

14: DATA FOR THE SOCIAL GOOD: TOWARD A DATA-ACTIVIST RESEARCH AGENDA

BECKY KAZANSKY, GUILLÉN TORRES, LONNEKE VAN DER VELDEN, KERSTI WISSENBACH AND STEFANIA MILAN¹

Introduction

'Big data' is a hyped buzzword - or rather, it has been for a while, before being supplanted by 'newer' acclaimed concepts such as artificial intelligence. The popularity of the term says something about the widespread fascination with the seemingly infinite possibilities of automatized data collection and analysis. This enchantment affects not only the corporate sector, where many technology companies have centered their business model on data mining, and governments, whose intelligence agencies have adopted sophisticated machinery to monitor citizens. Many civic society organizations, too, are increasingly trying to take advantage of the opportunities brought about by datafication, using data to improve society. From crowdsourced maps about gender-based violence ('femicide') in Latin America, to the analysis of audio-visual footage to map drone attacks in conflict zones, individuals and groups regularly produce, collect, process and repurpose data to fuel research for the social good. Problematizing the mainstream connotations of big data, these examples of 'data activism' take a critical stance towards massive data collection and represent the new frontier of citizens' engagement with information and technological innovation.

In this chapter we survey diverse experiences and methodologies of what we call 'data-activist research' - an approach to research that combines embeddedness in the social world with the research methods typical of academia and the innovative repertoires of data activists. We argue that such approach to knowledge production fosters community building and knowledge sharing, while providing a way to fruitfully interrogate datafication and democratic participation. By exploring what we can learn from data-activist projects and investigating the conditions for collaboration between activist communities and academia, we aim at laying the groundwork for a data-activist research agenda whose dynamics are socially responsible and empowering for all the parties involved.

The chapter is organized as follows. We begin offering a working definition of data-activist research. We explain how the notion has developed within the DATACTIVE research collective at the University of Amsterdam, whose work investigates the politics of datafication and massive data collection from the perspective of civil society. We describe how our commitment to 'engaged research' feeds into our ideas about data-activist research.² We build upon interdis-

1 * This project has received funding from the European Research Council (ERC) under the European Union's Horizon2020 research and innovation programme (grant agreement No 639379-DATACTIVE, awarded to Stefania Milan as Principal Investigator). See <https://data-activism.net>.

2 By engaged research we indicate systematic, evidence-based, social science research which is

ciplinary literature on datafication and the valuable insights shared by activists, civil society organizations and engaged researchers at the *Data for the Social Good* workshop (University of Amsterdam, November 2017).³ We discuss concrete examples of existing research projects and their novel tools and approaches. Since our main goal is to call for more interaction between activists and academics, we conclude with a reflection on the ethics of collaboration, as we deem these two elements to be central questions today. We hope that this discussion will encourage the two communities to appropriate and build upon the powerful approach of data-activist research.

Defining Data-activist Research

The label data-activist research emerges at the intersection of 'traditional' research and the set of critical and/or activist practices that deal or 'act upon' datafication.⁴ The roots of data-activist research are to be found in data activism itself, which critically engages with the manifold impact of datafication on social life.⁵ Processes of turning aspects of social life into data are of course not new and have always been at the core of the practices of science and knowledge production.⁶ Nor are efforts to challenge how social life is turned into data a new thing. For example, where statistics have long been used to steer city and health planning, official numbers and calculations have been challenged by 'statactivists' to produce impactful public policy reform.⁷ However, over the last decade, datafication has become a fundamental component of people's lived reality and a major driver of knowledge production. Whether it is through social media use, engaging with the government, buying online goods, or using public transport, people are continuously digitally 'measured', included in databases,⁸ and interact with such measurements through the feedback they get via apps and other devices.⁹ Furthermore, these data are afterwards used in various types of knowledge production activities that feed political and economic decision-making and governance processes.

designed to actively involve and possibly empower disempowered communities and people beyond the academic community. See Stefania Milan, 'Toward an Epistemology of Engaged Research', *International Journal of Communication* 4 (2010), p. 856.

- 3 DATACTIVE, *Workshop Report: Data for the Social Good*, University of Amsterdam, 2017, https://data-activism.net/wordpress/wp-content/uploads/2018/01/DATACTIVE_DataSocialGood2017_Report.pdf.
- 4 Sebastian Kubitschko, 'Acting on Media Technologies and Infrastructures: Expanding the Media as Practice Approach', *Media, Culture & Society* 40.4 (2018).
- 5 Stefania Milan, 'Data Activism as the New Frontier of Media Activism', in Goubin Yang and Viktor Pickard (eds), *Media Activism in the Digital Age*, Oxford: Routledge, 2017.
- 6 See Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Boston: Houghton Mifflin Harcourt, 2013; and Michel Foucault, *The Will to Knowledge: The History of Sexuality*, London: Penguin Books, 1998 (1976).
- 7 See Isabelle Bruno, Emmanuel Didier and Tommaso Vitale, 'Statactivism: forms of action between disclosure and affirmation', *Partecipazione e conflitto* 7.2 (2014).
- 8 Kevin D. Haggerty and Richard V. Ericson, 'The Surveillant Assemblage', *The British Journal of Sociology* 51.4 (2000).
- 9 Kashmir Hill, 'What Happens When You Tell the Internet You're Pregnant', *Jezebel*, <https://jezebel.com/what-happens-when-you-tell-the-internet-youre-pregnant-1794398989>.

The specific way in which civil society actors have responded to the new possibilities and risks brought about by datafication has informed our research into data activism, an umbrella term which embraces, for instance, socio-technical practices that provide counter-hegemonic responses to the discrimination, social exclusion and privacy infringement that go hand in hand with big data.¹⁰ Data activism 'interrogates the politics of big data',¹¹ and it does so in a variety of ways: for instance, '[t]he action repertoire of data activists includes examining, manipulating, leveraging, and exploiting data, along with resisting and meddling in their creation and use'.¹² In other words, data activism includes both the use, mobilization or creation of datasets for social causes (providing an alternative to what big data corporations or state agencies do with data), as well as the development and employment of technologies that frustrate massive data collection (providing protection to what big data corporations or state agencies do with data).¹³

Studying the methods and strategies of data activism led us to question our own *research processes, practices and relationships*. This is because data activism signals the emergence of innovative 'epistemic cultures',¹⁴ namely experimental and context-specific ways of producing knowledge about and with data. As Milan and van der Velden suggested, by '[p]ostulating a critical/active engagement with data, its forms, dynamics, and infrastructure, data activists function as producers of counter-expertise and alternative epistemologies, making sense of data as a way of knowing the world and turning it into a point of intervention'.¹⁵ Take for instance the artist Mimi Onuoha, who created a 'Library of Missing Datasets' to strategically draw attention to important issues of social justice which could benefit from more data,¹⁶ or the experience of the activists using drones to counter decades of injustice over oil exploitation in the Amazon rainforest.¹⁷ These cases signal that (data) activism is a powerful location for knowledge production able to fuel political projects, through practices that draw from institutionally entrenched approaches to research while simultaneously subverting, expanding and questioning their components.

10 Stefania Milan, 'Data Activism as the New Frontier of Media Activism', p 152.

11 Ibid p. 153.

12 Ibid p. 143.

13 See also Becky Kazansky, 'Privacy, Responsibility, and Human Rights Activism', *The Fibreculture Journal*, 26 (2015); Lonneke van der Velden, 'Leaky Apps and Data Shots: Technologies of Leakage and Insertion in NSA Surveillance', *Surveillance & Society* 13.2 (2015).

14 The notion of 'epistemic culture' is used in science studies and refers to the 'specific strategies that generate, validate, and communicate scientific accomplishments'. See Karin Knorr-Cetina and Werner Reichmann, 'Epistemic Cultures' in *International Encyclopedia of the Social & Behavioral Sciences*, second edition, Oxford: Elsevier, 2015 pp. 873-80. The concept highlights the diversity in scientific practices. Here we use it to discuss the diversity in knowledge making in the context of datafication.

15 Stefania Milan and Lonneke van der Velden, 'The Alternative Epistemologies of Data Activism', *Digital Culture & Society*, 2.2 (2016).

16 Mimi Onuoha, 'The Library of Missing Datasets', <http://mimionuoha.com/the-library-of-missing-datasets/>.

17 If not us then who?, 'Detecting Disasters'. <https://ifnotusthenwho.me/films/using-drone-technology-detect-oil-spills/>. See also Stefania Milan and Miren Gutiérrez, 'Technopolitics in the Age of Big Data', in Francisco Sierra Caballero and Tommaso Gravante (eds), *Networks, Movements & Technopolitics in Latin America: Critical Analysis and Current Challenges*, London: Palgrave Macmillan, 2017, pp. 95-109.

It is in collaboration with these novel epistemic cultures that we see possibilities for constructive interaction between activism and academia, and for a joint discussion about what 'data activist research for the social good' could look like. This entails not only a reflection *about* data activism that tries to locate its most innovative and empowering research practices, but also entails paying attention to what engaged and productive role academia could play in the process. In other words, can we do data-activist research ourselves, and if so, how? What could academia learn from these emerging practices and what could it offer back? What are the conditions of possibility for joint research projects? We argue that to provide the best answer to these questions it is necessary to move beyond doing research *about* (data) activism, towards conducting institutional boundary-crossing research that finds common grounds and opportunities for collaboration *with* (data) activists.¹⁸ In the next section we further explore this claim.

Data-activist Research is Engaged Research

Several members of the DATACTIVE research group have known or have been involved for long with the communities they study. They have faced an important question that arises when researching groups one is closely affiliated with: how to develop and deploy a research pathway that is most relevant for the community, making sure that the community itself can contribute to shape both the project's goals and practices? In other words, how can we do research that matters also to those being researched? These concerns are certainly not new in academia, and there are several examples of individuals and groups who approached research *in a different way*. Early attempts at co-producing knowledge while reflecting upon its connection to community empowerment can be found in the 1960s and 1970s. They were influenced by the writings of Brazilian educator Paulo Freire and the *con-ricerca* (co-inquiry) experiments in Italy, for example involving factory workers in analyzing the social impact of capitalism.¹⁹ Since the 1980s, Charlotte Ryan, co-director of the now dormant Movement/Media Research Action Project (MRAP) at Boston College (US), has been experimenting with producing recursive 'two-way, dialogic exchanges that create new, generalizable knowledge' expected to contribute to the 'democratization of theorizing'.²⁰ Because both theorizing and practice benefit if scholars 'embed themselves in movements, not simply as active citizens but as skilled learners',²¹ MRAP members encouraged activists and scholars to establish 'learning communities; based on shared learning practices and work routines. More recently, Lovink and Rossiter have pointed to the importance of working together with actors in the field, since 'collaborative concept production' is needed in order to keep theory up to date.²² Similarly, the *DataCenter: Research for Justice* organization in Oakland, California, have characterized its Research Justice Model as having three main tenets:

18 This argument was made earlier in a series of articles published in the Feature 'Making Communication Matter' of *the International Journal of Communication* (4/2010) edited by Stefania Milan (see <http://ijoc.org/index.php/ijoc/issue/view/5>).

19 Paulo Freire, *Pedagogy of the Oppressed*, New York: Continuum, 2007 (1968).

20 C. Ryan and K. Jeffreys, 'The Practice of Collaborative Theorizing', unpublished manuscript (2008).

21 Ibid, p. 3.

22 Geert Lovink and Ned Rossiter, *Organization after Social Media*, Colchester: Minor Compositions, 2018, p. 75.

1. It defines research processes as a collective endeavor and a shared knowledge creation process between and academic and community researchers;
2. It creates, maintains, and engages with the knowledge that is produced by community experts, traditional knowledge keepers, as well as cultural leaders in ways that envision research as a ceremonial act of mutual respect and co-sharing; and
3. Only research that is responsive to the social, legal, economic cultural, and political policy needs as identified by community experts should be conducted.²³

Drawing from these sources of inspiration, DATACTIVE proposes an 'engaged' approach to research that questions the impact that empirical inquiry has over people and communities, and strives to contribute to their causes.²⁴ Such an approach entails to do research 'with' instead of merely 'about', thus entering into a continuous dialogue with the fields of action and interaction being observed.²⁵ Nevertheless, an engaged approach to research does not lose sight of the wider context and maintains a sharp attention to the question of power.

In our view, data-activist research should thus emerge as the result of community endeavors whose perspectives and self-definitions can be located in specific and contested discourses about technology, information, activism, marginalization, exclusion and even selfhood, rather than being merely the result of the interaction between disembodied agents in a universal field of knowledge. In what follows we present four case studies that give a sense of what data-activist research might mean in practice.

Data-activist Research in Practice

Forensic Oceanography, *The Syrian Archive*, and the local instances of the *Alaveteli* software are good examples of data-activist research which succeed at performing a series of steps allowing activist-researchers to do 'research that matters'. The three projects managed to remain close to the problems they identified, to then take a step back to develop an abstract understanding and analysis of the reality, only to return to the field to address the community issues that had been identified. As we will see, what these projects demonstrate is that research processes are more productive when they are meaningful to specific communities rather than merely a product of 'disembodied scientific objectivity'.²⁶

23 Andrew J. Jolivet, *Research Justice: Methodologies for Social Change*, Bristol: Policy Press, 2015.

24 Stefania Milan, 'Toward an Epistemology of Engaged Research.'; Chiara Milan and Stefania Milan, 'Involving Communities as Skilled Learners: The STRAP Framework', in Norbert Wildermuth and Teke Ngomba (eds), *Methodological Reflections on Researching Communication and Social Change*, Cham: Springer, 2016, 9-28.

25 Cf. C. Ryan, V. Salas-Wright, M. Anastario & G. Camara, 'Making Research Matter... Matter to Whom?'

26 Donna Haraway, 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective', *Feminist Studies*, 14.3 (1988): 576.

Forensic Oceanography: In search of a 'disobedient gaze'

Forensic Oceanography is the maritime counterpart of *Forensic Architecture* focusing specifically on migration and bordering. *Forensic Architecture* started in 2011 as an ERC-funded research project at Goldsmiths, University of London (UK), to turn later into a stand-alone research agency focused on the production of evidence of human rights violations. Through the analysis of architecture, the environment, and its media representations, researchers have provided prosecution teams, political organizations and international institutions alike with evidence that can be used in court. The research agency has engaged in a variety of projects spanning from the analysis of deadly drone attacks in Syria, to the disappearance of 43 Mexican students, to the ecocide in Indonesia.²⁷ The data sources that *Forensic Architecture* relies upon to fuel its investigation are of varied nature - from satellite images, to publicly available data and media produced by the communities involved in the events under scrutiny.

The 'Left to Die Boat', a project by *Forensic Oceanography*, is a good example of the work of *Forensic Architecture*: it reconstructs the story of a vessel that left Libya with 72 people on-board in the midst of the NATO-led military intervention in the country. The boat ran out of fuel, drifted for two weeks and was finally washed back to the Libyan coast. Most of its passengers died. The survivors stated that they had contact with several ships and helicopters, but no one intervened to help. Using publicly available databases on sea traffic, the researchers traced and visualized the contacts made by the boat, proving that a number of ships, including military vessels, were indeed navigating close by, but chose not to intervene. The evidence allowed advocates to start a number of legal petitions against NATO member states, accused of the crime of non-assistance at sea.²⁸

Forensic Oceanography shows how it is possible to use as research input monitoring technologies, including those typically used by police forces, with the goal of bringing about a 'disobedient gaze' - a perspective that challenges the dominant narrative. As Pezzani and Heller explain it, this disobedient gaze performs a reversal of the surveilling action, turning its sight to the act of policing itself.²⁹ Through this inversion, *Forensic Oceanography* brings to light events and issues that the surveilling system prefers to hide. It also shows how monitoring technologies can be used to hold accountable the very agents who set them in place to exert power. Thus, this project - as well as the rest of *Forensic Architecture's* work - makes evident how the availability of data can foster the creation of new mechanisms of participation that take advantage of technologies designed for other purposes. In this sense, *Forensic Oceanography* is a great example of the diverse politics of datafication, since the data produced by surveilling technologies can also be processed to provide backing evidence to strengthen the politicization of contemporary social issues.

27 For the project about Syria, see: <http://www.forensic-architecture.org/case/al-jinah-mosque/>; for the case of the missing Mexican Students, see: <http://www.forensic-architecture.org/case/ayotzinapa/>; for the Indonesia case see: <http://www.forensic-architecture.org/case/ecocide-indonesia/>.

28 <http://www.forensic-architecture.org/case/left-die-boat/>.

29 Lorenzo Pezzani and Charles Heller, 'A Disobedient Gaze: Strategic Interventions in the Knowledge(s) of Maritime Borders', *Postcolonial Studies*, 16.3 (September, 2013).

Departing from *Forensic Oceanography's* work, it could be interesting for data-activist research to think about what other kinds of 'disobedient data politics' are possible,³⁰ and what their ethical implications are. For example, given that many of the technologies used by *Forensic Oceanography* have been designed with the goal of performing surveillance, it is important to remain attentive to the question of whether there are risks in using them. If data-activist research engages with scraping, data monitoring, etc., how can activist-researchers engage in these activities in a responsible way that does not reproduce the extractive and exploitative rationality of the mainstream discourse and practices? What we learn from this type of projects is that researchers and the communities they work with benefit from 'continuously reflecting upon whether their investigation contributes to a "disobedient gaze", rather than merely a vigilant one.'³¹

The Syrian Archive: Turning Open Source Intelligence Upside-down

The *Syrian Archive* is an Open Source Intelligence (OSINT) effort to document and preserve information about human rights violations committed by all sides of the Syrian conflict.³² Started in 2014, the project brings together developers and human rights activists focusing on the preservation of media evidence under threat of being deleted or censored from the online platforms where it is uploaded. Its main goals are to secure data, verify its authenticity, and categorize it. The resulting database allows the wider public to reuse the material for various purposes, although evidence gathering concerning human rights violations is the primary rationale.

The *Syrian Archive* aims at implementing ethical principles starting from the design of the technology that powers its activities, the methodologies, and the way its activist-researchers preserve findings. The tools built in the context of the project are open source and most of the code used to process and organize the data is made available in the software repository GitHub. The project also follows a user-centered approach maintaining regular contact with media sources, who have provided so far more than 1 million entries to the archive, all of which have undergone verification and categorization. The project's ultimate goals are to identify reliable sources of data collection, organize the material in a database, establish the trustworthiness of the content, and automatizing data collection and preservation.

The *Syrian Archive's* methodology makes evident that even working with publicly available data has severe ethical implications. For example, one of the many thorny issues its activist-researchers constantly reflect upon is how to acquire the consent of those depicted in the footage, or how to decide what should be preserved and what should be discarded. To

30 On disobedient sensing and listening see Charles Heller, Lorenzo Pezzani, and Maurice Stierl, 'Disobedient Sensing and Border Struggles at the Maritime Frontier of Europe', *Spheres: Journal for Digital Cultures* 4 (June, 2017).

31 Pezzani cited in DATACTIVE, 'Workshop Report', p. 14.

32 <https://syrianarchive.org/en/about>. Open Source Intelligence is data gathering based on publicly available sources.

guide decisions, the project follows a 'do no harm' approach,³³ taking care to exclude certain sensitive data. 'Do no harm' refers to a set of protocols intended to make sure that humanitarian practitioners do not end up further harming the situation they intend to improve.³⁴ Additionally, for the *Syrian Archive*, 'open source' does not only refer to the public availability of its materials, but also a specific approach to the transparency of protocols and practices. Keeping the software tools open allows other activists to replicate the work of the organization.

In conclusion, the activities of the *Syrian Archive* stress that, even if one pursues an urgent goal through activism - such as collecting data about human rights violations before it disappears or gets censored - it is still of paramount importance to pay attention to the consequences of data gathering, processing and sharing facilitated by digital technology. Furthermore, the project shows how it is possible to build databases with a rationality that does not aim at maximizing control or private benefit, but focuses on its political potential.

Alaveteli: Engaging with communities across borders

Alaveteli identifies a Freedom of Information (FOI) request platform and the community that emerged around it. It is currently implemented in 25 countries across the world.³⁵ The original platform was launched in 2008 in the UK under the name 'What do They Know'. The open source code of the platform, however, was quickly picked up by other civil society actors (the first in New Zealand), before the e-democracy project *mySociety* made its own iteration available for everyone under the name *Alaveteli*, offering support for groups who were interested in adopting it locally. *Alaveteli* enables citizens to openly request information from government institutions, allowing the whole process to be tracked online and the institutional replies to be available for everyone. In each local deployment, the success or failure of advocating for FOI and engaging civil society through the platform depends on a multitude of factors, such as cultural dynamics, political restrictions, and infrastructural limitations. For example, the backgrounds of the actors who have picked up the code locally are very diverse, ranging from political activists to journalists, from technologists to human rights organizations. The responses to context-related challenges are therefore also varied. While some platform implementations are deeply rooted in an activist ethos, with people spending significant proportion of their volunteer time into platform management and mobilization, other *Alaveteli* communities have been more efficient in pursuing social innovation grants to localize the platform and engage in awareness-raising. The long-term success of the platform usually depends on the ability of the actors involved to establish wider collaborations. For instance, if a group of techies has set up the platform, collaborations with advocacy groups help to create awareness, increase engagement and establish links with potential users such as journalists.

33 Mary B. Anderson, *Do No Harm. How Aid Can Support Peace or War*, Boulder CO/London: Lynne Rienner Publishers, 1999.

34 At the same time, this is not an easy approach when a researcher has to deal with unethical actors such as perpetrators of human rights violations.

35 See <http://alaveteli.org/>. Alaveteli is a good example of Civic tech activism, an emerging instance of organizing collective action that engages in institutionally regulated governance processes through the crafting of direct engagement spaces for civil society and, thus, pushing governing institutions toward more accountability.

Cycling back to our original quest for 'good' collaborations between researchers and the communities on the ground, *Alaveteli* well exemplifies the crucial role of human interaction and relationship-building around technological innovation if this is to become relevant for local communities. The platform's reliance on the local context and its specificities underscores also the importance of making use of the already existing infrastructure - encompassing both technology and human relations - and building on the previous experience of local activists. In sum, data-activist research requires paying attention not only to infrastructure practices, but also to local contexts and human dynamics.

One from the house: Studying collaboration in online communities

The DATACTIVE research group has contributed to the development of a computational research tool called *BigBang*, 'a toolkit for studying communications data from collaborative projects'.³⁶ Our interest in *BigBang* grew out of the desire to understand how the human rights discourse has evolved within multi-stakeholder discussions about the governance of international data flows. Thus far, this research has targeted a number of community mailing lists within the Internet Corporation for Assigned Names and Numbers (ICANN).³⁷

Among other functionalities, *BigBang* allows researchers to scrape large swaths of data from a mailing list database and easily search for keywords. *BigBang* has proven especially useful to the study of ICANN because the large majority of community interactions takes place on mailing lists. As a large community with thousands of contributors across the globe, ICANN produces many data traces. This amount of data can prove cumbersome for manual analysis, hence automating the search for keywords makes the task of investigating the discursive evolution in internet governance processes more manageable. However, the toolkit brings up some concerns because it facilitates research techniques in the realm of 'big data' analysis - a set of techniques which the DATACTIVE project investigates with a critical eye. Partaking in the development and use of this tool presents an interesting opportunity for us to reflect on our research ethics, the 'why' of our research, and our connection with the issues at stake.

Take for example the distinct understandings and expectations of privacy in different community-contexts - a question which is relevant to most data-activist research projects given their reliance on publicly available data. During DATACTIVE's internal discussions we have raised concerns about the expectations of privacy that can be found in different online contexts, and asked how these expectations are affected when the data can be more easily collected and analyzed by third parties - as *BigBang* makes possible. ICANN is a community which conducts much of its work 'in the open' - a fundamental requisite of its multi-stakeholder nature. Because of the open nature of the data the organization produces, DATACTIVE felt it was ethically sound to use it after producing a list of conditions guiding its acceptable use.

36 See <https://github.com/dataactive/bigbang>. *BigBang*'s initiator and Lead Developer is computer scientist Sebastian Benthall (UC Berkley & NYU).

37 See Stefania Milan and Niels ten Oever, 'Coding and encoding rights in internet infrastructure', *Internet Policy Review* 6.1 (2017).

These conditions are the result of questioning our goals and intentions: why do we harvest these data? Do we need it to achieve the goals of our research? Who is affected by our data collection and analysis, and how? Who benefits from our research? Among others, we learnt that when it comes to online content the level and the modalities of publicity, including academic publications, need to be determined on a case by case basis and in collaboration with the participant communities themselves.

Inspired by these examples and acknowledging that data activism, in its many forms, emerges from a plurality of social worlds and identities,³⁸ we deem crucial to reflect upon issues of collaboration. We now turn our attention to this aspect, in order to contribute to sketch the groundwork for a joint research agenda between data activists and academics.

What Collaborations for Data-Activist Research?

Reflecting on the politics of collaboration must be seen as a central methodological task when dealing with the production and use of data to fuel political projects in the interest of society at large. Such reflection has to be guided by the recognition of the existing difference in organizational cultures, modus operandi, goals and values that characterize activists on the one hand, and researchers on the others.³⁹ Collaborative data-activist research strategies can benefit from researchers and communities developing questions and research practices jointly from the start, remaining open for the exchange of different types of know-how despite the apparent difference in expertise.⁴⁰ Such approach aims to go beyond the 'distant reading' of the data points activists produce, moving instead towards a 'critical proximity' that remains close to the issues approached, participating in their development.⁴¹

The *researching with* that we highlighted as a crucial feature of data-activist research can benefit from the process of building a 'we'; a shared identity resulting from a set of iterative activities, dialogues and reflections connected to fundamental questions such as how do we, as a community, define what the issue at stake is? How do we identify mechanisms to address

38 Eliana Herrera Huérfano, Francisco Sierra Caballero and Carlos del Valle Rojas, 'Hacia una Epistemología del Sur. *Decolonialidad* del saber-poder informativa y nueva Comunicología Latinoamericana. Una lectura crítica de la mediación desde las culturas indígenas', *Chasqui. Revista Latinoamericana de Comunicación* 131 (April-June, 2016).

39 Milan, 'Toward an Epistemology of Engaged Research.'; Milan & Milan, 'Involving Communities as Skilled Learners: The STRAP Framework.'

40 Milan, 'The Ethics of Social Movement Research', in Donatella della Porta (ed.), *Methodological Practices in Social Movement Research*, Oxford: Oxford University Press, 2014. See also Graeme Chesters, 'Social Movements and the Ethics of Knowledge Production', *Social Movement Studies*, 11.2 (2012); Milan & Milan, 'Involving Communities as Skilled Learners: The STRAP Framework'; Donatella Della Porta and Elena Pavan, 'Repertoires of Knowledge Practices: Social Movements in Times of Crisis', *Qualitative Research in Organizations and Management*, 12.4 (2017).

41 See for instance Lorenzo Pezzani and Charles Heller, 'A Disobedient Gaze: Strategic Interventions in the Knowledge(s) of Maritime Borders'; Andreas Birkbak, Morten Krogh Petersen and Torben Elgaard Jensen, 'Critical Proximity as a Methodological Move in Techno-Anthropology', *Techné: Research in Philosophy and Technology*, 19.2 (2015).

it? What core values guide us in the process?⁴² From this perspective research is a social process that demands a careful consideration of 'for whom' and 'to what end' it is conducted.⁴³

How, then, do we enable collaborative data activist research? This question addresses how the relations between, and the engagement of researchers, activists, and wider civil society look like.⁴⁴ Charlotte Ryan has highlighted the importance of working in cycles of dialogue rather than a one-off exchange, continuously assessing the meaningfulness of one's research and the conditions of inequality between researchers and activist/communities.⁴⁵ As we have mentioned before, a collaborative, dialogue-based data-activist research methodology that fosters the process of community building and knowledge sharing has to depart from a joint reflection on what knowledge and its production mean, and what building a 'we' entails. However, no process by itself has the ability to erase power asymmetries - imbalances can very well occur within activist communities themselves along lines of race, class, gender, expertise, etc. Therefore, processes of collective research design and analysis need to take into account the power asymmetries prevalent among the actors involved and consciously reduce space for hierarchies. What are, then, the building blocks of an ethics of data-activist research?

The Ethics of Data-Activist Research

Within a data-activist research methodology, ethics should be understood as a process rather than a mere checklist. In conceiving of it as a process we take inspiration from the ethics guidelines by the Association of Internet Researchers (AoIR),⁴⁶ and feminist 'ethics of care',⁴⁷ which puts a caring relationship with research subjects at centre stage. In what follows, we offer a list of potential starting points in thinking about research ethics.

1. Do no harm

Data-activist research goes beyond the idea of attempting not to negatively impact the communities involved. The guideline is to collectively bring about a difference for such communities,

42 See Ryan in DATACTIVE, 'Workshop Report'.

43 The discipline of Social Movement Studies has to some extent engaged with the question of making research relevant for the research subjects. See e.g. David Croteau, William Hoynes and Charlotte Ryan (eds), *Rhyming Hope and History: Activists, Academics, and Social Movement Scholarship*, Minneapolis: University of Minnesota Press, 2005; the Special issue on 'The Ethics of Research on Activism', *Social Movement Studies*, 11.2 (2012); Milan, 'The Ethics of Social Movement Research'.

44 The concept of inclusive participation has been addressed in fields such as Critical Development Studies or communication for social change. See Alfonso Gumucio Dagron and Thomas Tufté (eds), *Communication for Social Change Anthology: Historical and Contemporary Readings*, South Orange NJ: Communication for Social Change Consortium, 2006. It is here that we also find an analytical/methodological account of dialogue: see e.g. Alejandro Barranquero, 'Rediscovering the Latin American Roots of Participatory Communication for Social Change', *Westminster Papers in Communication and Culture* 8.1 (May 2011); and Freire, 'Pedagogy of the Oppressed'.

45 See, for example, Ryan in DATACTIVE, 'Workshop Report'.

46 See AoIR, Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0), <https://aoir.org/reports/ethics2.pdf>.

47 Virginia Held, *Ethics of Care*, New York/London: Oxford University Press, 2005.

aiming for a positive impact as one of the main outcomes of the research.⁴⁸ Researchers should ask whose goals the research does or might further serve, as well as what harms might come from having particular experiences or vulnerabilities exposed and made public as research findings.

2. *Setting equitable research agendas*

If we talk about data-activist research from a perspective centered on collaboration, one key consideration comes to mind: data, where it is meant to be produced and used in the interest of activists or the wider civil society, has to be representative of the needs and interests of those it means to 'support'.⁴⁹ However, from a methodological perspective, the reflection around collaboration must go beyond a focus on representation; it builds on the idea that people are in charge of the decision-making processes on which their very realities are constructed.⁵⁰ Also in the emerging field of data activism, in which data forms the main currency of engagement in advocacy tactics,⁵¹ forms of collaboration and engagement with civil society in order to identify relevant tactics proves crucial for realizing representative data structures.⁵²

3. *Re-centering perspectives pushed to the periphery*

Researchers should be critical of overly focusing on expert opinions, as these can be used as proxies for the issue or groups being studied, while much of the labor of knowledge production is being done elsewhere.⁵³ To this end, researchers should strive to look beyond the most prominent names when 'sampling' and selecting research subjects, and adopt a conscious strategy of seeking out expert opinions from underrepresented populations such as women, people of color, affected populations, and other minority groups.

4. *Transparency of research objectives (and funding)*

Researchers should disclose the aims of their projects and communicate the 'why' of the research to those involved in any research activities - whether an interview, ethnographic participant observation or a joint policy advocacy project. Researchers should be clear that theirs is not a 'view from nowhere', but a situated perspective.⁵⁴ Issues of class, race, and locality of the researchers should be reflected upon within the research.

48 Anderson, 'Do no harm'; Milan, 'Toward an epistemology of engaged research'.

49 Linnet Taylor and Dennis Broeders, 'In the Name of Development: Power, Profit and the Datafication of the Global South', *Geoforum* 64 (August, 2015).

50 Freire, 'Pedagogy of the Oppressed'.

51 Milan and Velden, 'The Alternative Epistemologies of Data Activism'.

52 Geoffrey Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences*, Cambridge MA: MIT Press, 1999.

53 In DATACTIVE, 'Workshop Report'.

54 Haraway, 'Situated Knowledges'.

5. *Recognizing research as labor*

Researchers should understand that interacting with researchers and 'being researched' is a form of labor.⁵⁵ Sitting in for interviews or engaging in other research activities takes time away from urgent work, including gaining an income. Thus, researchers should strive to minimize disruptions caused by their participation in activities.⁵⁶ On the other hand, there may be instances in which researchers should also clearly lay out expectations around their own labor of research, for example, by explaining why it might take a certain amount of time for findings to be 'fed back' or published.⁵⁷

6. *Contextualizing data and data collection*

Researchers should examine the context and potential consequences of studying communities, identities, projects, networks, and dynamics. Some data that is considered public is actually just 'publicly available (sensitive) data'.⁵⁸ A minimum standard for much of social science research is to obtain the consent of research subjects. Yet in projects using 'big data', this can be difficult. Data-activist researchers should thus put adequate attention to strategizing how they will anonymize any data they use on online communities and consider if it is ethical to collect it in the first place - and they should be up to date with respect to data protection regulations which might prohibit its collection.

7. *Responsible data management and sharing*

Researchers should strive to create an information management plan prioritizing the privacy and security of research data. The development of a plan should root itself in the particular scenarios of the research life and should consider all phases of a research project.⁵⁹ This also includes a plan of how to store and back up research data; how to share data with other researchers; how to transport data while traveling across borders; how to guard data while at field sites; and how to communicate sensitive details within the research team as well as with research subjects.

8. *Fair attribution*

Researchers should provide correct attribution, anonymizing and pseudonymizing as necessary, or should mention interviewees by name if requested. This is a fundamental step in the recognition of social actors as knowledge producers in their own right, no less than external observers.

55 Arne Hintz and Stefania Milan, 'Social science is police science: Researching grassroots activism', *International Journal of Communication* 4 (2010).

56 Ibid.

57 See Ryan in DATACTIVE, 'Workshop Report'.

58 Jacob Metcalf and Kate Crawford. 'Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide', *Big Data & Society* 3.1 (January, 2016).

59 Milan and Milan, 'Involving Communities as Skilled Learners: The STRAP Framework'.

9. Sharing research results

An ethical stance forces researchers to 'share back' with their informants. Are research subjects able to access the work they have contributed to freely, or are publications beyond paywall? Are research subjects able to provide feedback and discuss findings (in terms of time and accessibility of language) before it is published? For example, our hope within the DATACTIVE project is that researching strategies to enhance privacy, digital security, and open source investigations in the midst of human rights and social justice related activities can provide useful information back to civil society actors for their own purposes.

Conclusions and Open Issues

In this chapter we have dealt with a number of methodological and ethical questions that need to be addressed while using and producing data to fuel political projects in the interest of society at large. With the help of four examples, we discussed several aspects from the field of data activism that researchers - particularly those aiming to work with (data) activists - could incorporate in their own work. We have taken a brief look over matters of (disobedient) data research, collaboration and empowerment, and data ethics. These examples have helped us to build a series of recommendations for researchers in light of our own interest in developing joint research projects between data activists and academia. Much work is however needed to expand the range of problems and solutions addressed in a data-activist fashion. Only a broad, collaborative discussion can help us moving this agenda forward: we thus call upon the engaged-researchers and researching-activists across the globe to experiment and share in a long-term exercise of re-thinking what doing 'research that matters' means in the age of datafication.

References

- Anderson, Mary B. *Do No Harm. How Aid Can Support Peace or War*, Boulder CO/London, 1999.
- Barranquero, Alejandro. 'Rediscovering the Latin American Roots of Participatory Communication for Social Change', *Westminster Papers in Communication and Culture* 8.1 (2011), 154–177.
- Birkbak, Andreas, Morten Krogh Petersen, Torben Elgaard Jensen, and Philosophy Documentation Center. 'Critical Proximity as a Methodological Move in Techno Anthropology', *Techné: Research in Philosophy and Technology* 19.2 (2015): 266-90.
- Bowker, G, and SL Star. *Sorting Things Out: Classification and Its Consequences*. Cambridge MA: MIT Press, 1999.
- Bruno, Isabel, Emmanuel Didier and Tommaso Vitale. 'Statactivism: forms of action between disclosure and affirmation', *Partecipazione e conflitto. The Open Journal of Sociopolitical Studies* 7 (2014), 198-220.
- Chesters, Graeme. 'Social Movements and the Ethics of Knowledge Production', *Social Movement Studies* 11.2 (1 April, 2012): 145-60.
- Dagron, Alfonso Gumucio, and Thomas Tufte (eds). *Communication for Social Change Anthology: Historical and Contemporary Readings*, South Orange NJ: Communication for Social Change Consortium, 2006.

- Dalton, Craig M, Linnet Taylor, and Jim Thatcher. 'Critical Data Studies: A Dialog on Data and Space', *Big Data & Society* 3.1 (2016).
- DATACTIVE. 'Workshop Report: Data for the Social Good', Amsterdam: University of Amsterdam, November 2017.
- Della Porta, Donatella, and Elena Pavan. 'Repertoires of Knowledge Practices: Social Movements in Times of Crisis', *Qualitative Research in Organizations and Management: An International Journal* 12.4 (2017): 297-314.
- Michel Foucault. *The Will to Knowledge: The History of Sexuality: 1*, London: Penguin Books, 1998 [1976].
- Freire, Paulo. *Pedagogy of the Oppressed*, New York: Herder and Herder, 1968.
- Haggerty, Kevin D., and Richard V. Ericson. 'The Surveillant Assemblage', *The British Journal of Sociology*, 51.4 (2000): 605-22.
- Haraway, Donna. 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective', *Feminist Studies* 14.3 (1988): 575-99.
- Heller, Charles, Lorenzo Pezzani, and Maurice Stierl. 'Disobedient Sensing and Border Struggles at the Maritime Frontier of Europe', *Spheres: Journal for Digital Cultures*, 4 (2017), <http://spheres-journal.org/disobedient-sensing-and-border-struggles-at-the-maritime-frontier-of-europe/>.
- Hill, Kashmir. 'What Happens When You Tell the Internet You're Pregnant', *Jezebel*, <https://jezebel.com/what-happens-when-you-tell-the-internet-youre-pregnant-1794398989>.
- Jolivet, Andrew J. *Research Justice: Methodologies for Social Change*, Bristol: Policy Press, 2015.
- Knorr-Cetina, Karin, and Werner Reichmann. 'Epistemic Cultures', in *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition), Oxford: Elsevier, 873-80, <http://www.sciencedirect.com/science/article/pii/B9780080970868104544>.
- Kubitschko, Sebastian. 'Acting on Media Technologies and Infrastructures: Expanding the Media as Practice Approach', *Media, Culture & Society* 40.4 (May 2018): 629-35.
- Lovink, Geert and Ned Rossiter. *Organization after Social Media*, Colchester: Minor Compositions, 2018.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Boston: Houghton Mifflin Harcourt, 2013.
- Metcalfe, Jacob, and Kate Crawford. 'Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide', *Big Data & Society* 3.1 (2016).
- Milan, Chiara, and Stefania Milan. 'Involving Communities as Skilled Learners: The STRAP Framework,' in Norbert Wildermuth and Teke Ngomba (eds), *Methodological Reflections on Researching Communication and Social Change*, Cham: Springer, 2016, 9-28.
- Milan, Stefania. 'Between Datafication and Encryption: Media Activism in Times of Big Data', Annenberg School of Communication: University of Pennsylvania, 2014.
- '. 'Data Activism as the New Frontier of Media Activism', in Goubin Yang and Viktor Pickard (eds), *Media Activism in the Digital Age*, Oxford: Routledge, 2017.
- '. 'The Ethics of Social Movement Research', in *Methodological Practices in Social Movement Research*, 446-64. Oxford: Oxford University Press, 2014.
- '. 'Toward an Epistemology of Engaged Research', *International Journal of Communication* 4 (2010): 856-58.

- Milan, Stefania, and Miren Gutiérrez. 'Technopolitics in the Age of Big Data', in F. Sierra Caballero & T. Gravante (eds.), *Networks, Movements & Technopolitics in Latin America: Critical Analysis and Current Challenges*, London: Palgrave MacMillan, 2017, pp. 95-109.
- Milan, Stefania and Niels ten Oever. 'Coding and encoding rights in internet infrastructure', *Internet Policy Review*, 6.1 (2017).
- Milan, Stefania, and Lonneke van der Velden. 'The Alternative Epistemologies of Data Activism', *Digital Culture & Society*, 2.2 (2016): 57-74.
- Pezzani, Lorenzo, and Charles Heller. 'A Disobedient Gaze: Strategic Interventions in the Knowledge(s) of Maritime Borders', *Postcolonial Studies* 16.3 (September 2013): 289-98.
- Ryan, Charlotte, V., Salas-Wright, M. Anastario and G. Camara. 'Making Research Matter... Matter to Whom', *International Journal of Communication* 4 (2010): 845-55.
- Ryan, C. and K. Jeffreys. 'The Practice of Collaborative Theorizing', Unpublished manuscript (2008).
- Tactical Technology Collective and Becky Kazansky. 'FCJ-195 Privacy, Responsibility, and Human Rights Activism', *The Fibreculture Journal* 26 (2015): 190-208.
- Taylor, Linnet, and Dennis Broeders. 'In the Name of Development: Power, Profit and the Datafication of the Global South', *Geoforum* 64 (2015): 229-37.
- Velden, Lonneke van der. 'Leaky Apps and Data Shots: Technologies of Leakage and Insertion in NSA Surveillance', *Surveillance & Society* 13.2 (2015): 182-96.
- Wildermuth, Norbert, and Teke Ngomba (eds). *Methodological Reflections on Researching Communication and Social Change*, Cham: Springer, 2016.
- Wolfinger, Emily. "'But it's Already Public, Right?": The Ethics of Using Online Data', http://datadriven-journalism.net/news_and_analysis/but_its_already_public_right_the_ethics_of_using_online_data.
- Zimmer, Michael. "'But the Data Is Already Public": On the Ethics of Research in Facebook.', *Ethics and Information Technology* 12.4 (2010): 313-25.