun/13/david-anderson-qc-investigatory-powers-report-gchq-undemocratic (accessed December 10, 2015).

Negroponte, Nicholas. *Being Digital.* New York: Alfred A. Knopf, Inc., 1996.

New York Times, Web site. "Reshaping New York," 2013. http://www.nytimes.com/newsgraphics/2013/08/18/reshaping-new-york/ (accessed October 16, 2014).

Newman, Bruce L. *A Review in Political Marketing: Lessons from Recent Presidential Elections.* Chicago: DePaul University, 2001.

Nicks, Denver. *Private: Bradley Manning, Wikileaks, and the biggest exposure of official secrets in American history.* Chicago. Chicago Review Press, 2012.

Nsehe, Mfonobong. "Africa's Most Successful Women: Ory Okolloh." *Forbes*, June 1, 2012. http://www.forbes.com/sites/mfonobongnsehe/2012/06/01/africas-most-successful-women-ory-okolloh/#33fe303beccd (accessed October 16, 2014).

O'Connor, Lydia. "San Francisco Train Passengers Too Distracted By Phones To Notice Shooter's Gun In Plain Sight." *The Huffington Post*, October 9, 2013. http://www.huffingtonpost.com/2013/10/08/san-francisco-train-shooting_n_4066930.html (accessed October 16, 2014).

ODI. "Resilience and humanitarian action," 2011. http://www.odi.org/projects/2359-resilience-humanitarian-livelihoods (accessed January 11, 2016).

—. "Building momentum in phasing out fossil fuel subsidies." Internal document, London, 2014.

—. "What is the future of official statistics in the Big Data era?" London, January 19, 2015. http://www.odi.org/events/4068-future-official-statistics-big-data-era (accessed January 20, 2015).

Organisation for Economic Co-operation and Development, Web site. "About the OECD," 2015. http://www.oecd.org/about/ accessed June 22, 2016).

—. Web site. "Open Data Berometer," 2016. http://opendatabarometer.org/3rdEdition/report/#executive_summary (accessed June 22, 2016).

Ogburn, William Fielding, ed. *Technology and International Relations.* Chicago: The University of Chicago Press, 1949. (accessed January 3, 2016).

Oil Change International, Web site. "About us," 2016a. http://priceofoil.org/about/ (accessed March 23, 2016a).

—. “The Price of Oil.” 2016b. http://priceofoil.org/ (accessed April 30, 2016b).

Okolloh, Ory. “Bear Witness,” *Kenyan Pundit Blog*, 2008. http://madkenyanwoman.blogspot.com.es/2008/01/bear-witness-kenya-pundit.html (accessed March 23, 2016).

Oliver, Guy. “NGOs against MONUSCO drones for humanitarian work.” *IRIN*, July 23, 2014. http://www.irinnews.org/report/100391/ngos-against-monusco-drones-humanitarian-work (accessed April 30, 2016).

Omenya, Rhoda. “Uchaguzi Kenya 2013: Monitoring and Evaluation.” *Ushahidi Blog*, July 8, 2013. https://www.ushahidi.com/blog/2013/07/08/uchaguzi-monitoring-and-evaluation-final-report-released (accessed January 27, 2016).

Open Street Map, Web site. “Humanitarian OSM Team,” April 20, 2016. http://wiki.openstreetmap.org/wiki/Humanitarian_OSM_Team (accessed May 4, 2016).

Open Knowledge, Web site. “What is open?” 2016. https://okfn.org/opendata/ (accessed May 4, 2016).

Ortiz-Ceberio, Cristina, and Rodríguez, María Pilar. “European cinema: Ethical debates on immigration laws.” In *The role of the humanities in times of crisis*, edited by María Pilar Rodríguez, 74-92. Madrid: Dykinson, S.L., 2016.

Owen, Taylor. *Disruptive Power: The Crisis of the State in the Digital Age.* Oxford: Oxford Studies in Digital Politics, 2015.

Oxfam, Web site. “Wealth: Having it all and wanting more,” 2015. https://www.oxfam.org/en/research/wealth-having-it-all-and-wanting-more (accessed May 4, 2016).

Oxford Dictionaries, Web site. “Word of the year is…,” 2016. https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2016 (accessed December 23, 2016).

Packer, Martin. *The Science of Qualitative Research.* Cambridge: Cambridge University Press, 2010.

Park, Robert E. *The Immigrant Press and Its Control.* New York and London: Harper & Brothers Publishers, 1922.

Parkin, Simon. “Operation Troll ISIS: inside Anonymous’ war to take down Daesh.” *Wired*, October 6, 2016. http://www.wired.co.uk/article/anonymous-war-to-undermine-daesh (accessed October 8, 2016).

Parker, Laura. "How 'Crisis Mapping' Is Shaping Disaster Relief in Nepal." *National Geographic* 22, May 1, 2015. http://news.nationalgeographic.com/2015/05/150501-nepal-crisis-mapping-disaster-relief-earthquake/ (accessed October 8, 2016).

Patterson, Orlando. "Freedom and 9/11." *Democracy: A journal of Ideas* (Summer Issue 2011). /http://democracyjournal.org/magazine/22/freedom-and-911/ (accessed October 8, 2016).

Pawlak, Patryk, and Ricci, Andrea, eds. "Crisis Rooms." *European Union Institute for Security Studies (EUISS)*, 2014. http://www.iss.europa.eu/uploads/media/Crisis_Rooms.pdf (accessed August 5, 2016).

Peluso, Nancy. "Whose Woods Are These? Territories In Kalimantan,. Indonesia. Counter-Mapping Forest." *Antipode* 27, no. 4 (October, 1995): 383-406. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8330.1995.tb00286.x/abstract (accessed April 18, 2016).

Peoples Global Action. "1st Conference of the Peoples' Global Action against 'Free' Trade and the WTO." Conference secretariat – Geneve, February 12, 1998. https://www.nadir.org/nadir/initiativ/agp/en/pgainfos/conference.htm (accessed April 18, 2016).

Periscopic, Web site. "Do good," 2016. http://www.periscopic.com/do-good (accessed May 4, 2016).

Pew Research. "Socially Responsible Algorithms for Data Science." *Data Science DC*, November 17, 2015. http://www.meetup.com/Data-Science-DC/events/226358392/ (accessed April 30, 2016).

Plage, Stefanie. *Mass Communication, Information Technology and Social Exclusion in Contemporary Society: Reconciling Public Sphere and Network Society?* Essex: Graduate Journal, 2009.

Poell, Thomas, Kennedy, Helen, and van Dijck, Jose. "Special Theme: Data & Agency." *Big Data & Society* (December 7, 2015). http://bigdatasoc.blogspot.com.es/2015/12/special-theme-data-agency.html (accessed July 12, 2016).

Populate, Web site. "Populate Tools," 2016. http://populate.tools/ (accessed July 12, 2016).

Porteous, Kimberley, Hudson, Michael, and Chavkin, Sasha. *International Consortium of Investigative Journalists*, December 4, 2014. http://www.icij.org/blog/2013/04/release-offshore-records-draws-worldwide-response (accessed February 23, 2016).

Portes, Alejandro and Rumbaut, Ruben G. *Legacies: The Story of the Immigrant Second Generation.* Berkeley: University of California Press: 2001.

Postigo, Hector. *The Digital Rights Movement.* Cambridge: MIT, 2012.

Prensky, Marc. "Digital Natives, Digital Immigrants." *On the Horizon - MCB University Press* 9, no.5, October 2001. http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf (accessed February 23, 2016).

Putnam, Robert D. *Bowling Alone: The Collapse and Revival of American Community.* New York: Touchstone Books by Simon & Schuster, 2001.

Pybus, Jennifer, Cote, Mark, and Blanke, Tobias. "Hacking the social life of Big Data." *Big Data & Society* (July-December 2015): 1-10. http://journals.sagepub.com/doi/pdf/10.1177/2053951715616649 (accessed February 23, 2016).

Radjawali, Irendra, and Pye, Oliver. "Counter-mapping Land Grabs with Community Drones in Indonesia." International Institute of Social Sciences, Conference Paper 80, Chiang Mai University, June 5-6, 2015. https://www.iss.nl/fileadmin/ASSETS/iss/Research_and_projects/Research_networks/LDPI/CMCP_80-Radjawali_and_Pye.pdf (accessed February 23, 2016).

Rafoth, Ellie. "Crisis Mapping: The Technology of Managing World Crises." *Milennial Magazine*, December 17, 2015. http://millennialmagazine.com/crisis-mapping-technologically-managing-world-crises/ (accessed August 12, 2016).

Rato, Daniel. "Radio Vallekas: 30 años agitación, periodismo y trabajo barrio." *Diagonal*, March 20, 2016. https://www.diagonalperiodico.net/global/29445-radio-vallekas-30-anos-agitacion-periodismo-e-integracion.html (accessed August 12, 2016).

Redaccion El Universal, Web site. "Asi hacemos periodismo de datos en El Universal," January 12, 2016. http://www.eluniversal.com.mx/articulo/periodismo-de-datos/2016/01/12/asi-hacemos-periodismo-de-datos-en-el-universal (accessed August 12, 2016).

ReliefWeb, Web site. "ReliefWeb," 2015. http://reliefweb.int/organizations (accessed December 11, 2015).

Reuters Institute. "Digital News Report 2015." *Reuters Institute*, 2016. http://www.lisboncouncil.net/component/downloads/?id=1208 (accessed August 12, 2016).

Royal Geographical Society. "Clark's Sector Model," 2015. https://www.rgs.org/NR/rdonlyres/53FAACFE-326B-4975-AC42-38D52792DB16/0/ClarkFisherModelPresentation.ppt (accessed August 12, 2016).

Rheingold, Howard. *Smart Mobs: The Next Social Revolution.* Cambridge: Perseus Publishing, 2002.

—. "Politics and Social Change." In *Technologies of Cooperation*, edited by Maureen Davis. Palo Alto, California: Institute for the Future, 2005.

Rifkin, Jeremy. *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism.* New York: Palgrave Macmillan Trade, 2014.

Rigobon, Roberto, interview by Thiago Ferrer Morini. "El tener mas datos no quiere decir que haya mejor informacion." *El Pais*, October 26, 2014. http://economia.elpais.com/economia/2014/10/24/actualidad/1414172504_719698.html (accessed July 21, 2016).

Ritzer, George. *Encyclopedia of social theory.* Thousand Oaks, CA: Sage Publications, 2005.

Robert Wood Johnson Foundation, Web site. "Semi-structured Interviews," Qualitative Research Guidelines Project, 2008. http://www.qualres.org/HomeSemi-3629.html (accessed July 21, 2016).

Roderick, Rick. "Habermas and the Fragile Dignity of Humanity." In *The Self Under Siege: Philosophy in the Twentieth Century*, Video Lecture 5, 2012 [1993]. https://www.youtube.com/watch?v=itGtf3ZSkyQ (accessed July 21, 2016).

Rodríguez, Clemencia. *Fissures in the mediascape - An international study of citizens' media.* Cresskill: Hampton Press Inc., 2001.

—. Interview by Miren Gutiérrez. "Special Report-Media:Alternative and Influential?" *Inter Press Service*, October 29, 2004. http://www.ipsnews.net/2004/10/special-report-media-alternative-and-influential/ (accessed July 21, 2016).

—. *De medios alternativos a medios ciudadanos: Trayectoria teorica de un termino.* Oklahoma: The University of Antioquia, 2009.

Rodríguez-Giralt, Israel. "Social Movements as Actor-Networks: Prospects for a Symmetrical Approach to Doñana's Environmentalist Protests." *Convergencia* (May-August 2011): 13-35. http://www.scielo.org.mx/pdf/conver/v18n56/v18n56a1.pdf (accessed July 21, 2016).

Rogers, Simon. "Anyone can do it. Data journalism is the new punk." *The Guardian*, May 24, 2012. https://www.theguardian.com/news/datablog/2012/may/24/data-journalism-punk (accessed February 14, 2016).

—. "John Snow's data journalism: the cholera map that changed the world." *The Guardian*, March 15, 2013. http://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map (accessed February 14, 2016).

Rojas, Fabio. "Big Data and Social Movement Research." *Mobilizing Ideas*, April 2, 2015. https://mobilizingideas.wordpress.com/2015/04/02/big-data-and-social-movement-research/ (accessed February 14, 2016).

Rosen, Jay. "Web Users Open the Gates." *Washington Post*, June 19, 2006a. http://www.washingtonpost.com/wp-dyn/content/article/2006/06/18/AR2006061800618.html (accessed March 6, 2016).

—. "The People Formerly Known as the Audience." *PressThink*, June 27, 2006b. http://archive.pressthink.org/2006/06/27/ppl_frmr.html (accessed March 6, 2016).

Rosling, Hans. *The magic washing machine. TED Talks*, March 2011. http://www.ted.com/talks/hans_rosling_and_the_magic_washing_machine (accessed March 6, 2016).

Rossetto, Louis. "Rebuttal of the 'Californian Ideology.'" *Alamut*, 1998. http://www.alamut.com/subj/ideologies/pessimism/califIdeo_II.html

Rousseau, Jean-Jacques. *The Social Contract.* Oxford: Early Modern Texts, 2010 [1751].

Rowley, Jennifer and Farrow, John. *Organising Knowledge: An Introduction to Managing Access to information.* Guildford: Ashgate, 2008.

Russell, Andrew L. '*Rough Consensus and Running Code*' *and the Internet-OSI Standards War*. IEEE Computer Society, 2006.

Russom, Philip. "Managing Big Data - TDWI Best Practices Report." *TDWI research* (Fourth Quarter 2013): 1-40. https://www.pentaho.com/sites/default/files/uploads/resources/tdwi_best_practices_report_-_managing_big_data.pdf (accessed April 3, 2016).

Rutter, Tamsin. "How big data is transforming public services – expert views." *The Guardian*, April 17, 2014. https://www.theguardian.com/public-leaders-network/2014/apr/17/big-data-government-public-services-expert-views (accessed April 3, 2016).

Ryan, Sara, and Runswick Cole, Katherine. "From Advocate to Activist? Mapping the Experiences of Mothers of Children on the Autism Spectrum." *Journal of Applied Research in Intellectual Disabilities* 22, no. 1 (January, 2009): 43–53. http://onlinelibrary.wiley.com/doi/10.1111/j.1468-3148.2008.00438.x/abstract (accessed April 3, 2016).

Sampedro, Víctor. *El cuarto poder en red. Por un periodismo (de código) libre.* Barcelona: Icaria, 2014.

Sandoval, Marisol. *From Corporate to Social Media: Critical Perspectives on Corporate Social Responsibility in Media and Communication Industries*. Abingdon: Routledge, 2014.

Sassower, Raphael. "Radical Public Intellectuals." *Social Epistemology Review and Reply Collective* 4, no. 1 (December 30, 2014): 57-63. https://socialepistemologydotcom.files.wordpress.com/2014/12/sassower_radical_pi.pdf (accessed April 3, 2016).

Savio, Roberto, interview by Miren Gutiérrez. "Everybody Leaves the Forum Happier, Wiser and Stronger." *Inter Press Service*, January 8, 2008. http://www.ipsnews.net/2008/01/qa-39everybody-leaves-the-forum-happier-wiser-and-stronger39/ (accessed April 3, 2016).

Schmidt, Ben. "Reading digital sources: a case study in ship's logs." *Sapping Attention*, November 15, 2012. http://sappingattention.blogspot.com.es/2012/11/reading-digital-sources-case-study-in.html (accessed April 3, 2016).

Schmidt, Eric, and Cohen, Jared. *The new digital age: reshaping the future of people, nations and business.* New York: Knopf, 2013.

School of Data, Web site. "Who we are," 2016. http://schoolofdata.org/team/ (accessed April 3, 2016).

Schradie, Jen. "5 reasons why online Big Data is Bad Data for researching social movements." *Mobilizing Idideas*, April 2, 2015. https://mobilizingideas.wordpress.com/2015/04/02/5-reasons-why-online-big-data-is-bad-data-for-researching-social-movements/ (accessed April 3, 2016).

Schroeder, John. "MapR CEO Offers His Predictions for Big Data Drivers in 2015." *MapR*, 2015. https://www.mapr.com/company/press-releases/mapr-ceo-offers-his-predictions-big-data-drivers-2015

Scott, Charles Prestwich. "A Hundred Years." *The Guardian*, November 29, 2002 [1951]. http://www.theguardian.com/commentisfree/2002/nov/29/1 (accessed February 15, 2016).

Sen, Amartya. "How Is India Doing?" *The New York Review of Books*, December 16, 1982. http://www.nybooks.com/articles/1982/12/16/how-is-india-doing/ (accessed February 15, 2016).

—. *Development as Freedom.* New York: Alfred A. Knopf, Inc., 2000.

Stanford Encyclopedia of Philosophy. "The Value of Knowledge," August 21, 2007. https://plato.stanford.edu/entries/knowledge-value/ (accessed February 15, 2016).

Sharp, Adam. "A new barometer for the election." *Twitter*, August 12, 2012. https://blog.twitter.com/2012/a-new-barometer-for-the-election (accessed February 8, 2015).

Shaw, Jonathan. "Why 'Big Data' Is a Big Deal: Information science promises to change the world." *Harvard Magazine*, March-April 2014. http://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal (accessed February 15, 2016).

Shearman, Sarah. "Data 'crucial' to eradicating poverty." *The Guardian.* September 28, 2015. https://www.theguardian.com/media-network/2015/sep/28/data-poverty-sustainable-development-goals-un (accessed February 15, 2016).

Shirky, Clay. *Here Comes Everybody: The power of organising without organisations.* New York: Penguin Group USA, Inc., 2008.

Singh, Arvind. "Is Big Data the New Black Gold?" *Wired*, February 14, 2013. https://www.wired.com/insights/2013/02/is-big-data-the-new-black-gold/ (accessed February 15, 2016).

Skoll World Forum, Web site. "Up for Debate: How Can Big Data Have a Social Impact?" 2015. http://skollworldforum.org/2013/03/19/big-data-as-societys-watchdog/?series=how-can-big-data-have-a-social-impact (accessed May 26, 2015).

Slater, Dirk. "What I've learned about mapping violence." *Ushahidi Blog*, March 18, 2014. https://www.ushahidi.com/blog/2014/03/18/what-ive-learned-about-mapping-violence (accessed August 14, 2016).

Small, Annika, and Anderton, Ed. "Believe the hype: Big data can have a big social impact." *The Guardian*, May 6, 2014. https://www.theguardian.com/social-enterprise-network/2014/may/06/data-open-social-impact-tips-tools (accessed August 14, 2016).

Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of Nations.* Feedbooks.com, 2007 [1776]. http://www.ifaarchive.com/pdf/smith_-_an_inquiry_into_the_nature_and_causes_of_the_wealth_of_nations%5B1%5D.pdf (accessed August 14, 2016).

Smith, Peter. "Cellphone and Internet access helps – and hinders – accurate reporting in Kenya." *The Christian Science Monitor*, January 31, 2008. http://www.csmonitor.com/World/terrorism-security/2008/0131/p99s01-duts.html (accessed August 14, 2016).

SocialBrite, Web site. “Social Advocacy Toolkit,” 2016. http://www.socialbrite.org/advocacy-toolkit/ (accessed April 6, 2016).

Spotfire. “Big Data Analytics: Shaping the Future of Public Service.” *TIBCO Spotfire's Trends and Outliers Blog*, November 26, 2013. http://www.tibco.com/blog/2013/11/26/analytics-shaping-the-future-of-public-service/ (accessed April 6, 2016).

Stallman, Richard M. *Free Software, Free Society: Selected Essays of Richard M. Stallman*. Boston: Free Software Foundation, Inc., 2010.

Standby Task Force, Web site. “About us,” 2016. http://www.standbytaskforce.org/about-us/ (accessed April 20, 2016).

Stanley, Matthew. “Where Is That Moon, Anyway? The Problem of Interpreting Historical Solar.” In *‘Raw Data’ Is an Oxymoron*, edited by Lisa Gitelman, 77-90. Cambridge, Massachusetts: The MIT Press, 2013.

Stencel, Mark, Adair, Bill, and Kamalakanthan, Prashanth. “Why digital tools are missing in most newsrooms.” *The Goat Must Be Fed*, May 2015. http://www.goatmustbefed.com/resources/pdf/goat-must-be-fed.pdf (accessed February 19, 2016).

Stiglitz, Joseph. “Of the 1%, by the 1%, for the 1%.” *Vanity Fair*, April 2011. http://www.vanityfair.com/news/2011/05/top-one-percent-201105 (accessed February 19, 2016).

—. Globalization and Its Discontents. New York and London: W. W. Norton Company, 2002.

Stokes, Bruce, Wike, Richard, and Carle, Jill. “Global Concern about Climate Change, Broad Support for Limiting Emissions.” *Pew Research Center*,November 5, 2015. http://www.pewglobal.org/2015/11/05/global-concern-about-climate-change-broad-support-for-limiting-emissions/ (accessed February 19, 2016).

Stokes, Kathleen, Baker, Richard, and Pigott, Rosie. “Which doctors take up promising ideas? New insights from open data.” *Nesta* (January, 2014): 1-66. https://www.nesta.org.uk/sites/default/files/which_doctors_take_up_promising.pdf (accessed February 19, 2016).

Stone, Martha L. “Big data for media.” *Reuters Institute for the Studyof Journalism* (November 2014): 1-35. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Big%20Data%20For%20Media_0.pdf (accessed February 19, 2016).

Strategy Analytics, Web site. “33 Billion Internet Devices By 2020: Four Connected Devices For Every Person In World,” October 14, 2014.

http://www.strategyanalytics.com/default.aspx?mod=pressreleaseviewer&a0=5609 (accessed April 10, 2015).

Streeter, Thomas. *The Net Effect: Romanticism, Capitalism, and the Internet.* New York: NYU Press, 2010.

Strohmeyer, Robert. "The 7 Worst Tech Predictions of All Time." *TechHive*, December 31, 2008. http://www.pcworld.com/article/155984/worst_tech_predictions.html (accessed April 10, 2015).

Suárez, Eduardo, and Ramírez, María. "13 lessons from ISOJ to innovate journalism according to the blog #nohacefaltapapel." *Journalism in the Americas Blog*, April 28, 2015. https://knightcenter.utexas.edu/blog/00-16068-13-lessons-isoj-innovate-journalism-according-blog-nohacefaltapapel (accessed April 30, 2015).

Suárez, Sandra L. "Mobile Democracy: Text Messages, Voter Turnout, and the 2004 Spanish General Election." Annual Meeting of the American Political Science Association, Temple University, Philadelphia, September 1-4, 2005. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1620062 (accessed April 30, 2015).

Sutter, John D. "'Where's Waldo'... but for science." *CNN*, September 30, 2014: 4. http://edition.cnn.com/2014/09/30/opinion/sutter-wildlife-crowdsource/

TacticalTech, Web site. "Tactical Technology Collective," 2009. https://informationactivism.org/en (accessed February 24, 2016).

Tavernise, Sabina, and Lehren, Andrew W. "A Grim Portrait of Civilian Deaths in Iraq." *The New York Times,* October 22, 2010. http://www.nytimes.com/2010/10/23/world/middleeast/23casualties.html?_r=0 (accessed February 24, 2016).

TED, Web site. "About Juliana," 2010. http://www.ted.com/profiles/9417/fellow (accessed August 5, 2016).

Tennison, Jeni. "Open data is a public good. It should not be confused with data sharing." *The Guardian*, May 12, 2014. https://www.theguardian.com/commentisfree/2014/may/12/response-confuse-open-data-sharing-government

Terp, Sara. "Open up as much social-good data as possible." *Opensource.com*, February 4, 2015 . https://opensource.com/business/15/2/open-crowdsourced-data (accessed August 23, 2016).

Terry, Paul R. "Habermas and Education: knowledge, communication, discourse." *Curriculum Studies* 5, no. 3 (1997): 269-269. http://inelmen.boun.edu.tr/amr/erol03/philo03/habermas.pdf

Tetuan, Web site. "Asambla Popular Tetuan," July 27, 2011. http://tetuan.tomalosbarrios.net/ / (accessed February 14, 2016).

The Economist. "Special report: Managing information - Data, data everywhere." February 25, 2010. http://www.economist.com/node/15557443 (accessed January 16, 2016).

—. "Elinor Ostrom, defender of the commons, died on June 12, aged 78." June 30, 2012. http://www.economist.com/node/21557717 (accessed January 16, 2016).

—. "The new politics of the Internet: Everything is connected" January 5, 2013. http://www.economist.com/news/briefing/21569041-can-internet-activism-turn-real-political-movement-everything-connected (accessed January 16, 2016).

—. "Smartphones: Planet of the phones." February 28, 2015a. http://www.economist.com/news/leaders/21645180-smartphone-ubiquitous-addictive-and-transformative-planet-phones (accessed January 16, 2016).

—. "Telocoms and society: The truly personal computer." February 28, 2015b. http://www.economist.com/news/briefing/21645131-smartphone-defining-technology-age-truly-personal-computer (accessed January 16, 2016).

The International Data Responsibility Group, Web site. "About the IDRG," 2016. http://www.responsible-data.org/ (accessed May 5, 2016).

The New York Times. "From The Editors; The Times and Iraq." *The New York Times*, May 26, 2004. http://www.nytimes.com/2004/05/26/world/from-the-editors-the-times-and-iraq.html (accessed May 5, 2016).

The Poverty Site, Web site. "Relative poverty, absolute poverty and social exclusion," 2012. http://www.poverty.org.uk/summary/social%20exclusion.shtml (accessed May 5, 2016).

Thomas, Gary. *How to Do Your Case Study: A Guide for Students and Researchers.* Thousand Oaks: Sage Publications, 2011.

Timms, Matt. "We must end the age of exploration subsidies." *World Finance*, March 11, 2016. http://www.worldfinance.com/markets/we-must-end-the-age-of-exploration-subsidies (accessed May 5, 2016).

Todo Sobre la Ciencia. "ADN Basura – ¿Refuta el diseño?" 2016. http://www.allaboutscience.org/spanish/adn-basura.htm (accessed March 19, 2016).

Toffler, Alvin. *The Third Wave.* New York: Bantam, 1984.

Tomnod, Web site. “About,” 2016. http://www.tomnod.com/ (accessed April 20, 2016).

Touraine, Alain. *The Post-Industrial Society: Tomorrow's social history: classes, conflicts and culture in the programmed society.* New York: Random House, 1971.

—. *The Self-Production of Society.* Chicago: University of Chicago Press, 1977.

—. *A new paradigm for understanding today's world.* Cambridge and Malden: Polity Press, 2007.

Trottier, Daniel. “!Big Data Ambivalence: Visions and Risks in Practice.” In *Big Data? Qualitative Approaches to Digital Research (Studies in Qualitative Methodology,* edited by Martin Hand and Sam Hillyard, 51-72. Bingley: Emerald Group Publishing Limited, 2014.

Tu Derecho a Saber, Web site. “La negativa las instituciones a responder por correo electrónico condena al cierre a la web Tu Derecho a Saber,” December 10, 2015. http://www.civio.es/2015/12/cierre-tu-derecho-a-saber/ (accessed May 5, 2016).

Tufekci, Zeynep. “Engineering the public: Internet, surveillance and computational politics.” *Fisrt Monday* 18, no. 5 (May, 2012). http://firstmonday.org/article/view/4901/4097 (accessed May 5, 2016).

Turkle, Sherry. *Alone Together: Why We Expect More From Technology and Less From Each Other.* New York: Basic Books, 2010.

Turner, Jonathan H, ed. *Handbook of Sociological Theory.* Riverside, California: Springer, 2006.

Twitter Web site. “Twitter Privacy Policy,” 2016. https://twitter.com/privacy (accessed May 23, 2016).

Tysiachniouk, Maria S. “Transnational Governance Through Private Authority: The Case of Forest Stewardship Council Certification in Russia.” Dissertation - Wageningen University, December 11, 2012. http://edepot.wur.nl/240492 (accessed May 23, 2016).

UAViators, Web site. “Welcome to UAViators,” 2016. http://uaviators.org/ (accessed May 25, 2016).

Ulanoff, Lance. “Amazon knows what you want before you buy it.” *Mashable*, January 27, 2014. http://mashable.com/2014/01/21/amazon-anticipatory-shipping-patent/ (accessed May 25, 2016).

UN News Centre. “Deputy UN chief calls for urgent action to tackle global sanitation crisis,” March 21, 2013.

http://www.un.org/apps/news/story.asp?NewsID=44452#.WHzyl1PhDcs (accessed May 25, 2016).

UNDESA, Web site. "Securing reliable data for development," March 4, 2014. http://www.un.org/en/development/desa/news/statistics/unstats2014.html (accessed January 30, 2015).

Ushahidi, Web site. "Implementation Toolbox," 2011. http://www.slideshare.net/Ushahidi/ushahidi-deployment-implementation-toolbox (accessed Ausgust 17, 2016).

—. "About us," 2016a. https://www.ushahidi.com/about (accessed August 6, 2016).

—. "Ushahidi announces USD 1.5m grant from Development Innovation Ventures." *Ushahidi Blog*, April 15, 2016b. https://www.ushahidi.com/blog/2016/04/15/ushahidi-announces-usd-1-5m-grant-from-development-innovation-ventures (accessed August 6, 2016).

—. "Ushahidi's new maps, timeline bring clarity to data." *Ushahidi Blog*, June 14, 2016c. https://www.ushahidi.com/blog/2016/06/14/new-maps-timeline-bring-clarity-to-data (accessed August 6, 2016).

van de Donk, Wim, Brian D. Loader, Paul G. Nixon and Rucht, Dieter, eds. *Cyberprotest: New Media, Citizens and Social Movements.* London and New York: Routledge, 2004.

van der Velden, Lonneke. "Forensic devices for activism: Metadata tracking and public proof." *Big Data & Society (Sage)* (July–December 2015): 1–14. http://journals.sagepub.com/doi/pdf/10.1177/2053951715612823 (accessed August 5, 2016).

van Dijk, Jan A. G. M. *The Network Society: Social Aspects of New Media.* London: Sage Publications, 2005.

Vaterlaus, J. Mitch, and Higginbotham, Brian J. "Qualitative Program Evaluation Methods." *The Forum - NC State University*, 2011. https://ncsu.edu/ffci/publications/2011/v16-n1-2011-spring/vaterlaus-higginbotham.php (accessed August 5, 2016).

Vegh, Sandor. "Classifying Forms of Online Activism." In *Cyberactivism: Online Activism in Theory and Practice*, edited by Martha McCaughey and Michael D. Ayers, 71-96. New York and London: Routledge, 2003.

Verity, Andrej. "The [unexpected] Impact of the Libya Crisis Map and the Standby Volunteer Task Force." *Ushahidi Blog*, December 19, 2011. http://www.standbytaskforce.org/2011/12/19/sbtf-libya-impact/ (accessed August 5, 2016).

Vesset, Dan et al. "Worldwide Big Data Technology and Services: 2012 – 2016 Forecast." *IDC Analyze the Future* (December 2012): 1-33. http://laser.inf.ethz.ch/2013/material/breitman/additional%20reading/Worldwide%20Big%20Data%20Technology%20and%20Services%202012-2016%20Forecast.pdf (accessed July 18, 2016).

Vidascontadas, Web site. "Vidas contadas," 2013. http://vidascontadas.org/ (accessed July 18, 2016).

Vizzuality, Web site. "Beautiful data design for a better world," 2016. http://www.vizzuality.com/ (accessed July 18, 2016).

Vota, Wayan. Dead Ushahidi: A Stark Reminder for Sustainability Planning in ICT4D. *ICT Works*, July 9, 2012. http://www.ictworks.org/2012/07/09/dead-ushahidi-stark-reminder-sustainability-planning-ict4d/ (accessed May 23, 2016).

Waddell, Steve, and Khagram, Sanjeev. "Multi-stakeholder global networks: emerging systems for the global common good." En *Partnerships, Governance and Sustainable Development: Reflections on Theory and Practice*, edited by P. Glasbergen, F. Biermann and A. P. J. Mol, 261-287. Cheltenham Glos: Edward Elgar Publishing, 2007.

Wagner, Kurt. "WhatsApp Keeps on Growing, Hits 700 Million Users." *Recode*, January 6, 2015. http://recode.net/2015/01/06/whatsapp-keeps-on-growing-hits-700-million-users/ (accessed April 10, 2015).

Walker, Marína, interview by Carmen Aristegui. http://aristeguinoticias.com/0504/entrevistas/no-somos-arma-de-ningun-gobierno-Marína-walker-del-icij-en-cnn/ (accessed April 10, 2015).

Watin-Augouard, Jean. "De la régulation collective à l'arbitrage personnel." *Prodimarques*, no. 64 (October, 2008). http://www.prodimarques.com/documents/gratuit/64/de-la-regulation-collective-a-larbitrage.php (accessed February 23, 2016).

Weber, Max. *Essays in Sociology*. New York: Oxford University Press, 1946 [1921].

—. *The Protestant Ethic and the Spirit of Capitalism*. London and New York: Routledge, 2005 [1930].

World Bank, Web site. "Defining Civil Society," July 22, 2013. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/CSO/0,,contentMDK:20101499~menuPK:244752~pagePK:220503~piPK:220476~theSitePK:228717,00.html (accessed February 23, 2016).

Webster, Frank. *Theories of the Information Society*. London and New York: Routledge, 2006.

Were, Daudi. "Using technology effectively to fight corruption." *Ushahidi Blog*, December 10, 2015. https://www.ushahidi.com/blog/2015/12/10/using-technology-effectively-to-fight-corruption (accessed August 6, 2016).

—. "Ushahidi: Accelerating the Journey to Scale." *Ushahidi Blog*, August 3, 2016. https://www.ushahidi.com/blog/2016/08/03/ushahidi-accelerating-the-journey-to-scale (accessed August 6, 2016).

West, Mollie, and Posner, Andy. "The Math of Social Change." *Stanford Social Innovation Review (SSIR)* (November 9, 2012). https://ssir.org/articles/entry/the_math_of_social_change (accessed August 6, 2016).

Wikipedia, Web site. "Panopticon," 2016. https://en.wikipedia.org/wiki/Panopticon#/media/File:Panopticon.jpg (accessed February 2, 2016).

Williams, Raymond. *Television: Technology and Cultural Form.* London: Routledge, 1990.

—. *Culture and materialism: selected essays.* London and New York: Verso, 2005.

Wilson, T.D. "Sociological aspects of information science." *International Forum on Information and Documentation* 6, no. 2 (1981): 13-18. http://www.informationr.net/tdw/publ/papers/socasp81.html (accessed January 2, 2016).

Wollstonecraft, Mary. *A Vindication of the Rights of Men A Vindication of the Rights of Woman An Historical and Moral View of the French Revolution.* Oxford: Oxford University Press, USA, 1993.

Woodside, Arch. *Case Study Research: Theory, Methods and Practice*. Bradford: Emerald Group Pub, 2010.

World Bank, Web site. "FAQs: Global Poverty Line Update," September 30, 2015. http://www.worldbank.org/en/topic/poverty/brief/global-poverty-line-faq (accessed April 25, 2016).

Worldmapper, Web site. "On the grid: Worldmapper and beyond," February 25, 2013. http://www.viewsoftheworld.net/?p=3328 (accessed April 25, 2016).

Yin, Robert K. *Case Study Research: Design and Method.* Thousand Oaks: Sage Publications - Applied Social Research Methods Series, 2002.

Zeitzoff, Thomas. "The Micro-dynamics of Reciprocity in an Asymmetric Conflict: Hamas, Israel, and the 2008-2009 Gaza Conflict." (November 2009): 1-60. https://earlywarning.files.wordpress.com/2009/11/zeitzoff_gaza_11_2_09.pdf accessed May 3, 2016).

Zeleny, Milan. "Integrated Knowledge Management." *Human Systems Management* 7, no. 1 (1987): 59-70. http://milanzeleny.com/Files/Content/Knowledge_X.pdf (accessed May 3, 2016).

Zer-Aviv, Mushon. "DataViz—The UnEmpathetic Art." *Responsible Data Forum*, January 2016. https://responsibledata.io/dataviz-the-unempathetic-art/ (accessed May 3, 2016).

Zins, Chaim. "Conceptual Approaches for Defining Data, Information, and Knowledge." *Journal Of The American Society For Information Science And Technology* 58, no. 4 (February, 2007): 479-493. http://www.success.co.il/is/zins_definitions_dik.pdf (accessed May 3, 2016).

Zittrain, Jonathan. *The Future of the Internet And How to Stop It.* New Haven and London: Yale University Press, 2008.

Annex I: Semi-structured and in-depth interviews

The guide employed in the semi-structured interviews includes these basic questions, which were adapted to each interviewee, as follows:

- Take me back through the history in your career that brought you to XXX. What types of professional developments have you previously experienced? What are your background and strengths?
- Can you describe some details of what you have gained from this work? Is this experience similar to your expectations?
- What gap did XXX fill? What have been its biggest successes? Why do you think it succeeded? How about growth potential?
- What is the role of data in XXX? Big data infrastructures have expanded to the non-experts in recent years, have data changed anything in the way you think or work? Do you think data activism exists as an independent phenomenon? How would you define it? Do you think data could have the ability to emancipate people? If so, how?

For the case study, I employ a different set of questions around the Ushahidi platform, which include:

- Is Ushahidi a civil society organisation? How would you define it?
- Does Ushahidi do any advocacy work?
- How many of the organisations you support are civil society organisations?

- Does Ushahidi have a theory of change, that is, a mission and a vision about how the world should change to make it better, what the problem is and what it is proposing to improve things? If so, can you articulate it?
- Does Ushahidi promote activism? If so, how?
- Does Ushahidi have an explicit theory of change? If so, could you formulate it?
- What is the relationship between Ushahidi and the deployments that use it as a platform?
- What is 'activist mapping' and 'crisis mapping'? Is that what Ushahidi does?
- What is the role of participation and collaboration internally at Ushahidi?
- How many of you there are in the core team?
- Does Ushahidi aim to change practice or and policy in terms of how crises are tackled?
- Has is achieved any changes in practice or policy? Which ones?
- By innovating, is Ushahidi a systemic change agent in terms?
- Does it learn from mistakes? If so, can you illustrate with an example?
- Do Ushahidi-based projects involve an empowering process? How?
- What about empowerment at an individual level?
- What about information getting published having negative impacts on victims?

The interview guide includes, too, a clear set of instructions per each question, like this one: 'Probe to see if the interviewee mentions the introduction or adoption of data infrastructures as a defining factor in success' or 'Probe to see if the interviewee describes him or herself as data activist.'

The approximate length of the interview is from thirty to a hundred minutes, around the above mentioned major questions, which are slightly adapted to each interviewee depending on her or his background and where the conversation leads, but remain basically the same so comparisons can be made and data can extracted from them.

Table 7: List of semi-structured interviews

Name	Position	Date and place	Means
Aguilar, Luis Hernando	Knowledge Management and Digital Information Specialist, Nansen Centre for Peace and Dialogue (formerly at OCHA Colombia)	August 15, 2016, Bogota	Skype
Almiraat, Hisham	Medical doctor, community leader at Global Voices, frontrunner of the Arab Spring in Morocco	September 22, 2016, Amsterdam	Email and face to face
Álvarez Leiva, Sergio	Chief Product Officer, CARTO (CartoDB)	March 9, 2016, New York	Email and face to face
Beltrán, Esteban	Executive Director, Amnesty International	June 6, 2016, Madrid	Skype and face to face
Blat, Fernando	Director and co-founder, Populate Tools	July 13, 2016, Madrid	Email
Cabo, David	Director, Civio	May 10, 2016, Madrid	Email
Cabra, Mar	Coordinator Panama Papers, International Consortium of Investigative Journalists (ICIJ)	May 27, 2016, Madrid	Telephone and face to face
Cairo, Alberto	Data visualisation expert and author, Knight Chair in Visual Journalism, School of Communication, Miami	September 23, 2014, Miami (US)	Telephone and face to face
Carrión, Jordi	Journalist, author	July 12, 2016, Barcelona	Email and face to face
Clark, Duncan	Data Journalist, co-founder, Kiln	August 22, 2014, London	Face to face
Congosto, Mariluz	Data analyst, Researcher of the University of Carlos III	March 21, 2016, Madrid	Email and face to face
Cucianelli, Sandra	Data journalist, author, trainer of Knight Centre's MOOC, Knight International *Journalism* Fellow	June 21, 2016, Bahía Blanca (Argentina)	Skype
Downing, John	Founding Director of the Global	April 14, 2016,	Skype

	Media Research Centre, Professor Emeritus of International Communication at the College of Mass Communication & Media Arts, Southern Illinois University, and North-western University, Qatar	Qatar	
Faleiros, Gustavo	Head of InfoAmazonia	September 13, 2016, Rio de Janeiro	Skype
Halimanjaya, Aidy	Data analyst, and Climate Funds Update researcher	September 16, 2014, London and Bandung (Indonesia)	Email and face to face
Haque, Usman	Founder of Umbellium	March 14, 2016, London	Skype
Hogan, Mél	Assistant Professor of Communication, Illinois Institute of Technology	May 25, 2016, Illinois (US)	Email
Innerarity, Daniel	Author, researcher and Professor of Political and Social Philosophy,	September 29, 2014, Madrid	Email and face to face
Mitchell, Sebastian	Solutions Team Manager, Ushahidi	May 5, 2016, Granada	Telephone
Marín Miró, Oscar	Data Researcher and founder, Outliers Collective	August 6, 2016, San Sebastian	Email and face to face
Mills, Craig	Vizzuality, CEO	September 2, 2016, Madrid	Skype
Oduor, Angela	Director of Community Engagement, Ushahidi	April 1, 2016, Nairobi	Skype
Onigbinde, Oluseun	Public budget expert, strategist, co-founder, BudgIT and Knight International Innovation Fellow	October 2, 2014, Abuja	Email
O'Reilly, Matthew	Researcher and expert in open data	September 30, 2014, London	Email
Prest, Emma	General Manager at DataKind UK	February 24, 2016	Skype
Rank, Rachel	Director, Publish What You Fund	September 10, 2014, London	Telephone
Sampedro, Victor	Data journalist, author and director of the Communication, Culture and Digital Citizenship Master Programme, *King* Juan Carlos *University*	January 25, 2016, Madrid	Skype and face to face
Schwochow, Jan	Chief Executive Officer, Golden Section Graphics	February 25, 2016, Berlin	Email
Tulp, Jan Willem	data experience designer, TULP Interactive	June 17, 2016, The Hague	Email
Velasco, Cesar	epidemiologist and medical	May 5, 2016,	Skype and

	assistant manager of the Hospital Clínico Universitario	Zaragoza (Spain)	face to face
Were, Daudi	Executive Director, Ushahidi	May 11, 2016, Nairobi	Skype
Whipkey, Katie	Researcher and author of *Guidance for incorporating big data into humanitarian operations*, MSPPM '16 Carnegie Mellon University Heinz College	June 20, 2016, Pittsburgh (US)	Email
Whitley, Shelagh	Researcher and Team Leader of the Green Growth Programme at the Overseas Development Institute (ODI), and coleader in the fossil fuel subsidies advocacy efforts	April 15, 2016, London	Skype

Source: Elaboration by the author.

Annex II: Examples cyber activism and mobilisation with mobile phones

Examples of cyber activism:

1998: The Mexican EZLN (Zapatista Army of National Liberation or *Ejército Zapatista de Liberación Nacional*) used communications technologies, such as cell phones and the internet, to contact activists around the globe and to create the anti-globalisation network Peoples Global Action (PGA) to protest the World Trade Organisation (WTO) in Geneva. The PGA sustained the campaign and called for 'global days of action' in other anti-globalisation initiatives afterwards (Peoples Global Action 1998; Burbach 2001, 138).

1998: MoveOn.org[427] was created as an email group in order to pass around a petition to the US Congress to 'censure President Clinton and move on.' The petition gathered half a million signatures, but initiative continued similar campaigns.

1999: A worldwide network of online activist sites, gathered under an umbrella organisation called Indymedia, was created 'for the purpose of providing grassroots coverage of the WTO protests in Seattle' that year (Indymedia 2016). Indymedia –the Independent Media Centre— describes itself as 'a network of collectively run media outlets for the creation of radical, accurate and passionate *tellings* of the truth' (ibid.). In Seattle, Indymedia 'produced its own newspaper, distributed throughout Seattle and to other cities via the internet, as well as

[427] See front.moveon.org/ [accessed on April 18, 2016].

hundreds of audio segments, transmitted through the web and Studio X, a 24-hour micro and internet radio station based in Seattle' (ibid.).

2007: Avaaz[428] is cofounded by MoveOn, Res Publica[429] and Service Employees International Union, with a similar purpose but an international focus.

2011: Egyptian ruler Hosni Mubarak was forced to step down under pressure from a popular revolution, which made heavy use of cyber activist tactics. 'This revolution was characterised by the instrumental use of social media, especially Facebook, Twitter, YouTube, and text messaging by protesters, to bring about political change and democratic transformation' (Khamis and Vaughn 2011).

2012: *Kony 2012*,[430] a short documentary film released as part of the 'Stop Kony' campaign to force Joseph Kony –an African leader of the Lord's Resistance Army, a guerrilla group that initially operated in Uganda, and International Criminal Court fugitive— widely known in order to have him arrested by the end of 2012, when the campaign ended. The film disseminated virally, to the point that millions watched it globally. The campaign resulted in a resolution by the U.S. Senate and contributed to the decision to send troops by the African Union.

2014: A hashtag activist campaign was initiated (#BringBackOurGirls) after 276 schoolgirls were kidnapped in April from the Government Secondary School in Chibok, Nigeria. Most of them were never seen again (a few dozen managed to escape over the next few months). The responsibility was claimed by Boko Haram, a terrorist organisation. The campaign is still alive.

---

[428] See secure.avaaz.org/en/ [accessed on April16, 2016].
[429] See respublica.org.uk/ [accessed on April16, 2016].
[430] See invisiblechildren.com/kony-2012/ [accessed on April16, 2016].

Examples of social mobilisation using mobile technologies:

1999: ‘The use of mobile phones during the riots at the 1999 World Trade Organisation (WTO) meeting in Seattle starkly illustrates the potential for using mobile devices to coordinate political action’ (Harkin 2003, 61).

2001: Millions Filipino citizens mobilised using SMS and the internet to force Joseph Estrada out of presidential office. Into the second year of his six-year term in 2000, Estrada was inundated by reports of corruption, inexplicable affluence and participation in illegal gambling syndicates.

2004: The Spanish general election occurred in the wake of an unprecedented terrorist attack. The ruling Popular Party (PP) was due to win, according to most surveys. But its orchestrated reaction to the tragedy –trying to blame it on the Spanish terrorist group ETA (which would have reinforced its anti-ETA stance) — produced an unprecedented public reaction that turned the tables in favour of the main opposition party, the socialist PSOE. ‘Mobile phones have the potential of fostering political mobilisation. Like the internet, mobile phones facilitate communication and rapid access to information. Compared to the internet, however, mobile phone diffusion has reached a larger proportion of the population in most countries, and thus the impact of this new medium is conceivably greater’ says Suárez. The elections outcome in Spain (against all odds, Jose Luis Rodríguez Zapatero was elected as Prime Minister) ‘reflects the potential mobile phones have to provide the user with independent information and bring about voter mobilisation’ (2005, 2).

2011: The ‘Indignad@s Movement,’ also in Spain, is also another example of massive mobilisation facilitated by smartphones. Monterde and Postill conclude that smartphones were key as ‘new articulators of online spaces and occupied physical spaces, especially via Twitter

and live streaming' (2013, 1). The 15-M Movement or Indignad@s Movement entails a series of massive demonstrations in Spain, whose origin can be traced to social networks such as Real Democracy NOW (*Democracia Real YA*) or Youth Without a Future (*Juventud Sin Futuro*), among other civilian digital platforms and 200 other small associations. The demonstrations mobilised millions of citizens across Spain.

2011: Occupy Wall Street was the name of a protest movement that started on September 17, 2011, in Zuccotti Park, in New York City's Wall Street financial district. It obtained global attention and generated the Occupy movement against social and economic inequality worldwide. It was inspired by anti-austerity protests in Spain coming out of the 15-M movement. 'Social movements during the past decade have taken advantage of ubiquitous computing and social media, facilitating mobility and the flexibility to organise actions and create and share digital content' say Haimson and Cartagena in *Information Occupation: Using Information Science to Explore Occupy Wall Street* (2013). In this study, the authors use the same technologies used to mobilise people to examine them. They determine that technological mediation 'gives rise to, and provides researchers new insight into social movement practices' (ibid.).

2011: Claims about the role of internet, social media and mobile technologies in the political upheaval in the Middle East and North Africa in 2011 –called the Arab Spring— are yet as unclear as the outcomes of some of these national struggles years after the rebellions began to unfold. The suicide of street vendor Mohamed Bouazizi is generally considered as having triggered the rebellion that brought down the Tunisian government and then spread to Egypt, Libya and beyond. In a study about the Tunisian uprising, Lim looks at why Bouazizi's death and the demonstrations that followed were not as easily dismissed by the authorities as earlier

events had been, and that the internet and mobile technology had a role in it. The suicide was filmed, and facts were ‘adjusted’ to frame the death in a way that appealed to a broad range of Tunisians (Lim 2013, 926-927). Twitter and Facebook were key in the coordination of the revolt, but also ‘made an unmistakable impact on the reporting of the Arab Spring revolutions’ (Duffy 2011, 53; EMiróglu 2012,).

Annex III: Additional documentation - Short biographies

Ory Okolloh is an open-government activist, lawyer and blogger who co-founded Ushahidi. She is now Director of Investments at Omidyar Network. Previously, she had founded Mzalendo, a website that helps Kenya's electorate keep track of the activities of their representatives in parliament. After Ushahidi, she became Policy Manager for Google Africa. She also had a role as a World Economic Forum Young Global Leader (Nsehe 2012).

Erik Hersman is a technologist, blogger and commentator who specialises in the impact and application of technology throughout Africa. He is one of Ushahidi's co-founders. He has also started up many web, mobile and community projects harnessing Africa's boundless spirit of innovation through organizations like the iHUB. He is also the founder of the BRCK, a backup generator for the internet, and is a general partner in the Savannah Fund (Hersman 2016). In 2009 he was awarded a TED Fellow Fellowship, and the following year named a Senior TED Fellow (ibid.). He is still at Ushahidi's board.

Juliana Rotich is an information technology professional and a blogger. She is still at Ushahidi's board. As a blogger, she has authored articles on Afrigadget.com, acted as Environmental Editor of *Global Voices Online*. As a public speaker, she is known for her commentary on technology in Africa and voicing concerns about the loss of indigenous forest and water catchment areas in Kenya. She is a Senior TED Fellow. In 2011, Rotich was named Schwab Foundation Social Entrepreneur of the Year in Africa by the World Economic

Forum. She co-founded Mobisoko, a mobile marketplace for language and location relevant apps in Africa (TED 2010). She is still at Ushahidi's board.

David Kobia is the Director of Software at BRCK, co-founder of Ushahidi and Trustee of the iHUB. He is still at Ushahidi's board. David is an MIT Fellow and in 2010, he was a recipient of MIT Technology Review's TR35 award (35 top innovators under 35) and the Humanitarian of the Year award (Kobia 2016).

Daudi Were is Ushahidi Executive Director. He is a technology strategist, entrepreneur, social activist. He was previously Project Director at Ushahidi, managing key partnerships and custom Ushahidi deployments for organisations ranging from multinational organisations to grassroots NGOs. Daudi blogs at Mentalacrobatics, is a TED Africa fellow (Mentalacrobatics 2016).

Patrick Meier is an expert and consultant on humanitarian technology and innovation, and part of the initial core team. His book *Digital Humanitarians* is a book of reference. Over the past 14 years, Meier has worked in the Sudan, Somalia, Kenya, Uganda, Liberia, India, Philippines, Kyrgyzstan, Nepal, Timor-Leste, Turkey, Morocco, Western Sahara, Haiti, Vanuatu and Northern Ireland on a wide range of humanitarian projects. He is the Executive Director & Co-Founder of WeRobotics, which scales the positive impact of humanitarian aid, development and environmental projects through the use of robotics solutions. In addition to leading WeRobotics, he serves as a consultant to the World Bank and Facebook. He blogs at iRevolutions (2016).