DATACTIVE

Big Data from the South

Towards a Research Agenda

Stefania Milan & Emiliano Treré (eds.)

Data activism: The politics of big data according to civil society

Table of Contents

Big Data from the South: Towards a Research Agenda

University of Amsterdam 4–5 December 2018

4 December, 3pm
Can data be decolonized?
by Nick Couldry (LSE) & Ulises A. Mejias (SUNY Oswego)
UvA library, Belle van Zuylen

4 December, 8pm
Big Data from the South:
Decolonization, Resistance & Creativity

with Payal Arora (Erasmus University Rotterdam), Nick Couldry (LSE),
Merlyna Lim (Carleton University) & Ulises A. Mejias (SUNY Oswego)
SPUI25

5 December, 10am–5pm
immersive research workshop
Department of Media Studies (invitation only)

data-activism.net/bigdatasur   #BigDataSur

BIG DATA SUR BIG DATA FROM THE SOUTH

erc European Research Council Established by the European Commission

ACES AMSTERDAM CENTRE FOR EUROPEAN STUDIES

ASCA

ACGS SPUI25

DAT ACT IVE
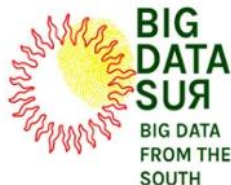
Introduction

What would datafication look like seen… 'upside down'? What questions would we ask? What concepts, theories and methods would we embrace or have to devise? These questions were at the core of the two-day immersive research workshop 'Big Data from the South: Towards a Research Agenda' (University of Amsterdam, 4-5 December 4-5 2018). The event was the third gathering of the Big Data from the South Initiative (BigDataSur)[1], a research network and program launched in 2017 by Stefania Milan (University of Amsterdam) and Emiliano Treré (Cardiff University).

BigDataSur is a collective effort to interrogate 'Big Data from the South', moving beyond the Western centrism and 'digital universalism' (Chan, 2013) of much of the critical scholarship on datafication, digitalization and Artificial Intelligence. The research workshop allowed the BigDataSur network to advance in charting its field of inquiry, including in the conversation activists and practitioners from various corners of the globe as well as scholars from disciplines as diverse as media studies, development studies, law, globalization studies, philosophy, science and technology studies, critical data studies—and counting.

In the context of BigDataSur,

> we maintain a flexible, expansive, and plural definition of the South(s) that casts it as a place of (and a proxy for) alterity, resistance, subversion, and creativity, embracing the dynamism and the multiplicity of interpretations while going beyond the geopolitical denomination. On the basis of this definition, which goes beyond the restrictive notion of the 'Global South,' we can find countless Souths also in what would normally be inscribed under the West, as long as people suffer discrimination, and/or enact resistance to injustice and oppression and fight for better life conditions against the impending data capitalism. With our definition, we engage in an exercise of disaggregation of the geographical dimension: while, on one hand, geography per se loses centrality in favor of a 'broader' and more imaginative definition of the South as also metaphor and proxy, on the other hand, we recenter geography, recognizing the differences it harbors and especially the shaping roles of power inequalities, old and new, still visible today in, for example, digital and data infrastructure. Simultaneously, we want to avoid the temptation to romanticize the South and reify the North (Milan & Treré, 2019, p. 325).

Methodology

The immersive research workshop consisted of working-group sessions designed for participants to engage in 'blue sky thinking' towards a conceptual agenda on 'Big Data from the South'. The methodology, devised by Stefania Milan, aimed at promoting collective thinking and cross-fertilization of ideas and distinct disciplinary perspectives.

Participants were invited to address **five thematic areas**, namely
1. data colonialism/decolonizing data
2. gendering data and feminism(s)
3. infrastructure and materiality
4. agency, activism and practices
5. methods, ontologies, epistemologies.

Each thematic area featured one or more assigned **facilitators** with subject matter expertise. While facilitators were assigned to a specific 'table', workshop participants covered all five thematic areas, rotating for a total of five slots of 45 minutes each. Participants reassembled in different configurations each time they joined a thematic table. Participants are listed on pages 23-29.

Each table was presented with a pre-defined set of **guiding questions**, as follows:
- What do we mean when we talk of [__]?
- How is the concept/understanding of [__] changing in the age of datafication?
- What can we learn by looking at datafication from a [__] perspective?
- What is the role of practices in the study of [__]? What practices should be taken into account?
- What are the main obstacles in adopting/promoting the [__] conceptual lens?
- What type of dialogue should we promote in our work?

In addition, each thematic area started with two sample questions, which allowed participants to situate the topic in relation to the overarching theme, as follows:

#1 Data Colonialism/Decolonizing Data (facilitators: Nick Couldry, Monika Halkort and Ulises A. Meijas)
- What are the differences and similarities between traditional forms of colonialism and data colonialism?
- What do we mean by decolonizing data?

#2 Gendering Data and Feminism(s) (facilitators: Payal Arora and Anja Kovacs)
- What are the main concepts and debates we should look at when dealing with gendering data?
- What is the added value of framing the debate in terms of feminism?

#3 Infrastructure and Materiality (facilitators: Anita Say Chan. ad Niels ten Oever)
- What do we mean when we talk of the materiality of data?
- What infrastructure should we addresses when studying Big Data from the South?

#4 Agency, Activism and Practices (facilitator: Guillén Torres)
- How is agency redefined in the datafied society?
- What are the main obstacles of data agency in the [Global] South(s)?

#5 Methods, Ontologies, Epistemologies (facilitators: Anna Berti Suman and Kersti R. Wissenbach)
- What does it mean to do 'engaged research'[2] with data?
- What does it mean to do 'engaged research' in the South(s)? Which steps should we take?

This working paper summarizes the conversations prompted by the event. It has been edited by the event organizers, with distinct sections being compiled by the facilitators of the thematic tables, when available, or by the editors on the basis of the minutes.

#1 Data Colonialism/Decolonizing Data
Emiliano Treré (Cardiff University)

Participants of the research workshop underline the need to situate data colonialism within the various local contexts and historical trajectories in relation to more traditional forms of colonialism. Hence, the **language we use** to address these dynamics is fundamental: What kind of metaphors and visions do we need to devise to address these kinds of processes? For instance, is data the new 'English', that is the master's language and therefore the master's infrastructure too? And if so, how do colonial subjects and multiple colonial subjectivities relate to this language? Do they need the master's language to think and un-think their subjectivity within this context? But before posing these questions, perhaps we should take a step back and ask who/what counts as a **colonial subject** in datafication.

In order to think about our **local contexts** and inquire how data colonialism unfolds differently at the borders and the margins, we need to investigate what we mean and how we articulate 'data' in our 'own' specific languages. Such understandings are localized and do not always translate to larger contexts and have different pathways of adoption. Therefore, it is key to think about vulnerabilities and resistance, different ways of conceptualizing vulnerabilities, and how we envision and enact **resistance**. Resistance is possible and necessary, both from inside and outside, but building solidarity is difficult when the condition of data colonialism is so subjective. Besides, not all colonial subjects are subordinated in the same way. Moreover, the colonizer can be more dehumanized than the colonized.

The **role of education** is pivotal in promoting diversity of views and radical alternatives. However, within our current neoliberal universities and 'standard' institutions, it is increasingly hard to promote and nurture forms of resistance to the current data regime.

Policy, regulation, as well as **ethical frameworks**, for instance in relation to artificial intelligence, are mainly limited to perspectives from the US and the EU. How can we include the rest of the world in these debates? For instance, how can perspectives from the global South, such as indigenous knowledge practices, inform research on algorithmic biases? Or, as another example, what are the consequences of the Chinese infrastructure and its deployment in Africa? We need to shed light on the cultural clashes between infrastructural development and the needs and interests of people, highlighting the exploitative extraction of resources and the fact that the benefits often go outside the place of the original investment. In dealing with this scenario, we should carefully think about the forms of repression that are likely to emerge in the near future, as well in 20 or 30 years. Consider, for example, Big Data in Latin America. Big Data has not only economic, but also real political impacts, as illustrated by its role within the Cambridge Analytica scandal during political elections, or the massive deployment of bots and political uses of WhatsApp within discussions on migration. Big Data has invaded Latin America without being properly understood, leading to questions about managing societal impacts and new forms of exploitation. Data colonialism is giving rise to a new social order currently unfolding in Latin America which can be observed in new reconfigurations of politics.

At the same time, digital technologies such as end-to-end encryption may afford new capabilities that allow for reclaiming some control over the **infrastructures** of 'civil space'. Hence, infrastructure can constitute a space of resistance, but how can we reclaim civic spaces within public and private platforms? Can we trust platforms not to share data with governments or corporations? As privacy seems a luxury good in the contemporary scenario, shall we accept the lack of privacy as a new reality?

The key point is we can only engage in a process of **decolonization** if we are able to disentangle how colonialism (in relation to data) operates, including the relationship between colonizer and colonized, the role of the state, of the government, of finance, of the military, the power imbalances, and the

specific forms of oppression and subalternity it engenders. As (de)colonization are historical processes (Fanon, 1963), datafication should similarly be understood in relation to historical processes.

In relation to this last point, many questions arise. Who is the colonizer and who is the colonized? Is Europe at the receiving end? What about our minds, are they colonized? What kind of order is generated through this process? Perhaps **colonial power** is much **more diffuse** than it used to be and as such is harder to see and to critique nowadays. Corporations are colonizing the public sector (education, health) through datafication, they are no longer just media companies. This is something that goes beyond surveillance capitalism. Data colonialism is of a larger order and scale, highlighting how new ways of understanding the world are fused with new forms of capital accumulation. It works at both the material level and at the level of discourse, language, and as a mythology. The real struggle is how we can describe this newly configured world to allow for alternative ways of knowing.

#2 Gendering Data & Feminisms
Zhen Ye & Stefania Milan (University of Amsterdam)

Feminism(s) vs. Gender
How does feminism challenge current discourses on bid data and datafication? Feminism provides a useful lens to look at data and datafication-related phenomena from a Southern perspective. To be productive, however, feminism should be tied to and grounded on specific issues to fully understand the problem at hand, because we ought to pay close attention to the context in which social phenomena occur. We should also consider the context in which data are generated, including algorithms and the surrounding economy, starting from the close-up and the detailed to move to the broader societal context.

As an interpretive lens, feminism is more productive than gender, as it offers **a way to interpret and read power and power relationship**, while challenging the status quo. Gendering, on the other hand, might reproduce the stereotypes, biases and assumptions associated with it. In this sense, feminism includes an embedded notion of activism as a strategy of change, whereas gender is more tied to a set of analytical approaches and tools, which could reproduce the existing state of affairs. Feminism, as activism, allows, for example, to present specific narratives of inequality in terms of gender differences and contextualize these narratives in different layers. Simultaneously, however, we ought to ask whether feminism, too, can be exclusionary (e.g., because of the excess of jargon and the over-reliance on insider knowledge). We also should interrogate the language used in the study of communicative practices. A number of distinct and complex questions arise. For instance, how do the existing power relationships work? Should we create a relationship between gendering and feminism? How can this discourse contribute to communication of complex theories and formation of analytical tools? How can we reconcile the tension between the activist discourse and its simplification and generalization with the scholarly discourse?

Feminism in Data and Tech
What do we mean when we talk about 'gendering data'? At the outset, when discussing Big Data and datafication generally, we should establish a conceptual link between feminism and social transformation. It is also key to bring an **intersectional perspective** to the debate on datafication, asking questions such as: who does it benefit, which relations of power/oppression emerge from it?

Is it possible to encode feminist values in, for example, algorithms? This question counters the argument that technology is 'neutral' and provides 'a neutral ground' to social interactions. Moreover, current mainstream practices intervene in the normative discourse, too. A case in point is represented by apps for 'women's health', which cater to particular 'kinds' of women only. In India, for example, a

certain health app required women to select symptoms, but the lists of available options did not represent the actual experiences of women. Part of the challenge was that these women struggled to express exactly what their problem was, for instance because cultural taboos. Other examples include sexual violence apps, safety apps and fertility apps. A feminist approach can help to question the representation of bodies in data and digital technologies. Should we seek to 'gender' data, platforms and the representation of contemporary issues or concentrate instead on confining women's concerns to traditional paradigms? Furthermore, we should also consider the fact that apps are often the equivalent of a 'digital leash', enabling new forms of surveillance.[3] How do these new technologies intersect with, and reinforce, oppressive family relations, for example? Other apps are similarly problematic, such as those tracking menstrual cycles. These apps typically visualize pregnancy and fertility in a teleological way. How do these technologies construct the female body, for example by evoking 'rosy backgrounds' and 'optimal days to get pregnant'? One of the key questions concern who is behind the construction of these mobile applications?

The dominant narrative usually pictures women as excluded from digital empowerment tools. But actually, women themselves may choose not to use the 'tools of empowerment' that society imposes on them.[4] Therefore, we need to be careful of technological determinism, which draws clear distinctions between digital empowerment and gender inclusion. Decoupling such assumptions is of fundamental importance in terms of research approaches because the traditional marginalization could be seen as an opportunity for activism. Relatedly, we should critically reflect on the power relations involved in the archiving of activism. Archiving involves collective activities within and beyond organizations to provide records and resources to bring about social justice and changes. As another example, the imprisonment of activists in various parts of the world requires us to reflect on the wider social environment and how to work for justice in such contexts.

Accessibility and (Dis)Empowerment
To discuss the power relations and technology with a feminist lens, we need to consider the accessibility of the Internet in today's societies. At the most basic level, some ethnic minorities may not have valid access to the Internet. Moreover, the so-called digital divide is still highly relevant to the gender and technology debate, for instance due to the lack of sufficient information on utilizing the Internet. For instance, indigenous women in Colombia are contributing to public and social projects, aiming to facilitate communication among indigenous communities and enhance the voices and participation of women. These women construct collaborative knowledge by embodying themselves as the 'informational nodes' of their respective families, actively sharing and formulating knowledge with others.

Historically, the role of women in society has been framed in sexist and gendered terms. For example. it is not until recently that India has passed the adultery law and decriminalized adultery. It is in the development of technology and the digital practices that the duality of such issues is reflected. On the one hand, digital media allows for new forms of surveillance, as communication is archived and used as legal evidence to protect and support women. On the other hand, digital media failed to improve legal frameworks in specific contexts, for instance due to the stigma resulting from videos of gang rapes. In societies where the gendered culture of shaming is more powerful than justice, digital tools may result in disempowerment of the women (e.g. sexualized blackmail or revenge porn). We ought not neglect the deadly consequences for victims in terms of their bodies, their mental states, and their social lives. We need to take seriously the urgent call to de-romanticize locality, especially when the locality is intertwined with patriarchal norms: because it is 'local' does not mean it is inherently good. The embodiments and interactions of women call for reconsideration of the relationship between gender and technology.

De/Gender Data

The starting point for reflection on the importance of de-gendering data is to bring the body back into debates about data. This is powerfully illustrated by discussions about health care and the digital representation of women's bodies, such as the algorithmic gender bias in health and social care. We need to reflect on whether certain platforms serve men and women in the same way. If not, what are the different implications for men and women participants? For instance, the misogyny prevalent in online fora such as 4-chan requires us to consider how anonymity facilitates misogyny. At the same time, direct action and civil disobedience may require a publicness that puts some groups at greater risk than others. Such publicity might in turn make them vulnerable as targets. It is always the case that we address people collectively, but we rarely take individual subjects that need to be protected from being discriminated against on the collective level into consideration (cf. 'group privacy').

Regarding research, we notice that the data used for engineering research is normally associated masculinity, while the data used for emotional sensitive research are typically associated with femininity. How do such associations interact with the design and outcomes of research? How can we address such biases? Affect might offer us another way of thinking. Translating feminism into affects allows us to talk about the wider breadth of human experience, instead of the binary feminine and masculine.

Universalism vs. Local Values

From the perspective of engineering, feminist theories, too, have a contribution to offer. In the realm of Artificial Intelligence (AI), among others, data is closely related to the way we conceptualize ourselves as human beings. The resistance to it happens in a clash, where the clash becomes proof of visibility and opens up the possibility of intervention. Instead of the western ways of talking about our values as relative (e.g. religions position women behind men), can we talk about the relative universality of human rights and the plurality of human values? Engineers should tailor to the needs of mapping out these values. Using feminism as a way to read power means that the goal is to have equality. While power plays locally, questioning power is universal, and by questioning power, different values can be brought to effect.

In establishing values, context matters. It is important to think about who are the interpreters of universal values and local contexts and how we are shaping those interpreters. Translation is key to answering this question. Although some argue that algorithms are merely functioning as neutral translations, we should not neglect the fact that values are translated into codes—the debate about making AI more ethical is about the reconfiguration of human values. Ethical concerns necessitate better datafication practices. Yet, there is a mismatch between fairness and algorithmic society (boyd, 2017). Defining fairness is a political act. The suggestion of universal freedoms or universal values can be recognized in the discourse of Silicon Valley, where, for instance, Microsoft has formulated a new research agenda on gender, race and power in AI (i.e. AI Know). Indian tech companies also adapted the discourse of universal values and dismantlement of power relations. Thus, claims to universal values and the concept of digitalization in effect serve the purpose of centralization.

Race and Decolonization

It is not possible to use decolonial terminology without putting race on the agenda, where it offers possibilities to intersect with feminism. A powerful example is the book *Algorithms of oppression*, which provides ways of thinking about algorithms, data discrimination and questions of representation (Noble, 2018). The way Google represents black girls as pornographic objects is associated with the colonial histories where blackness was seen through a narrative of sexuality, a racist projection of animalistic visions of sexuality. Nowadays, colonial legacies are turned into algorithms creating certain representations of reality. Similarly, search results in Google Images for the terms 'Muslim women' in Google Image are also stereotypical. Such stereotypes are perpetuated

from western imaginaries and are reproduced as algorithmic outcomes. As a result, they lead to the 'colonization' of non-western imaginaries.

Despite claims of benefiting universal freedoms, digital environments are not an inherently free place for feminist and decolonial perspectives. For instance, a media project in Mexico once wanted to include YouTube videos about Latin American feminism and traditional knowledge, but related search results were full of misogynist content. In online public spaces, feminism sometimes is related to anti-feminist content (e.g. trolling and harassment). We need intermediate spaces, that are safe and encourage women to participate. Indeed, there are similarities between online and offline spaces: specific discrimination offline is also reflected online and reinforces what is happening in our life. There are examples how online environments can provide safe spaces for people who do not have access to physical public space (e.g. the Indian queer movement that started online in the 90s). However, we also note trends of fragmentation in terms of different platforms providing niche services (e.g. specific dating apps for Muslims). How to position feminist apps? The first message to convey is that these apps are for women, but of course we can repurpose these apps or platforms for different values.

Activism and the Everyday
Capitalism has the capacity to satisfy desires in a way that is affectively and emotionally supplicated, while many alternative approaches, such as resistance and activism, tend to neglect the affective or sensual dimension of capitalism. Is there a way to rethink capitalism for activist purposes? For instance, how can a hacker group achieve its goal of attracting people by proposing alternatives that are not appealing at all? From this perspective, queer collectives are doing very well: they reject an exclusive focus on development, reclaiming the right to pleasure and the importance of romance. Pleasure is a privilege, because poverty reproduces a 'leisure divide'. In a non-western context, the Internet is considered more serious, for learning or improving skills, but in the majority of cases, the use of Internet is associated with fun.

Thus, we need consider how politics and lifestyles intersect in digital spaces. For instance, Instagram introduces problems of depression and anorexia, as well as the issues of self-branding into our lives. However, in some regions, such as Thailand and Indonesia, Instagram is not about self-expression but more related to reclaiming online spaces and the creation of informal market relations.

#3 Infrastructure and Materiality
Niels ten Oever (University of Amsterdam)

The Goal of the Discussion
Infrastructures structure ways of knowing, feeling, living, learning, governing, and being governed. Through their materiality, they both inhibit and enable certain behaviors. They reflect both past and current power structures, providing a lens to reflect on relations of repression and emancipation. Through the discussions in this working group we have explored how infrastructures can he designed, shaped, and researched in progressive ways. We situated the discussion between critical feminist, de/post-colonial, and human rights-based approaches in order to interrogate the relation between interoperation and sovereignty, independence and communication, access and control.

Colonial Infrastructures
If one observes Internet infrastructures in Africa, and the traffic flows they enable, one cannot but be struck by how much of the African traffic flows directly from and to Europe and the United States. This resembles the flow of raw resources extracted in Africa, only to be shipped to other continents for processing and sale at a high mark-up. Similarly, the process of value creation on the Internet, through

the gathering, processing, and re-selling of personal data, takes place outside of Africa, based on raw data that is gathered on the continent. This can be attributed to a lack of data centers in Africa, regional politics that problematize regional peering, and digitally literate diasporas. Next to that, privacy awareness in Africa generally is very low. An important exception to this case is Rwanda, which highlights how historical experience are reflected in people's relation to infrastructure.

Decolonizing infrastructures
We start from Fanon's premise that decolonization is process (Fanon, 1963). One cannot rebuild a city while people are living in it, because this would also erase the history of the people in it. One needs to acknowledge history and struggles in and over infrastructure. The only approach to decolonizing infrastructure is not to ignore the colonized roots, but see can be done with them, and thus use them as a means for reflection. This perspective also prevents essentialization of a specific version of perfect infrastructure and allows for a productive tension between values in the technical, governance, economic, and cultural realm. In order to build on this tension, one should also consider the breakdown and disruption of infrastructures, instead of defaulting to a paradigm of hyper-connection, with a bias for optimization, efficiency, and economic development. To protect a community, one should also be able to resist interoperability, to have a politics of incomparability and compatibility, to reject adopt other people's standards, nor be forced to document one's own indigenous standards for the benefit of others. There are many ways of decolonizing infrastructure, but the process is always both material and relational. Rather than a static artifact, it is a journey that goes against the grain of a historical process of oppression. The decolonizing of infrastructure is a pedagogy that charts a trajectory to the future.

Feminist Infrastructures
Similarly, feminism allows us to critically reflect and rethink traditional infrastructures, emphasizing the plurality of ways of thinking, constructing alternative realities and power structures. Feminism and infrastructure studies both make invisible work, such as materiality and maintenance, visible. Another point of analytical convergence is that labor is disproportionally done by people of color and women. To understand these relations, research needs to take into account historical and geographical contexts. If one creates knowledge about infrastructure, it needs to be situated in how it is locally involved/has evolved in order to foreground how it plays a role in or reflects a web of control and domination. This exercise allows us to decentralize the notion of the liberal givenness and recognize power relations in infrastructures, to ensure we do not repeat the past and reproduce the given.

Human Rights and Infrastructures
Universal values, such as human rights, can help to set standards for infrastructures, but these can only be understood and analyzed in specific contexts. The impacts of an infrastructure should always be analyzed from the perspective of the rights-holder, not from the perspective of the infrastructure. One needs to be able to reflect on symbolic elements carried beyond the materiality of infrastructure, which itself reflects, and sets, a standard. Implicit in infrastructure is that it allows for thinking about all kinds of scales and processes. Rather than all-encompassing, infrastructures should also accommodate smaller forms of association, and non-participation, otherwise it becomes a repressive social reality from above. At their best, infrastructures are an expression of collectivity, and collective agency: and can be understood as a collective form of expression. On the other hand, one cannot avoid wondering about individual ability to exercise their rights, because the expression of collectivity itself can become repressive. In that case, an individual should be able to disengage, but this should not have significant consequences. This makes equality in a time of interconnection and interoperability very hard - because those who disconnect do not benefit from the (imagined) connection-dividend, total autonomy, or autarky, should not be over-romanticized. The interesting tension, and challenge, is to facilitate and accommodate for local practices and in transnational infrastructures.

#4 Agency, Activism, and Practices
Guillén Torres (University of Amsterdam)

The Goal of the Discussion
The discussion in this working group was oriented towards figuring out how agency, practices, and activism are redefined within datafication. It emphasized the importance of people, given that much work has already been done in relation to the role of technology, and particularly in relation to the perils of automation. On the face of this apparently unavoidable displacement of agency towards non-governable autonomous technologies, we chose to discuss instead in what sense people and their agency have more importance now than ever. To achieve this goal, we considered it necessary to find out how to decenter the interest in the 'technical' within the 'sociotechnical', as well as locating new methodologies to research algorithms, datafication, and automated technologies, which emphasize people's involvement in the process.

To locate the discussion theoretically we chose to approach agency from three different perspectives: agency in relation to a particular subject, agency in relation to its context, and agency in relation to specific technologies. However, throughout the discussion it became evident that subjects, contexts, and technologies are difficult to analyze in disconnection from one another. Thus, the categories of 'subjects', 'contexts' and 'technologies' could be thought just as different viewpoints to choose from. In each of them, a specific element is emphasized in order to highlight its relative influence/relevance and its specific concerns.

Agency and Subjects
The question of what agency is, brought us to reflect upon ways of acting: is there a way of doing activism that everybody agrees upon? After so many years of working and sharing best practices, are we doing activism in the same way as 20 years ago? While trying to find an answer we noticed that doing activism is also impacted by processes of self-definition as well as by the perception of others about our agency. In some cases, being called an activist might imply compliance with a specific discourse, mobilizing specific technologies, or reproducing a few recognized practices. However, this might become problematic due to its normativity. Therefore, to avoid contributing to build a normative notion of what activism is by reaching a definition or a specific set of features that constitute activism, we rather reflected upon the necessity to remain open to recognize alternative ways of being an activist, and more importantly, only employing the label in those cases in which people actually want to be recognized as such.

The discussion around ways of acting also touched upon whether agency in the current context of datafication necessarily implies or is equal to resistance, and what connection it holds with the idea of social change. Unfortunately, certain features of activism make difficult the construction of narratives and concepts to foster resistance. For example, activists' goals are often project driven, which makes difficult both considering long-term scenarios and having a historical grounding. In fact, in many cases we could talk about re-activism rather than activism. This has been intensified by to the specific context of datafication, since some of the referents used by previous activists are missing; old activists used to struggle with the state, but the counterparts now are multifarious and their responsibilities blurry. Thus, we also identified a need to locate the relation between agency and history. In many cases activism reacts to daily problems without noticing there is a continuation of historical problems acting upon the here and now. In order to locate the continuum agency-resistance, it was proposed to pay attention to mundane practices that could amount to small, aggregated change in the long run.

Agency and Context

Apart from ways of acting, we considered it necessary to discuss the spaces in which to act. This process may go through a sort of inventory of the spaces we have left, those which need to be created or gained and those which need to be defended to keep them accessible. In fact, in some cases it is also necessary to remain open to finding spaces in which we would not expect agency could be exercised. This is the case, for example, for platforms that in the global North have already been extremely delegitimized, but in the South still have contentious potential despite their clear flaws.

One problem regarding the location of agency is that, in the global South, datafication seems very often extremely far away, at least initially. Isolated communities who have problems accessing basic resources might not be immediately aware of how datafication (or the lack thereof) influences their lives. A specific concern arisen from the discussions in the working groups was the possibility that this perception of distance to datafication is related to the lack of first-hand experience with quantification or with technology in general. Many people are just barely consciously entering the 'digital', although that does not mean that they have not been affected already by what happens *there*, if we continue the spatial metaphor. In that sense, it may be necessary to nuance the common critique that the current data-hype makes people think that everything can be solved with data. Yes, indeed, not all the social problems can be solved by implementing some kind of data-driven solution, but this becomes even worse if we only consider certain things as data and data-practices. Framing a problem as a data-problem might be prejudicial for communities if we ignore their own view on what data and data practices are. If external ways of working with data are imposed, that will probably create more problems rather than solving them. As a guide to identify relevant data-practices within communities, the group thought of, for example, researching through analogies such as the management of cattle as a sort of database management, or the searching for water as a mapping effort. Building upon existing practices was seen as a much better way than imposing solutions by introducing external mechanisms.

While reflecting upon the process of knowledge exchange, we started to consider that perhaps a common lexicon is necessary to set the basis for a conversation among different fields, such as ICT for development/social change, critical data studies, digital activism, media studies, etc. And perhaps one of the first terms to be commonly defined with input from all of these closely interrelated fields is that of activism. As many of the attendees suggested, it is easy to get caught in tropes of what we consider activism to be, and consequently what we think of as agency. This may result in us ignoring everyday acts and decisions that people understand as activism –and thus examples of agency-, since we are focused on finding things that fit our previous assumptions. In academia, it is important to recognize that we often register certain practices and miss others, which is connected to our own convictions and the fitness of our approaches. Thus, a major task to undertake is defining what we think activism means in practice within the specific domain of datafication. This endeavor needs to take into account the fact that the experience of datafication is very different depending on locality. For example, while some people experience datafication in a context of human rights void, others have some defensive tools at their disposal.

Paying attention to the local might require specific redefinitions of what data and datafication are and how do they work in different localities. Particularly in relation to how this is experienced in the global South, we focused on the fact that data does not necessarily need to be connected to high tech. To have a specifically Southern perspective it is necessary to look at what people in the South do to solve their problems producing, processing and repurposing data. Therefore, what constitutes agency depends on this redefinition.

Considering the local level when thinking about agency and datafication is crucial, and it signals the need to reflect upon the connection between agency and structure. For example, the practices of civil

society throughout Latin America tend to be influenced by the expectations of the founders and the collective construction of the discourse of ICT for empowerment. This makes certain matters appear more relevant than others given the attention they receive from local organizations, but this may be more connected to the ideas local NGOs have about what funders would fund, than with the real needs of the local population.

The relevance of the local does not necessarily imply that thinking (from the South) about agency within a context of datafication has to forgo concerns about the global. The idea is rather to not lose touch with the local origins of social problems and find out how different layers pile up one after the other, tracing different struggles through different scales.

Agency and Technology

Given these differentiated experiences of the influence of technology over social life, it is necessary to also consider that some communities may not see technology as a tool for activism and empowerment, but rather as a tool for domination and surveillance. An interesting example of how datafication is experienced differently in different places is the idea, relatively widespread in developed areas of the world, that the outcome of datafication is an accurate representation of who we are, particularly in relation to individuals' likes and dislikes when it comes to marketing. Both corporations and the public discourse tend to represent the picture emerging from our online habits as an accurate representation of our interests. However, in the global South, where it is usual to experience faulty technologies regardless of who is in control of them, there is no expectation of accuracy and there is a certain disbelief about technologies being able to 'datafy' people correctly. This non-working datafication may have terrible consequences, such as in India or Chile where the collection of biometric information is connected to the provision of public services.

Agency and Big Data from the South

This topic was approached from the need to define what is the most urgent aspect to tackle when thinking about datafication from the South: should we emphasize the role of infrastructure (or the need to build grassroots infrastructure)? Or should we talk about how the contexts of the South influences the very specific ways in which datafication occurs? A suggestion coming out of one of the working groups was to perhaps think of agency from a specific position in relation to those main concerns: what is agency in relation to infrastructure/context and how has it changed/been displaced?

A main obstacle brought up by the participants is that most infrastructure has been built by the west, and that efforts to de-colonize are now also becoming a sort of colonization, for example China's success in building all kinds of infrastructures different countries around the African continent.

The relationship between agency and technology was of interest in terms of activist's experiences of technology, that is, what is the affective dimension of being an activist in the time of datafication and what is the role of emotions within their work. This implies, for example, finding out what emotions are triggered in activists when they tackle issues and how these emotions relate to their people's agency.

#5 Methods, Ontologies, and Epistemologies
Anna Berti Suman (Tilburg University) & Kersti Wissenbach (University of Amsterdam)

The Goal of the Discussion and Our Approach

Our discussion mainly aimed at disentangling questions related to methods, ontologies, and epistemologies in our practices of research and diverse forms of engagement in data and activism (and adjacent fields). In approaching the topic, we decided both to address the issues from our

disciplinary standpoints, and to discuss them from a cross-discipline perspective. In addition, considering the 'applied' nature of the discussion group, we welcomed experiences from practice. We adopted both a 'reflexive' lens, discussing the topic by looking back to our stories of research and engagement, and a 'predictive' lens, looking ahead to our future challenges and formulating expectations on how practices in the study of data and activism will evolve in response. We also adopted a 'normative' standpoint at instances, inquiring how we as researchers, but also engaged members of society, *should* face the questions arising in our research arena.

Among the discussed topics, we identified seven themes: power structures inherent to the academic environment and our role in the shaping of power dynamics through our approaches to research; embracing relativism and engaging with 'bad data practices', including on the resistance side, in order to overcome the 'big good' versus 'big evil' lens; acknowledging that grasping the ''many Souths' cannot happen as long as we stick to practices of pre-definitions; as a consequence, exploring the need to embrace typologies and categories as results, not as points of departure; the relevance of embracing failures and biases in our research, yet trying to avoid data biases; striving to overcome disciplinary 'siloing' in order to grasp realities more thoroughly; and, lastly, the need to wonder whether data (activist) research is currently in a bubble and whether there are adjacent activist/research groups at (higher) risk.

In dealing with the dark side of data practices, we discussed how some of these bad practices are the result of power imbalances. We wondered what our role is in reshuffling power dynamics in academia that lead to 'bad data practices', for example in terms of *who* gets the chance to define 'the Souths'. We are facing the continuous (also academic) tendency of 'the North' defining 'the South'. In doing so we may miss important nuances. We rather have to take an 'experiential' approach by looking at (daily) data practices and at the communities around them to understand the roots of 'the Souths' therein. Such 'Souths' will differ and show power dynamics that are contextual to each unique case. Our responsibility as researchers *and* as teachers is to contrast and dismantle such power dynamics, both inside and outside academia.

The themes illustrated here are not to be seen as disconnected discussions but are intertwined in many layers and ways. In what follows we will briefly elaborate on the conversations we had around each of those themes, suggesting what emerged from the vibrant conversations, and what we would consider the ways forward towards meaningful data activist research and practice.

(Power) Structures Within Academia and Our 'Role' in Shaping Power Dynamics Through Our Research
Pretty much right at the start of our conversations we turned 'inwards' by recognizing the central role we, as members of academic institutions, play in shaping power dynamics through our research. This relates to power (im)balances within academia, related to different funding structures and subsequent opportunities of meaningful research and voice. It also very much relates to aspects such as the enduring dominance of English language within the academic sphere. We are emphasizing the need for more awareness of such, and other, power dynamics and how they build patterns of exclusion into our 'resistance work' as research activists.

Our discussion also addressed power dynamics between different sources of information, embracing the relevance of certain knowledge materials produced outside academia, such as NGO reports. The recognition of failure when working with tech, such as online and mobile platforms, has been emphasized in the practical field, within organized civil society. Academia would benefit from taking such knowledge into account, which can be found in public (online) discussions, conferences, grey literature, newspapers, reports, and other fora. We acknowledged the need to recognize those sources as valid bases of knowledge.

This also directly brought us to the discussion about the need to break with the conventional, unidirectional flows within the academic (research) system. We perceive a persistent and widespread assumption that the researcher has a privileged vantage point, coming 'from somewhere else'. We critically discussed the idea that the 'distant' researcher, even if going into the field, can be the right person to study distant dynamics. We critically inspected the fact that local civil society actors have their own positioning and perception with regard to the dynamics of the 'outsider' coming in. A participant brought up the example of Asian civil society getting increasingly fed up with people coming from Europe/US and concluding that people across Asia may have a very different understanding of and attitude towards privacy.

We then discussed the requirements for moving beyond the 'outsider' coming in, which has much to do with how we build relationships with our fields of study, places and people as well as well as how we justify it. In deciding on how to write 'about people', as data activists and researchers we may want to strive as much as possible to 'let people speak for themselves' but also accept that we might not always been able to fully and really understand certain cultures, and be transparent about such limitations.

Moving from Extremism to Relativism
We tackled the need to go beyond a certain 'extremism' that we register in our field. We consider this important in order to overcome the 'big good' versus 'big evil' lens. As becomes clear from developments in the research field, we have to look for ways to question what we fit into the categories 'the good' and 'the bad' in order to make perceptions more critical and granular. In addition, the relativity of both categories should be acknowledged. Similar to acknowledging that there are many 'Souths', we need to recognize that there are many declinations of the good and the bad and having a 'doing good' agenda does not intrinsically mean that related actors cannot have questionable practices, particularly when discussing good and bad data practices. We cannot be so limited to think that the resistance is the equivalent to good data practices, also if those practices are not deliberate or conscious. As examples, we discussed the troubling data practices of many NGOs, exploitative data exploitation, and the use of software/collaborating/benefitting from software companies/social media actors known to act in unethical ways. We do find various levels of risk-awareness within the NGO environment about the risks associated with such practices.

Grasping the Many Souths Cannot Happen Through Predefinitions
Bringing up the event's main focus on the need for the recognition of many Souths within academia, and data activist research in particular, brought us back to our discussion about who has a voice in setting agendas. We consented that grasping the many Souths needs to not only move beyond traditional geographical divisions but also take a critical stance towards *who* defines the many Souths, the diverse realities, the challenging conditions, and overall, the roots of power dynamics. This stimulated us to express the need to *let* define the roots of power dynamics and doing so for every single research context. This, then, implies the need to move away from pre-definitions since grasping the many Souths from within exactly such multitude of contexts cannot emerge from pre-defined concepts. We therefore need to look at how data and data practices come into being and are shaped by a variety of power dynamics situated in contexts (think of actors, infrastructures, political, socio-cultural, economic, and so on). An example that was brought up is the case of Basilicata in the South of Italy: a post in the Big Data from the South blog demonstrates the many forms of injustice and exclusion harbored also in the 'geographically Northern' region (Berti Suman, 2018).

Over the course of the discussions, we also asked ourselves whether we are lacking focus by meshing together too many ontologies, making things difficult to grasp. We discussed the potential of other approach, such as trying to identify a meaning of Souths that is specific to data and encompasses the vulnerabilities and inequalities that are interrelated with data realities and data practices. These

considerations, and the need for a 'many Souths approach' when dealing with data, follow directly from the starting point that raw data is an oxymoron (Gitelman, 2013). Acknowledging this, we realized that we need 'accountable categorizations': new data-related methods have to account for *multiplicity.* One example we discussed related to the research of harms and how we make sure to grasp the real, local causes and challenges of sticking to data-driven methods of analysis, which are not capable of grasping the roots of the issue. Motives of why people act in certain ways, local contexts underlying certain behaviors *etc*, do require qualitative approaches in order to be surfaced.

Typologies/categories as result, not the point of departure
Limitations on grasping realities on the ground and/or the roots of issues we observe are not to be reduced to data-related methods, however. We also deepened our discussion around the need to distance ourselves from predefining categories. When predefining categories or typologies, we run the risk of trying to fit research observations and findings into those categories. Instead, when we enable the emergence of new categories from local actors and realities, these categories would be defined from the grassroots rather than from the academic ivory tower. Thus, whilst we do need epistemological categories, the question is who gets to define them. We agreed that a shared typology is important in order to compare and to organize knowledge. Putting more emphasis on the question of who has a say leads us to considering typology to be the result, rather than the driver. This would mean to not predefine how we build our analytical frameworks and to be conscious of assumptions we make about certain contexts.

Enabling space and agency for people to identify what is important to them, and collectively unpacking the roots of certain power dynamics, is a participatory process, as suggested before. We can draw inspiration from Fals-Borda's participatory action research (Fals-Borda, 1987) and DATACTIVE's work (Kazansky, Torres, van der Velden, Wissenbach, & Milan, 2019). Participatory approaches require us to reflect on what our presence as researchers does, no matter if we are critically aware. We need to understand different cultural contexts, e.g. with regards to how people engage or speak up, as well as the vastly diverging understandings of concepts (such as e.g. privacy or citizenship) that people can have. We need to go beyond established categories and assumptions in order to be able to observe and understand the experience of citizens, e.g. when they appropriate mainstream tech. In sum, designing research based on our (pre)conceptualizations can be problematic.

We need to embrace the need to learn and adjust, getting to the roots of certain practices or non-practices rather than remaining focused on symptoms. To do so requires the means, space, and freedom to adjust our research agendas along the journey. This is where we identified some dilemmas. The speed of academia does not allow us to deeply understand the field. Institutional decision-making and other formalities, such as funding or ethical review, typically demand predefined concepts. Furthermore, research is unpredictable to an extent because everyone has culturally specific histories and, thus, sensitivities to how we absorb realities, practices, et cetera.

Discussing Failures and Biases in Our Research
Whereas it seems imperative to address data biases in our research, it becomes more urgent every day to discuss openly the failures in our research (from making wrong hypothesis/research questions to applying wrong methods/wrongly applying good methods). Failures need to be embraced as this can strengthen our own research and overall research quality in the academic and non-academic worlds. Accepting the failure of a hypothesis, for example, can be also a way to re-valuate methods of social or human sciences vis-à-vis possible critiques. In addition, we have to acknowledge and be transparent about our own biases when we go to the field (and before/after). In going to the field, we need to strive to find roots of certain phenomena and understand the real problem(s), but often we look for the problems that we already expect. This is the result of our biases influencing our research approach. These biases can be of a different kind, like moral and political biases, or biases deriving

from implicit values held in certain research communities. Such biases may be unavoidable, differently from data biases, which is what we would want to reduce. As research on data and activism is highly context-driven, we need to pay attention to biases deriving from our perception of the context in which the research is situated. To address this risk, more 'small data' and context-aware research may be needed.

Overcoming Disciplinary Siloing to Better Grasp Reality: From Un-Disciplinary to Un-Theoretical Research?
When engaging in discussions about the need to not pre-define concepts, thus for more explorative and participatory approaches in order to grasp realities and avoid building bias into our research, we began to realize the dangers of disciplinary 'siloing'. If we, as data activist researchers, frame everything as datafication, we believe that we are missing out. How people make use of certain means, the motivations behind certain practices and non-practices are not data or infrastructure per se. The risk of silos is that they determine how we look at things. Therefore, we considered a move from un-disciplinary to un-theoretical research. We discussed whether we are freer in the contextual unpacking of the roots of power dynamics (as earlier elaborated) if we do not have to channel everything through particular theoretical lenses. Theorizing can happen in many ways and is *hegemonic* in itself. Every discipline hangs on to certain methods and approaches. Breaking through those and learn to grasp realities more thoroughly, choosing the methods best, not those most strongly related to our discipline, would mean to rethink the realms of theory as enabling participatory ways of unpacking reality.

This consideration also derived from discussion about the crucial role of not pre-defining concepts. This implied the need, as outlined before, of giving agency to local voices, and to avoid pre-defined analytical frames. However, being un-theoretical, then, does require that we have a shared foundation of understanding whilst being contextual and aiming at perceiving typologies as results rather than points of departure; a language that can also fit the 'terminology' of civil society in ways that convey the position of participants. Furthermore, it raises the need for a common set of outlets, physical and non-physical spaces, in order to bring these conversations forward. Our academic common set of outlets and language does not necessarily match those of civil society and participants. This raises some potential challenges to further explore.

Data (Activist) Research Is in a Bubble?
As researchers of data and activism, we sometimes risk being 'far' from the reality of activism where actors can be at (higher) risk. We should ask ourselves the question on whether we engage (sufficiently) with adjacent activist/research groups that could benefit from our knowledge and ideas when facing their challenges. We discuss 'data activism', but how often do we look at the actual threats that activists working on other topics, like environmental activists, are confronted with? Many of those actors are working with communities at the highest risk, and in the highest risk environments, without mastering necessary practices, such as data security.

Therefore, we need to inspect how activists are threatened by increasing means of surveillance. However, these questions are often not on the agenda of activists, for instance because of a lack of awareness or capacity. Which role do we have in shedding light on such threats? In addition, risks and coping strategies may depend to a large extent on what makes sense in the specific contexts of these activist worlds. We discussed the need to find a 'common denominator' to address these questions that maybe could be found in the *concern* rather than in specific methods, ontologies or epistemologies. We also realized that, although we need to somehow reach a 'common denominator', we do not have to fall into the risk of predefining a problem before experiencing it in the field: a listening attitude should be dominant, and each typology (of possible threats to activists, for example) should be the result rather than the starting point of inquiry.

Notes

[1] Visit https://data-activism.net/publications/big-data-from-the-south/.
[2] To know more about 'engaged research', see Milan (2010) and Milan & Milan (2016).
[3] See for example the Internet Democracy Project's research on safety apps in India. https://genderingsurveillance.internetdemocracy.in/safety-app/.
[4] See for example the initiatives against facial recognition, such as Crooks (2019).
[5] https://data-activism.net/2017/07/datactive-presents-big-data-from-the-south-in-cartagena-july-15/.
[6] Visit https://data-activism.net/2018/11/workshop-big-data-from-the-south-towards-a-research-agenda-amsterdam-december-4-5/.

Cited Works

Berti Suman, A. (2018). Civic resistance to environmental failures from the South (of the North...): The AnalyzeBasilicata initiative. *Big Data from the South: The Blog*, 14 December. Retrieved from https://data-activism.net/2018/12/civic-resistance-to-environmental-failures-from-the-south-of-the-north-the-analyzebasilicata-initiative/

boyd, danah. (2017). Toward Accountability: Data, Fairness, Algorithms, Consequences. *Data & Society: Points*, 12 April. Retrieved from https://points.datasociety.net/toward-accountability-6096e38878f0

Chan, A. S. (2013). *Networking Peripheries: Technological Futures and the Myth of Digital Universalism*. Cambridge: MIT Press.

Couldry, Nick, and Ulises Ali Mejias. 2019. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, California: Stanford University Press.

Crooks, R. (2019). What We Mean When We Say #AbolishBigData2019. *Medium*, 22 March. Retrieved from https://medium.com/@rncrooks/what-we-mean-when-we-say-abolishbigdata2019-d030799ab22e

Fals-Borda, O. (1987). The application of participatory action-research in Latin America. *International Sociology*, *2*(4), 329–347.

Fanon, F. (1963). *The Wretched of the Earth*. New York: Grove Press.

Gitelman, L. (2013). *"Raw Data" Is an Oxymoron*. Cambridge, MA: MIT Press.

Kazansky, B., Torres, G., van der Velden, L., Wissenbach, K., & Milan, S. (2019). Data for the social good: Toward a data-activist research agenda. In A. Daly, S. K. Devitt, & M. Mann (Eds.), *Good Data* (pp. 244–259). Amsterdam: Institute of Network Cultures.

Milan, S. (2010). Introduction: Toward an Epistemology of Engaged Research. *International Journal of Communication*, *4*, 856–858.

Milan, Chiara, and Stefania Milan. 2016. "Involving Communities as Skilled Learners: The STRAP Framework." In *Methodological Reflections on Researching Communication and Social Change*, 9–28. Cham: Springer.

Milan, S., & Treré, E. (2019). Big Data from the South(s): Beyond Data Universalism. *Television and New Media*, *20*(4), 319–335.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

The Big Data from the South Initiative

The Big Data from the South Initiative is an interdisciplinary space for theoretical and empirical exchange on the challenges of datafication, massive data collection and automation as they unfold in the plurality of South(s) inhabiting our increasingly complex world. It brings together researchers, practitioners and activists and engages in various activities, including events, presentations at conferences, and a dedicated multilingual blog.

The Big Data from the South Initiative was launched in July 2017 by Stefania Milan (University of Amsterdam) and Emiliano Treré (Cardiff University). Milan and Treré co-organized a one-day event in Cartagena, Colombia, as a pre-conference to the annual meeting of the International Association for Media and Communication Research (IAMCR). Entitled *Big Data from the South: From media to mediations, from datafication to data activism,* the event interrogated the mythology and universalism of datafication and big data from an epistemology of the South perspective and was attended by 50+ participants between academics, practitioners and activists.

The Big Data from the South Initiative has received funding by the European Research Council through the DATACTIVE project, the Amsterdam Center for Globalization Studies (ACGS), the Amsterdam School of Cultural Analysis (ASCA), and the Amsterdam Center for European Studies (ACES).

Events

- **Big Data from the South: From media to mediations, from datafication to data activism** (Cartagena, Colombia, 15 July 2017). The event was made possible by the funding of DATACTIVE and by the generous engagement of Guillén Torres (DATACTIVE) and *Fundación Karisma* (Bogotá, https://stats.karisma.org.co), and the assistance of the IAMCR local organizing committee.[5]
- **Three panels at the annual conference of the Latin American Studies Association** (Barcelona, Spain, 25 May 2018), organized by Anita Say Chan **(University of Illinois, Urbana-Champaign).**
- **Big Data from the South: Towards a Research Agenda,** a two-day workshop at the University of Amsterdam. The event was funded by the ACGS, ACES, ASCA and the European Research Council.[6]

Publications

- **'Big Data from the South', Special Issue of the peer-reviewed journal** *Television & New Media* (volume 20, issue 4, May 2019) edited by Stefania Milan and Emiliano Treré. The issue features articles by (in order of appearance) Stefania Milan (University of Amsterdam) and Emiliano Treré (Cardiff University), Nick Couldry (LSE) and Ulises A. Mejias (SUNY Oswego), Paola Ricaurte (Tecnológico de Monterrey), Payal Arora (Erasmus Rotterdam), Monique Mann and Angela Daly (Queensland University of Technology), Jean-Marie Chenou (Universidad de los Andes) and Carolina Cepeda-Másmela (Pontificia Universidad Javeriana), and a commentary by María Soledad Segura (Universidad Nacional de Córdoba) and Silvio Waisbord (The George Washington University). Visit https://journals.sagepub.com/toc/tvna/20/4.
- The **multilingual blog 'Big Data from the South'**, accessible at https://data-activism.net/big-data-from-the-south/.

The Immersive Research Workshop

The two-day research immersive workshop 'Big Data from the South: Towards a Research Agenda' took place at the University of Amsterdam on December 4-5, 2018.

Day #1 was a combination of two public events and a close informal session:

Open session: **Can Data be Decolonized? Data Relations and the Emerging Social Order of Capitalism,** with Nick Couldry (London School of Economics and Political Science) & Ulises A. Mejias (State University of New York at Oswego)

The talk examines how contemporary practices of data extraction and processing replicate colonial modes of exploitation. The concept of 'data colonialism' is introduced as a tool to analyze emerging forms of political control and economic dispossession. To that effect, the analysis engages the disciplines of critical political economy, sociology of media, and postcolonial science and technology studies to trace continuities from colonialism's historic appropriation of territories and material resources to the datafication of everyday life today. While the modes, intensities, scales and contexts of dispossession have changed, the underlying function remains the same: to acquire resources from which economic value can be extracted. Just as historic colonialism paved the way for industrial capitalism, this phase of colonialism prepares the way for a new economic order. In this context, the authors analyze the ideologies and rationalities through which 'data relations' (social relations conducted and organized via data processes) contribute to the capitalization of human life. Their findings hold important implications for how we study the internet, and how we may advocate for the decolonization of data in the future (cf. Couldry & Mejias, 2019).

Roundtable **Big Data from the South: Decolonization, Resistance and Creativity**, with Payal Arora (Erasmus University Rotterdam), Nick Couldry (London School of Economics), Merlyna Lim (Carleton University) and Ulises A. Mejias (State University of New York, College at Oswego). In collaboration with SPUI25 (spui25.nl). Chair: Stefania Milan (DATACTIVE, University of Amsterdam). Moderator: Emiliano Treré (Data Justice Lab, Cardiff University). The roundtable was followed by an invitation-only informal research session in collaboration with Terre Lente (terrelente.com).

Datafication has dramatically altered the way we understand the world around us. Understanding the so-called 'big data' means to explore the profound consequences of the computational turn, as well as the limitations, errors and biases that affect the gathering, interpretation and access to information on such a large scale. However, much of this critical scholarship has emerged along a Western axis ideally connecting Silicon Valley, Cambridge, MA and Northern Europe. What does it mean to think datafication from a Southern perspective? This roundtable interrogates the mythology and universalism of datafication and big data, moving beyond the Western centrism and 'digital universalism' (Say Chan, 2013) of the critical scholarship on datafication and digitalization. It asks how would datafication look like seen… 'upside down'? What problems should we address? What questions would we ask?

Day #2 consisted of working-group sessions designed for participants to think along towards a conceptual agenda on 'Big Data from the South'. It took place at the Department of Media Studies of the University of Amsterdam on December the 5th, 2018. The participants are listed below, in alphabetical order. The list of participants includes also the authors of the reports, listed however in the next session.

Participants

Carla Alvial Palavicin is a researcher at NUMIES, a research group in Santiago de Chile focusing on Sustainability Transitions. Her work relates to the governance of technologies, particularly the science/expert/politics interphase and forms of metrics/datafication/and valuation used by policy makers in the context of environmental technologies and environmental problems. She is also interested on speculation and imagination as tools to discuss the governance of emerging issues and large-scale problems, from climate change to data governance. She collaborates with Open Science and Open Knowledge groups in Chile. She holds a PhD on Governance of Knowledge from University of Twente (NL).

Keywords: Valuation, open knowledge, environmental datafication, data in policy making, citizens science.

Payal Arora is an Associate Professor at Erasmus University Rotterdam and Founder of Catalyst Lab, a digital activism organization. Her research focuses on digital cultures in the Global South. She is the author of several books including the upcoming, "The Next Billion Users: Digital Life beyond the West" with Harvard University Press. She sits on multiple boards including the Facebook Advisory Committee, Columbia University's Earth Institute
Connect to Learn, and The World Women Global Council in New York. She has held Fellow positions at NYU, GE, ITSRio and ZeMKI and is the Section Editor for the University of California Press journal - Global Perspectives.

Keywords: Digital cultures, privacy, data economies, poverty, inequality, Global South

Sérgio Barbosa holds a BA degree in Sociology (2014) and a *licenciatura* in Social Science from the University of Brasilia (2015). He has a MA degree in Political Sociology from the Federal University of Santa Catarina (2017). Currently, he is a PhD Candidate in 'Democracy in Twenty-First Century, at the Centre for Social Studies in partnership with the Faculty of Economics of the University of Coimbra, Portugal. His research explores the role of WhatsApp for activism and social mobilization.

Keywords: WhatsApp, digital activism, resistance

Davide Beraldo is a postdoc in the DATACTIVE project, University of Amsterdam. He is a sociologist with a computer science background, and his research focuses on the intersection between social movements and the digital. In his PhD, Davide explored Anonymous and Occupy with computational methods, and reflected on the capability of 'contentious brands' of this sort to act as a common interface for highly diverse (sometime contradictory) social movement issues. In relation to BDS, he is particularly interested in counter-epistemologies of big data, emancipatory data practices, algorithmic resistance, and issues of global political economy of data.

Keywords: Counter-epistemologies, political economy, algorithmic resistance, computational methods, social movements

Enrico Calandro's research explores the relationship between digital access and development with a focus on digital inequalities. Over the last nine years, his professional career has been research driven. He has conducted Internet policy and regulation research across the African continent to facilitate evidence-based and informed policy making for improved access, use and application of ICTs in Africa. His research has focused on the political economy underpinning telecommunications markets in developing countries context, with a focus on global and regional internet governance structures and processes.

Keywords: Research, digital access, inequality, development, cyberpolicy

Fabien Cante is a former team member of the DATACTIVE project, and currently a postdoc at the University of Birmingham. His interest in data from the South comes from three sources. First, from his research on media (old and new) as infrastructures of city life in Abidjan, Côte d'Ivoire. Fabien am particularly interested in how media configure the city as a space of livelihood, social connection and political agency. Second, he am interested in data because of the way it is implicated in tracking and inflicting violence on migrants. Third, he is interested in how social threats are used to justify and contest new modes of surveillance in the Global South.

Keywords: Infrastructural life, institutional violence, cities, race, ethnography

Bernardo Caycedo is a Ph.D. Researcher at the Department of Philosophy of the University of Amsterdam. His research explores new forms of political engagement related to the Internet (in particular, whistleblowing, hacktivism, and radical open access initiatives). Bernardo is part of the project *Transformations of Civil Disobedience: Democratization, Globalization, Digitalization*. He is a member of the Amsterdam School for Cultural Analysis, the Dutch Research School of Philosophy, and the Amsterdam Center for Globalization Studies. Bernardo obtained his MA in Philosophy from Universidad de los Andes, Bogotá, Colombia and his BA in Philosophy from Pontificia Universidad Javeriana, Bogotá.

Keywords: Hacktivism, surveillance, social movements, social justice

Alberto Cossu is a sociologist who does research at the intersection between digital media and activism, qualitative and digital methods, collaborative and digital economies. He holds a PhD from the University of Milan. Alberto conducted research on the mobilization of knowledge and art workers in Italy, on peer-to-peer models of organization and production in Italy and France, on digital economy and co-working spaces in Italy and Thailand. He co-edited the book 'artWORK: Art, Labour and Activism' (Rowman & Littlefield, 2018). He is co-founder of PLANK (Politically-Led Art and Networked Knowledge), a London-based international network of artists, activists and academics, and is part of Macao 'The New Center for Arts, Culture and Research' in Milano since 2012.

Keywords: Art activism, digital methods, collaborative work, p2p

Nick Couldry is Professor of Media, Communications and Social Theory at the Department of Media and Communications, London School of Economics. His work is on media power, voice and increasing questions of data power and data justice. He co-led the chapter on Media and Communications in the International Panel on Social Progress whose report is published in 2018.

Keywords: Data colonialism, data relations, freedom, social order

Álvaro Andrés Crovo Godoy is a law student at the University of Rosario, Colombia. He is also a member of the Centre of Internet & society (ISUR, University of Rosario). Two articles have been published by him: one about Smart Contracts, their opportunities and risks; and the other one, about the Colombian regulation of financial related data. His major interest is to learn and write about Internet Governance and, more specifically, about corporate surveillance and its harmful effects.

Keywords: Surveillance*,* Internet Governance, Regulation

Monika Halkort is Assistant Professor of Digital Media and Social Communication at the Lebanese American University in Beirut, Lebanon. Her research explores the intersectional dynamics of coloniality, racialization and enclosure in data infrastructures, focusing on the specific context of political struggle and humanitarian governance. Published works include a study on the biopolitics of datafication in Palestinian refugee camps and the geopolitics of mobile phone networks in the context of kidnapped Eritreans in Sinai. The main geographic focus of my work is the Arab world. Prior to my academic career I have worked as a journalist for public broadcasters in Austria, Germany and the UK.

Keywords: Datafication, ontologics of colonialiality, ecologies of struggle, global south

Becky Kazansky is a PhD candidate with the DATACTIVE project at UvA. She has a joint affiliation with the Media Studies and Political Science departments, additionally serving on the ethics committee of the Informatics Institute. Her research focuses on resistance to surveillance amidst datafication, looking specifically at defensive strategies (practices and infrastructures) that try to predict, preempt, and anticipate surveillance. Alongside her academic research, she has worked for a decade with different human rights and social justice organizations on technology related issues.

Keywords: Resistance, surveillance, STS, critical data studies

Anja Kovacs directs the Internet Democracy Project in India, which works towards an Internet that supports freedom of expression, democracy and social justice, in India and beyond. Her research focuses on questions regarding cybersecurity, surveillance and privacy, and regarding freedom of expression. This includes gender and online abuse, gender and surveillance, and bodies and dataveillance, both at the theoretical and applied level (as in, trying to understand what the theory could or should imply for, e.g., data protection legislation in the context of a country like India). Feminist theory and practice are enduring building blocks and inspirations in Anja's work. Anja holds a PhD in Development Studies from the University of East Anglia.

Keywords: Bodies, dataveillance, feminism, data justice, political economy

Merlyna Lim is a Canada Research Chair in Digital Media and Global Network Society and Associate Professor in the School of Journalism and Communication at Carleton University, Ottawa (Canada). Lim's research and publications revolve around the mutual shaping of technology and society, and political culture of technology, especially digital media and information technology, in relation to issues of power, justice/equality, democracy and citizen engagement.

Keywords: Power, global south, citizen participation, activism, social movements

Joan López is a privacy researcher at Fundación Karisma (Colombia) and a Master student of Sociology at the Bielefeld University (Germany). Karisma is an organization is an NGO invested in protecting human rights in the digital environment. Joan's work relates to the creation of massive databases for digital government analyzing the interactions of the public sector, private sector, and international organizations. He has analyzed the creation of biometric identification databases in Mexico and the centralization of the diverse databases and the datafication of the healthcare system in Colombia.

Keywords: human rights, digital government, biometrics

Aaron Martin, PhD is a tech policy expert specializing in cybersecurity, privacy and digital identity. In 2018 he joined the Tilburg Law School as a Postdoctoral Research Fellow on the ERC Global Data Justice project. His research interests include cyber policy, critical infrastructure protection, surveillance, biometrics, and technology regulation. He is particularly interested in exploring these topics across the Global South. In 2015-2018, he was a Vice President of Cyber Policy at JPMorgan Chase in New York. He has also worked in tech policy roles at the OECD, European Commission, Vodafone Group, and elsewhere. Aaron has a PhD in Information Systems and Innovation from LSE.

Keywords: Biometrics, identification, regulation, policy, cybersecurity

Silvia Masiero is a Lecturer in International Development at the School of Business and Economics, Loughborough University. Her research focuses on the use of information and communication technologies in the field of socio-economic development. In particular, she studies the multiple forms of embeddedness of the IT artefact in development policy and governance, with a specific interest in its participation in the politics of anti-poverty programs. Her current work revolves around the role of digital technologies in reshaping the politics of the Public Distribution System (PDS), the biggest food security scheme in India.

Keywords: ICT4D, datafication, anti-poverty programs, Aadhaar, India

Ulises Ali Mejias is Associate Professor of Communication Studies and director of the Institute for Global Engagement at the State University of New York, College at Oswego. Ulises is the author of *Off the Network: Disrupting the Digital World* (University of Minnesota Press, 2013), and with Nick Couldry, of *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism* (forthcoming in 2019 from Stanford University Press). He has published many articles, including 'Liberation Technology and the Arab Spring: From Utopia to Atopia and Beyond' in *Fibreculture* (2012), and 'The Limits of Networks as Models for Organizing the Social' in *New Media and Society* (2010).

Keywords: Critical internet studies, network theory and science, philosophy and sociology of technology, and political economy of digital media.

Hellen Mukiri-Smith is a PhD Researcher on Dr. Taylor's ERC Global Data Justice project at Tilburg Institute for Law, Technology & Society, Tilburg Law School. Her research interests include digital surveillance law, big data and power, data value chains, data governance and ethics, and the impact of data technologies on people's privacy and other areas of human development. Hellen's PhD research focuses on Kenya as a case study.

*Keywords***:** Big data, biometrics, data regulation and human development

Nelli Piattoeva is Associate Professor in new social research in education based at the University of Tampere, Finland. She has researched the policies and practices of datafication in formal (school) education as particularly manifested through the proliferation of national large-scale assessments of learning outcomes. In parallel, she has done research on the coloniality of academic knowledge production in/on the post-Soviet region and its education systems. These two intersecting lines of research prompted a recent interest in engaging with the proliferation of big data on education through the lenses of (post)colonial and decolonial theorizing. Her main geographic focus is Russia, and she also aspires to look into the global education data production as well as extend my research into the Central Asian post-Soviet countries.

Keywords: Education, governance, large-scale assessments, Russia, coloniality

Marie-Therese Png is a PhD candidate at the Oxford Internet Institute and PhD research intern at DeepMind Ethics and Society. Her research interests are epistemic (in)justice, decolonization, geographic inclusion, etc. in AI Ethics after spending a short stint in AI Policy. She recently joined the IEEE Ethically Aligned AI Classical Ethics Committee, critiquing the monopoly of Western ethics in AI ethics / value alignment of autonomous systems with 'human values'. Previously, she was Research Affiliate at the MIT Media Lab. She did her undergraduate in Human Sciences at Oxford, and her Master's in Developmental Cognition and Intergroup conflict at Harvard.

Keywords: AI ethics, digital divides, decolonization theory, critical data studies, intercultural information ethics

Anita Say Chan is an Associate Professor in the Department of Media and Cinema Studies at the University of Illinois, Urbana-Champaign. Her research interests include globalization and digital cultures, innovation networks and the periphery, and science and technology studies in Latin America. She received her PhD in 2008 from MIT. Her book *Networking Peripheries: Technological Futures and the Myth of Digital Universalism,* was released by MIT Press in 2014, and in Spanish with the *Instituto de Estudios Peruanos* in 2018. She is a faculty affiliate of the School of Information Sciences and the Illinois Informatics Institute and is the founder of the Recovering Prairie Futures Research Network at UIUC.

*Keywords***:** Digital universalism, innovation at peripheries, decentering big data, local data, data care

Gabriela Sued is PhD Student at *Instituto Tecnológico de Monterrey* in Mexico. She was born, studied and worked in Buenos Aires. She taught in Communication and Media School of University of Buenos Aires until she moved to Mexico to make her PhD. She has also a MA degree in Social Studies of Science and Technology. She is specialist in Digital Cultures. Her PhD thesis is about Digital Cultural Objects and Digital Methods. She is interested in deepening and disseminating digital methods in Mexico and Argentina in both empirical and critical way to understand the current entangles between society and data. She is also researcher in gender and technologies areas.

Keywords: Data practices, north-south translations (and also south-north), data critical studies, citizen participation, critic of algorithms and automatization

Güneş Tavmen has recently submitted her PhD thesis at Birkbeck, University of London. The research is an investigation of implications of open data discourses and infrastructures in the context of smart city making and emerging citizenships. While her research focused on London, most of these discourses and technologies are intended to be exported to different geographies such as South Asia, Middle East and Sub-Saharan Africa. Prior, Güneş worked with a Turkish NGO.

Keywords: Critical data studies, smart city, infrastructure, open data, digital citizenships

Linnet Taylor is Associate Professor at the Tilburg Institute for Law, Technology, and Society (TILT). Her research focuses on digital data, representation and democracy, with particular attention to transnational governance issues. She leads the ERC Global Data Justice project, which aims to develop a conceptual framework for the ethical and beneficial governance of data technologies on the global level. The research is based on insights from technology users, providers and activists worldwide.

Keywords: Data justice, data markets, governance, political economy, privacy

Quynh Tu Hoang is a Research Master student in Media Studies at University of Amsterdam. Her research focuses on the intersection of human rights issues, surveillance and privacy, information controls, and cyber security. Trained as a new media researcher, Quynh has researched issues of textile labor, digital economy labor, data activism, and information controls in. She is a member of the 88 Project, whose work focuses on human rights advocacy in Vietnam.

Keywords: Information controls, decolonizing privacy universalism, resistance

Étienne von Bertrab is a Mexican academic and activist based in London. Teaching Fellow at the Development Planning Unit, University College London, where he has taught various postgraduate modules around the political ecology of infrastructure, water politics, and more recently, communication, technologies and social power. He is the co-founder of several organizations and networks around the right to the city and the struggle for deeper democracy in Mexico. He recently launched Albora: Mexico's Geography of Hope, a research and communication project aimed at restoring trust and cultivating hope in his home country.

Keywords: Social movements, territory, solidarity networks, hope

Norbert Wildermuth is Associate Professor, Communication & Arts, Roskilde University, Denmark. He has 30 years of fieldwork experience in the Global South, including Kenya, India, Nepal, Bhutan, Zambia, Sierra Leone, Mozambique, Vietnam and Brazil. In recent years, his qualitative research focus on young people and their everyday media uses has developed into a core interest in questions of communication for social change, digital inclusion/empowerment, online activism, e-participation, e-governance and open data initiatives. He has organized the symposium 'Smart Data and Sustainable Development in Eastern Africa: A Participatory Communication Perspective' in Nairobi, in April 2016.

Keywords: ICT-facilitated social accountability and transparency; non-profit crowd-mapping; citizen-driven election monitoring; big/smart data for sustainable development; ICT-facilitated participatory governance.

Authors' Bios

Anna Berti Suman is a PhD researcher at the Tilburg Institute for Law, Technology, and Society (NL), currently also Visiting Researcher at the EC Joint Research Center. Her PhD project aims at investigating how Citizen Sensing – environmental monitoring 'from below' – can challenge and ultimately influence environmental risk governance. Anna has work and research experience in environmental conflicts (Unión des Afectadxs por Texaco, Ecuador), water management (Comisión Económica para América Latina y el Caribe and Fundación Chile, Chile) and the public health sector (Royal Brompton Hospital, London). Besides academia, Anna volunteers as pro-bono environmental lawyer for non-governmental organizations (e.g. Greenpeace International).

Tomás Dodds is a PhD researcher at the Institute of Cultural Anthropology and Development Sociology at Leiden University. His research is concerned with how newsmakers are appropriating digital technologies and how these innovations are transforming the infrastructure, temporality and form of the newsmaking process. Keywords: journalism studies, media practices, digital technologies.

Sander van Haperen is a postdoc fellow in the DATACTIVE project and lecturer in the Media Studies Department of the University of Amsterdam. He studies the development of social movements, with a particular interest in leadership, governance, and digital networks. His research advances computational methods, drawing on social media, complexity, and network analysis, integrated with qualitative inquiry.

Stefania Milan is an Associate Professor of New Media at the University of Amsterdam. Her work explores the intersection of digital technology, governance and activism, with focus on grassroots data practices and autonomous infrastructure. Stefania is the Principal Investigator of the DATACTIVE project, exploring the evolution of citizenship and participation vis-à-vis datafication and surveillance (data-activism.net, ERC Starting Grant no. 639379), and of the ALEX project (ERC Proof of Concept grant no. 825974). She is the author of *Social Movements and Their Technologies: Wiring Social Change* (Palgrave Macmillan, 2013/2016) and co-author of *Media/Society* (Sage, 2011).

Niels ten Oever is a PhD candidate in the DATACTIVE project at the University of Amsterdam. His research focuses on the evolution of the notion of public interest in the Internet architecture. Niels tries to understand how invisible infrastructures provide a socio-technical ordering of our societies and how that might influence the distribution of wealth, power and possibilities. Previously Niels has been Head of Digital for ARTICLE19 where he designed, fundraised, and set up the digital programme which covered the IETF, ICANN, IEEE and ITU. Before that Niels designed and implemented freedom of expression projects with Free Press Unlimited. He holds a *cum laude* MA in Philosophy (UvA).

Guillén Torres is a PhD researcher at the Media Studies Department of the University of Amsterdam. His research focuses on how civil society's use of public sector information and open data triggers resistance from governmental institutions. He also serves as an advisor for the Mexican organization *ControlaTuGobierno* (ControlYourGovernment), which develops social accountability processes in collaboration with marginalized urban communities.

Emiliano Treré is a Senior Lecturer in Media Ecologies and Social Transformation in the School of Journalism, Media and Culture at Cardiff University. He is the *author of Hybrid Media Activism: Ecologies, Imaginaries, Algorithms* (Routledge, 2018), winner of the Outstanding Book Award of the ICA Activism, Communication and Social Justice Interest Group. He has published widely on media ecologies, protest movements, and algorithmic resistance. He is a member of the Data Justice Lab, the co-founder of the Big Data from the South Initiative, and the current vice-chair of the 'Communication and Democracy' Section of the European Communication Research and Education Association.

Kersti R. Wissenbach is a researcher and senior consultant working on the crossroads of communication, governance, responsible data, and civic tech. She is specialized in participatory methods and has worked with activist groups, NGOs, and government institutions in over 15 countries. Since 2011 Kersti runs her own consulting firm providing strategy and policy advice and facilitating co-creation processes. Kersti lectures in the Communication for Development Master of Malmö University. She is a researcher with the DATACTIVE project where she merges social movement and communication for social change scholarship for her study of power dynamics within transnational civic tech activism.

Zhen Ye is a Research master Student at Media Studies Department, University of Amsterdam. She is also a student assistant in the DATACTIVE project. Her research interests focus on (gendered) digital labor on social media platforms.

Acknowledgements