UNPLUGGED

The Fading Gloss of Data Science: Towards an Agenda that Faces the Challenges of Big Data for Development and Humanitarian Action

Miren Gutierrez[1] · John Bryant[2]

Abstract

Different UN and international agencies are busy trying to leverage big data to unlock its value for evidence-based decision-making in development and humanitarian action. But many vulnerable people are invisible to the data infrastructure, while just integrating their data without understanding the consequences can make them even more vulnerable. This article unpacks the challenges presented by data science for development and humanitarianism.

Keywords Humanitarianism · Development · Sustainable development goals · Algorithmic bias · Data infrastructure

Data Promises and Pitfalls in the Context of Development

The development sector is progressively taking advantage of the opportunities brought about by data science and datafication –understood as the transformation of everything we do into data (Baack 2015)— to improve developing societies. Different United Nations and international agencies are busy trying to leverage machine learning (ML) and other automated learning techniques to unlock the value of big data for evidence-based decision-making in development and humanitarian action. ML is an automatic learning system made of software packages that, fed by big data, make artificial intelligence (AI) possible (Cukier 2014). The UN system was one of its early adopters. 'The UN has been advocating for a data revolution since before it had a name and began working on big data applications just as the concept was emerging', proclaimed UN Chronicle in 2018 (2018a).

However, initial enthusiasm seems to have given way to scepticism. In 2019, UN Special Rapporteur on Extreme Poverty and Human Rights Phillip Alston cautioned that digital technologies could be a 'trojan horse' for forces that seek to dismantle and privatize economic and social rights, undermining progress toward the Sustainable Development Goals

✉ Miren Gutierrez
m.gutierrez@deusto.es

[1] University of Deusto, San Sebastián, Spain

[2] Humanitarian Policy Group, ODI, London, UK

(SDGs) instead of speeding it. 'As humankind moves, perhaps inexorably, towards the digital welfare future it needs to alter course significantly and rapidly to avoid stumbling zombie-like into a digital welfare dystopia' (Alston 2019). Besides, the COVID-19 pandemic has had two main effects; on the one hand, making people with access more dependent on digital platforms for log on to health, education, banking, entertainment, and food and other product delivery services, and on the other, exacerbating digital divides between those with access and those without (United Nations 2021). In 2020, UN Special Rapporteur on Racism Tendayi Achiume warned that technology is shaped by and frequently worsens existing social inequalities.

So, what has happened? Next, we review some of the opportunities, challenges of the data infrastructure –understood as the processes, software, and hardware necessary to turn data into actionable information. It provides an overview of how the understanding and critical reactions to big data approaches have changed in these sectors over time, and how concerns of 'digital divides', privacy, opaque algorithms, and more have impacted how big data has been and should be effectively used to advance development and humanitarian goals.
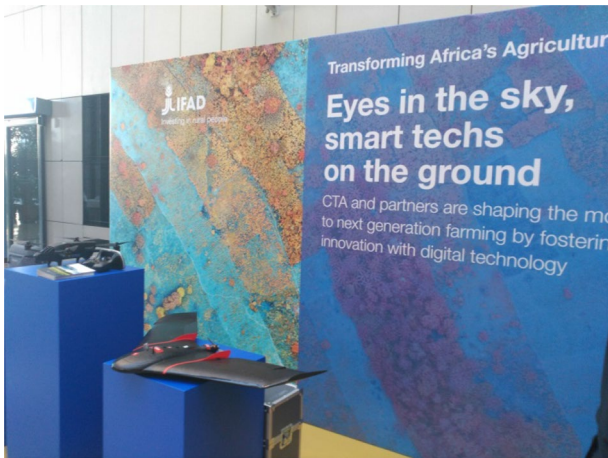
The Opportunities of Big Data

Today satellite imagery and ML are employed to determine poverty by identifying the proportion of thatch-roofed houses (UN Chronicle 2018a). Data from mobile phone networks serve to establish the displacement of people after a

Fig. 1 Drones for development *Source* Picture taken by the author at the Governing Council of the International Fund for Agricultural Development, Rome, 2019

disaster hits and help forecast the expanse of transmittable diseases (UN Chronicle 2018a). Changes in debit card usage are used to ascertain the impact of a crisis (UN Chronicle 2018a). Postal records can be observed to estimate trade flows and access to provisions (UN Chronicle 2018a). Crowdsourced citizen data have been employed during the COVID-19 pandemic to bridge patients in remote areas and supplies (Lilja et al. 2020). Other remote sensing technologies, such as uncrewed aerial vehicles (drones), are used for agricultural management as well, as Fig. 1 shows.

The UN agency The Food and Agriculture Organization and social enterprises are proposing, for instance, to use Distributed Ledger Technologies, commonly referred to as 'blockchain' –a sequence of archives linked by cryptography that can serve to secure transactions— to offer 'smart contracts' to farmers and fisherfolk networks that go past intermediaries (Sylvester 2019). Blockchain is already being leveraged to launch independent supply chains for fair trade coffee, for example (Koffman 2019). The potential gains in efficiency and traceability of transactions or information flows through such technology have also been noticed by the humanitarian sector, with the World Food Programme, World Vision, and IFRC all involved with at least pilot projects to improve cash transfer schemes (Coppi and Fast 2019).

Beyond institutions and companies, people and non-governmental organizations are exploring the potential of these data-based technologies to generate diagnoses and solutions to social problems. For example, cartographic platforms, able to visualize verified citizen data in near real-time, are being deployed to assist in humanitarian operations; satellite data from fishing vessels operating at sea have been mapped to reveal irregular activity impacting coastal communities, as well as to trace refugees as they attempt to cross the Mediterranean from Northern Africa, and maps that change in real-time are used to alert people of flooding or insufficient quality water and air (Gutierrez 2018a).

The opportunities are such that UN Global Pulse has proposed a taxonomy of data applications for development to steer data-centred efforts. This taxonomy includes two types of cases: in the first, data are employed in early warning systems and real-time awareness, conveying what is happening and may happen in the future; in the second, data can feed real-time feedback, adding the reasons why something is happening and shedding light 'on what could be done about it' (Letouzé 2012). This division echoes albeit blurred separation between humanitarian agencies –responding to immediate needs— and development organizations –focused on long-term causes.

Datafication, a platformed internet, the omnipresence of the internet of things, and the growing accessibility of tools and skills to transform data into relevant and actionable information are transforming development and humanitarianism. Most literature is focussed on the data promise; however, the data infrastructure presents different challenges.

The Challenges of Big Data for Development

Vinuesa et al. say that the emergence of AI and its impact 'requires an assessment of its effect on the achievement of the Sustainable Development Goals (SDGs)' (Vinuesa et al. 2020: 1). Concretely, AI can enable 134 targets across all the SDGs, but 'it may also inhibit 59 targets' (Vinuesa et al. 2020: 1). Concerns include that data-driven approaches for policing can hinder equal access to justice because of algorithm bias; therefore, legislation and ethical standards that guarantee transparency and accountability of AI are needed (Vinuesa et al. 2020). These authors conclude that 'the fast development of AI needs to be supported by the necessary regulatory insight and oversight for AI-based technologies', with the potential to both enable sustainable development and hinder it, by generating 'gaps in transparency, safety, and ethical standards' (Vinuesa et al. 2020: 1).

Bringing different layers or datasets together for new insights is the main offering of big data, but that process can also lead to new harms for already marginalized and at-risk communities (Bell et al. 2021: 3). There is consensus that individual rights, privacy, identity and security, and the presumed infallibility of data anonymization, are general areas where potential harm can occur when personal and other data are exploited (Letouzé 2014; Wes 2017; Data Protection Commission 2018). The 'uniqueness of mobility' that everyone with a phone exhibits makes it possible to identify 95% of individuals with only four spatiotemporal points (de Montjoye et al. 2013). By correlating information from several different sources, intruders can patch together a profile and use that information to connect more sensitive

pieces of data. So many 'anonymized' datasets have been compromised using these tactics that some in the field are declaring 'anonymization is dead' (Zibuschka et al. 2019). Some experts now talk about pseudo-anonymized data (Pandit et al. 2018). There is a body of evidence showing that the combination of large personal datasets can open the door to deanonymization, even if each dataset is individually inoffensive (Bettilyon 2019). For instance, de-identified health data from the Australian Medicare Benefits and the Pharmaceutical Benefits schemes were re-identified, employing public information about people (Culnane et al. 2017). The implications for vulnerable people affected by humanitarian crises or in poor communities are still to be explored thoroughly. For example, Oliver denounced in 2014 that when the UN Stabilization Mission (MONUSCO) in the Democratic Republic of Congo proposed to the humanitarian community that drones could be shared with the military for information collecting in 2014, people's identities and the principles of neutrality, which allow relief organizations to operate, could be compromised (Oliver 2014). Zibuschka et al. are calling for a new paradigm that prioritizes transparency regarding data collection over attempts to anonymize the data (Bettilyon 2019).

The platformization of personal data offers a specific challenge. One example is the European public service media (PSM) that has the remit to serve the public interest with equality, universal access, and social solidarity. In a platformed environment, PSM has chosen to share content on platforms such as Facebook and YouTube, submitting to the muscle of technology giants (Sørensen and Van Den Bulck 2010). After a series of scandals related to leaks of user data and distribution of fake news, and pressures from EU legislation to improve privacy, PSM in UK, Netherlands, and Finland started to reconsider their strategies (Sørensen and Van Den Bulck 2010). Facebook's and other platforms' data mining and usage practices are under scrutiny because of their impact on privacy and democracy. Embracing corporate data-gathering platforms without questions could make vulnerable people even more vulnerable as they become visible to the data infrastructure (Niklas and Peña Gangadharan 2018). For example, in 2018, ethnic cleansing in Myanmar was incited on Facebook, leading to the recent filing of a $150 bn lawsuit against the firm by Rohingya communities in the US and Europe (Whittaker et al. 2018; Chandran and Asher-Schapiro 2021). This matter is especially relevant when these platforms are taking steps to intervene in global issues interlinked with development and humanitarianism. For instance, Facebook, Microsoft, Twitter, YouTube, Dropbox, Amazon, LinkedIn, and WhatsApp are part of an independent online counter-terrorism forum to combat terrorism (Kwan 2019). The Global Internet Forum to Counter Terrorism (GIFCT) aims to work on 'sustaining and deepening industry collaboration and capacity' to 'prevent terrorists and violent extremists from exploiting digital platforms' (Global Internet Forum to Counter Terrorism 2010). How they intend to enforce and mitigate terrorism is still to be seen.

Citizen initiatives have been criticized too for embracing corporate technologies. An example is Ushahidi –testimony in Swahili—, a platform that allows mapping crowdsourced citizen data for humanitarian support in crises, which depends on mobile technology, whose business models are based on low production costs sometimes supported by semi-slave work (Palmer 2014).

One answer to these challenges could be to include ordinary people –the beneficiaries of development and humanitarian programmes— in the decision-making processes of data-based solutions. As argued by Kennedy (2018), debates of datafication often draw on the views of the elite and techno activists; however, 'it is important to take account of what non-expert citizens themselves say would enable them to live better with data' (Kennedy 2018). Moreover, as datafication may produce potentially discriminatory outcomes (Gangadharan 2012; Tufekci 2014a; van Dijck 2014; Eubanks 2018;), it is also relevant to explore and pre-empt the challenges presented by data science for development and humanitarian action. However, the integration of ordinary people in data processes is not innocuous and presents new challenges. Due to access to skills, tools, and opportunities, data processes impose new asymmetries within data agents even if they are citizens (Gutierrez 2019). For instance, collaborative mapping exercises can be manipulated by those in charge of making the map and verifying crowdsourced data (Halkort 2019).

This article explores four main areas of concern that data science poses around data absence/presence and data integration processes and narratives. The analysis highlights the contested nature of the data infrastructure and raises questions about the appropriateness of datasets and data processes for development and humanitarianism. These areas of concern include (a) the lack of data that results in the invisibilization of people and groups; (b) the targeting or excess of data that result in discrimination of targeted groups; (c) and the challenges of combining data, often insufficient data, from different mining methods; and (d) the need for new narratives on data-driven development. The analysis shows that these areas should be integrated when humanitarian and development programmes are being designed, not as an afterthought. This analysis draws mainly on secondary data and case studies from previous analyses (Gutierrez 2018b; Halkort 2019) and critical data studies, and decolonization theory literature.

🕸

Current Policy and Practice

In 2013, the UN Department of Economics and Social Affairs Under-Secretary-General Wu Hongbo demanded more data in development. 'Statistics is shaping our understanding of the world', he said, addressing the UNSC (UNDESA 2014). In its 2014 report, the International Telecommunication Union noted the UN Statistical Commission (UNSC) and national statistical organizations were looking into 'ways of using big data sources to complement official statistics and better meet their objectives for providing timely and accurate evidence for policy-making' (ITU 2014: 173). The *big data revolution* was understood as offering great opportunities, and new data-based development-focused organizations started to sprout globally. Global Pulse –an 'innovation initiative' on big data promoted by the UN— was launched in 2014.

Other development organizations outside the UN started to focus on big data solutions. These include Data-Pop Alliance (created by Harvard Humanitarian Initiative, MIT Media Lab, and ODI in 2014); Global Partnership for Sustainable Development Data (set up in 2015 and funded by the UK Department for International Development, Hewlett Foundation, US President's Emergency Plan for AIDS Relief, World Bank and US Millennium Challenge Corporation), and Big Data for Development (launched in 2018 with support from Canada's International Development Research Centre).

After a period of zeal, during which the accent was on maximizing the adoption of big data, now the UN says it is 'actively working to accelerate the discovery, development, and adoption of privacy-protecting big data applications' (UN Chronicle 2018a). In 2017, the UN adopted the Data Privacy, Ethics and Protection Guidance Note on Big Data for Achievement of the 2030 Agenda (United Nations Development Group 2017). UN Global Pulse also co-founded and chaired the inter-agency UN Privacy Policy Group (UN PPG), which in 2018 developed a set of principles on Personal Data Protection and Privacy (UN Chronicle 2018a).

The UN has recently started to incorporate data ethics into its innovation projects by conducting a Risks, Harms, and Benefits Assessment to 'identify anticipated or actual ethical and human rights issues that may occur at any stage of a data innovation process' (UN Chronicle 2018a). A 14-page Risks, Harms, and Benefits Assessment tool, available online and shaped as a questionnaire, asks new projects whether they use data that directly identifies individuals or 'sensitive data', and if not, whether the data can be used to single out a concrete person applying 'accessible means and technologies', as well as questions about the legitimacy and fairness of the access to and use of the data (UN Global Pulse 2016). 'Transparency is a key factor in helping to ensure accountability and is generally encouraged', says the

document (UN Global Pulse 2016). A due diligence process is to be employed when selecting partner third parties with access to data; 'you should only transfer personal data to a third party that will afford appropriate protection for that data', says the directive (UN Global Pulse 2016). Similar principles around data sharing and shared action on agreeing data responsibility among agencies were recently codified in the OCHA Data Responsibility Guidelines (2021).

Nevertheless, it is still to be seen how these principles and safeguards work in practice. For example, in February 2019, the World Food Programme (WFP) signed a $45 million partnership agreement with Palantir, a software company financed by the US Central Intelligence Agency and involved in intelligence contracts with US military and police forces (Parker 2019; World Food Programme 2019). The deal raised concerns about data management from the ninety million people affected by conflict and disasters that the WFP supports in different parts of the world. In response to criticism, WFP has since defended the partnership on the grounds of increased efficiencies and so more people fed, and lives saved and insisted the personal information of aid recipients will not be shared. As critics argue, however, de-anonymization through merging datasets are one of the tools Palantir offers its other clients, and the lack of transparency over the deal sets a dangerous precedent for other aid organizations (Igoe 2018).

Beyond the UN, the Global Partnership for Sustainable Development Data, a global network that includes governments, businesses, and civil society organizations, has generated initiatives such as the Collaborative Data Innovations for Sustainable Development (Global Partnership for Sustainable Development Data 2016). This initiative has developed a 'Data4SDGs Toolbox' (Data4SDGs Toolbox 2016; Global Partnership for Sustainable Development Data 2016). DataSwift –an initiative of CIVICUS in partnership with Wingu, the Open Institute, and Restless Development Tanzania— builds the capacity of civil society organizations to produce and use citizen-generated data. The ICRC and governments of Switzerland and Norway have led on a process to articulate the notion of 'digital dignity' that respects the 'agency, autonomy and identity' of individuals when their data is used to inform humanitarian action (Wilton Park 2019: 3). Much of the recent drive toward a better articulation of good practices and basic standards on was sparked by concerns of the aid sector's lack of understanding of potential data harms by the ICRC and Privacy International (2018) and Harvard Humanitarian Initiative (Greenwood et al. 2017).

Data governance –rules, laws, actors, relations, and strategies to collect, manage, share, store, and analyze data— has become an important issue as well. And within data governance, regulation can have a role in protecting people from predatory data practices. In the European Union, the General

Data Protection Regulation (GDPR), which came into force in May 2018, is considered relatively stringent by global standards. The regulation applies to systems considered 'solely' automated, includes a blanket ban on fully automated decision-making that 'significantly affect' people, and rules that people affected by these systems have the 'right to explanation' and are provided 'meaningful information' about them (European Commission 2018). EU regulators identified categories of criteria that are likely to result in risk and the need for a data protection impact assessment (DPIA), which includes the categories 'automated decision-making' with legal effects, 'sensitive data or data of a highly personal nature', 'data processed on a large scale', 'data concerning vulnerable subjects', 'interference with rights or opportunities' (European Commission 2017: 9–10).

However, none of these categories above is well-defined. European regulation requires that people have the right not to be subject to a decision-based solely on automated processing 'without human involvement'. However, algorithms are neither created *ex nihilo* nor are they *natural occurrences*; they are created precisely by humans. Besides, Goodman & Flaxman note that this regulation 'potentially' forbids 'a wide swath of algorithms currently in use in recommendation systems, credit and insurance risk assessments, computational advertising, and social networks' (Goodman and Flaxman 2017). Precisely how far this prohibition goes is yet to be seen, in terms of both the data and the organizations to which it applies. Sensitive data, prohibited by the GDPR, include racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometrics, health, sex life, and sexual orientation unless those data are 'manifestly public', the 'data subject' gives their consent or it is processed in the name of 'public interest' (European Commission 2019). Some of these categories (e.g., ethnic origin) are distinct, but others (e.g., religious, and philosophical believes or public interest) are not clear-cut. The applicability of GDPR to international organizations conducting data collection as part of humanitarian activities is also disputed, with various exemptions and privileges making enforcement unlikely. But a reputational 'soft' pressure to uphold at least the spirit of the regulations has led to various UN agencies and ICRC creating their own good practice data guidance (Gazi 2020; Kuner 2020).

The European experience in data protection is being observed in other regions, and court cases are also offering interesting learnings. For example, the Centre for the Study of the Economies of Africa Director of Research Adedeji Adeniran says in an interview:

> What we have learned from the European experience on data governance is that to establish suitable rules for the game, you must be sitting at the right table. The European Union is trying to regulate companies

which are mostly based in the United States and struggling to do so because decision-makers do not sit at the same table as these large platforms and Europe does not have the right software champions to join the debate. We must learn from this experience and make sure we create a seat at the right table when it comes to establishing data governance rules (Adeniran 2021).

This is a vital area of concern. A platform is a technological base, sustained by data and algorithms, 'on which complementary add-ons can interoperate, following standards and allowing for transactions amongst stakeholders, within the platform-centric ecosystem' (Sun and Keating 2015). Whether they charge for services and delivery of products or provide a stage for communication and transactions, all platforms have a common business model based on the extraction of personal data that they have not generated. The platform economy –estimated at $7 trillion in 2018– is based on algorithms' ability to scrutinize everyone linked to the platform, reap their data, and translate them into valuable, monetizable information, which can serve to generate user profiles or to be sold to third parties, sometimes abroad. In developed economies, access to communication, services, and content has been platformized (Poell et al. 2019), a process accelerated by the response to COVID-19 and the need to stay home to avoid contagion. Though platforms are ruled by conglomerates headquartered mainly in the US or China, their millions of users generating data do so from across the world, making regulation at national levels difficult to uphold (Consultancy.org 2018).

The growing dependence on platforms is a challenge because welfare platforms are increasingly affecting fundamental rights. In 2021, the Dutch government had to resign over an algorithmic system designed to detect fraud (System Risk Indication), employed in the poorest neighbourhoods where immigrant populations tend to concentrate (Abu Elyounes 2021). A Hague Court decision annulling the data collection and the creation of risk profiles of poor Dutch citizens decided the system was grounded on an infringement of Article 8.2 of the European Convention on Human Rights (Battaglini 2020).

There is strong literature on datafication (McNeill 2021). Recent work classifies the multi-layered supremacy of algorithmic power (Andersen et al. 2016) while identifying impenetrable algorithmic systems called *black boxes* (Yang and Pandey 2011; Letouzé and Sangokoya 2015), and epistemological disproportionateness between powers of critique and scientific practice (Elwood 2007; Baack 2015). Much scholarship tackles data biases or how regulation can be developed around an ethical computation (Goodman and Flaxman 2017). Other theorizations include exploring fair practices that advance data science to create ethical data science (Awan 2016) or data justice (Dencik et al. 2016). More

general concerns about regulation and guidelines on ethical data include the lack of consensus or a commonly accepted definition of 'fair' in data mining and usage. Next, the four areas of concern regarding data mining and processing for development are explored more in detail.

The Four Gaps in Data Integration

The specific areas of concern of data science for development include (1) the fact that many of those that most need assistance live in the data infrastructure's shadow; they are invisible to the data infrastructure. However, (2) just integrating them in datasets poses a second challenge, which connects to the general concerns about privacy and individual rights mentioned earlier. (3) The data integration and processing itself must be dealt with as well, as different data gathering methods, formats, and gaps can interrupt research on development issues, with consequences for policymaking and humanitarian action. Meanwhile, (4) data science for development and humanitarianism seems to have adopted the corporate *newspeak*. Data and the ML algorithms that feed on them are *sold* by the platforms that create them as inevitable, automatic, and spontaneous, and their outputs as more *objective* than human decision-making (Peng 2017; Naughton 2018; Smith 2018). But the data infrastructure is not innocuous.

To face these challenges, adopting new epistemologies, narratives, and practices that are development- and community-focussed seems prudent in the face of the delicate task of integrating vulnerable people into the data infrastructure, as seen next.

Data Presence: No One Left Behind?

Millions of people on this planet live in the data infrastructure's shadow. They do not own a smartphone or a bank account, do not surf online, or live in cities, so they leave no digital traces behind, and their physical movements are not captured by closed-circuit television cameras, sensors, or satellites. For example, the Word Bank wonders whether we can trust 'smartphone mobility estimates in low-income countries' (Milusheva et al. 2021). Besides, women's mobile phone ownership is much lower than that of men (Klapper 2019). Namely, millions of people are invisible for the data infrastructure. Crawford, founder of AI Now Institute –an interdisciplinary research institute in the US—, notes that there is 'little or no signal coming from particular communities' (Crawford 2013). Therefore, in any data analysis about development issues, we need to ask which people are excluded, which places are less visible, what happens if you live outside big datasets (Crawford 2013). These questions are especially relevant when they refer to vulnerable

people in need of development. More recently, the COVID-19 pandemic has made digital divides especially apparent, with increasingly digitized means of tracking, monitoring, and financial support threatening to exclude the 'data poor' (Roese 2021; Lupton 2021).

Similar gaps have long been observed in climate adaptation. For instance, looking at Africa, Adenle et al., who conducted a series of interviews with stakeholders, note that that '(climate) adaptation faces many constraints', including lack of climate data, resulting in impact models that are insufficient for supporting adaptation, particularly as they relate to food systems and rural livelihoods (Adenle et al. 2017). Adaptation refers to the strategies adopted to face irreversible transformations due to climate change; most developing nations, despite their small contribution to climate change, are vulnerable and need to acclimate to its irreversible impacts (Adger et al. 2003). However, the lack of data on both needs and programmes' impacts, among other factors, makes it challenging to create effective programming that allows people to adapt to changes (Adenle et al. 2017).

Datasets are missing because, among other factors, those with the resources to gather data might lack incentives or might not perceive the benefits of doing so, while they also can remove or conceal them (Onuoha 2019). The realities that need to be quantified might resist datafication, too (Onuoha 2019). The shortage of climate finance for climate adaption, for instance, is due, among many reasons, to the lack of uniform standards in labelling projects, the shortage of information on private adaptation (which also includes small household investments), and the different scales and methodologies employed by scholars to look at climate finance (Gutierrez and Gutierrez 2019). Another way of leaving people behind is to work with aggregated datasets that, for example, do not separate data for women or vulnerable groups (Gangadharan 2012). Thus, the challenge of applying data science to diagnose and solve development problems is that many of those who are in need are invisible to the data infrastructure.

During a Thomson Reuters Foundation workshop with professional journalists of the Global South, reporters from Tanzania told participants how they had to resort to cajoling guards at cemeteries and hospitals to tally coffins and guestimate the number of excess deaths during the COVID-19 pandemic because of the lack of official reports (Thomson Reuters Foundation 2021).

The danger here is to try to fill the gaps with the wrong datasets. And indeed, some scholars have questioned data representativeness and the validity of extrapolating, for example, from digital users' datasets to entire populations (Berry 2011; Crawford 2013; Trevisan 2013; Innerarity 2013, 2016; Andrejevic 2014; Zelenkauskaite and Bucy 2016). Talking about why social platforms' big data are 'bad data' for researching populations and social movements,

Schradie lists several flaws, including that 'hashtag data are often cherry-picked' (Schradie 2015: 1). A problem in social movement studies has been choosing case studies based on high levels of internet and social networking platforms employment and using 'too small' big data that exclude 'those who are on the other side of the digital divide', leaving out 'the poor and working-class' (Schradie 2015: 2). Such divides are also a concern in the humanitarian sector: studies of crowdsourcing efforts to track needs in post-earthquake Haiti and Nepal for example show resulting 'crisis maps' tend to reproduce the density of people able to participate online, rather than the severity of needs (Mulder et al. 2016). Without active efforts to counter divides, crowdsourcing projects risk portraying the 'situated knowledge of powerful groups' rather than any 'wisdom of the crowds' (Cinnamon 2019: 9).

Data Discrimination: When People are Singled Out in Datasets

The act of integrating people into data infrastructure can itself pose risks. For example, an AI Now Institute report describes how, in 2018, ethnic cleansing in Myanmar was incited on Facebook; Google built a secret engine for Chinese intelligence services and helped the US Department of Defence analyze drone footage. Microsoft signed contracts with US Immigration and Customs Enforcement (ICE) to use facial recognition on migrants (AI Now Institute 2018). Critical scholars, including Braman (2009), Tufekci (2014b), and van Dijck (2014), describe how the data infrastructure is employed to profile people, discriminate against vulnerable groups, and promote constant, omnipresent, and preventive monitoring.

Embracing data-gathering processes and data integration without question may make people who are otherwise digitally 'invisible' even more vulnerable. In the humanitarian sector, much of the conversation around the risks of data-gathering without adequate consent processes has focused on biometric registration – highlighted by the case of UNHCR handing over sensitive biometric data of Rohingya refugees to Bangladesh and then Myanmar, the government responsible for their persecution (HRW 2021). Yet personal data is not necessary to risk harm to marginalized people, especially with the focus of big data on interoperability. The risks of 'mosaicking' –where different, less sensitive datasets are brought together to create highly sensitive data that can identify particular groups – has been raised as a key concern in the sector (Capotosto 2021). GIS mapping may also risk, through making visible those fleeing conflict or in insecure land tenure arrangements, further persecution. Instances of land clearance shortly after information on informal settlements and displacement camps were made public has been recorded in Jordan and Kenya (Bryant 2021). Contrary to a

language often used by aid organizations around the importance of providing a 'digital identity,' the safety of many persecuted communities in humanitarian contexts is dependent upon an invisibility that risks being compromised through these data-gathering processes.

Not all unfair or erroneous results in algorithmic processes are down to input data's failings; the code can also be limited, opaque, or inaccessible. These issues present different challenges (Gutierrez 2021). Rather than bias-free decisions, being taught using datasets unrepresentative of the population they are then used on has led to discriminatory practices reflecting human biases in sensitive fields including healthcare, policing, and welfare (Zarsky 2016: 126; Iliadis 2018: 222). This is compounded by corporate secrecy and intellectual property regulation which can make code unreachable to inspection, generating a 'black box effect' (Whittaker et al. 2018). Edionwe (2017) argues that algorithms are impenetrable 'for proprietary reasons or by deliberate design' (para. 16). The lack of transparency makes it difficult to dispute algorithmic outcomes and makes decisions based on them seem arbitrary (Zarsky 2016: 130). A different matter is the 'unintelligibility' of some algorithms, which makes it impossible to detect problems or meaningfully engaging with 'algorithmically mediated aspects of life' (Niklas and Peña Gangadharan 2018: 14; Iliadis 2018: 222). Wronkiewicz (2018), Naughton (2018), Edionwe (2017), and Rahimi (2017), among others, suggest that coders themselves do not comprehend how their algorithms operate.

Consequently, groups are campaigning for opening code when that code is acquired and developed by public administrations with taxpayers' money (Civio 2019). That is, the public sector should instruct that providers relinquish code protection mechanisms (Whittaker et al. 2018). The same principles could apply to the UN system, funded by public funds.

Thus, data inclusion appears to pose new challenges; meanwhile, combining different datasets poses a new challenge. As noted for data on climate finance, duplication, data quality, standardization, and consistency for data collection, classification, processing, access control, usage, release, and security is an issue (Yao 2018).

Questions About Mining and Integration

There are many associated challenges regarding the way data are mined and integrated into datasets; this review –which is not comprehensive— focusses on making data compatible and comparable, the issue of anonymization, and the possibility of integrating citizen data in decision-making.

The COVID-19 pandemic has also exposed how difficult it is to make comparisons when the data collection and analysis methodologies are different, incomplete, or faulty.

One example, in the absence of comprehensive official data, has been the spike of pneumonia cases in Tanzania, which, according to journalistic sources, might have been mislabelled COVID-19 cases. There has been a 'paucity of data for hospitalized African patients suffering from COVID-19' notes another study (Kassam et al. 2021). Even open data facilities in developed countries with a track record of open data practices have been criticized for changing formats during the pandemic, making the integration of datasets impossible, datasets that are introduced hurriedly, with erratum, and introducing mistakes. A study of open COVID data in Spain concludes that just making data available is not enough, as:

> There are tensions between (a) the actual conditions in which open data supporters work within the administration and the expectations and needs generated in times of an emergency; (b) the perceptions of a lack of curiosity or need on the part of citizens and the genuine interest exposed by the projects submitted to the Open Data awards, and (c) the data literacy of people and the challenges of data agency (Gutierrez and Landa 2021: 23).

Anonymization is another challenge in data integration, especially when working with vulnerable populations and a lack of standards and protocols. Several studies demonstrate that you can re-identify most individuals in anonymized datasets with few demographic traits (Rocher et al. 2019). Solutions to this problem are emerging both from corporate platforms promising to keep data 'governed' or 'masked'[1] and non-governmental organizations (CyberGhost Research Team 2021). However, this is by no means easy, as digital traces are ubiquitous. Even printers leave invisible code of yellow dots in copies of documents that reveal the printer where the copies were made (Bradbury 2018).

Finally, one space where citizen data integration is being adopted successfully in digital humanitarianism, especially in mapping areas with less interest from for-profit entities. UN agencies such as OCHA have sought to integrate Open Street Map data into crisis mapping, having done so for the first time in 2011, and the digital mapping of hazards, services, and needs is now a key task in the initial phase of the humanitarian response (Gutierrez 2018b). Initiatives that use OSM rose to prominence in the wake of the Haiti earthquake response in 2010, with these 'digital humanitarians' using data gathered through social media and other sources to create crisis maps for the benefit of organizing responses. Beyond this work, open-source mapping has also been used across the world as a means by which locally-based mappers and affected communities themselves – assisted by advocacy and coordination groups like Humanitarian Open Street Map Team (HOT) – can make themselves visible and advocate for improved assistance and services.

Data Storytelling: The Need for New Narratives

Although it seems more practical to focus on what is possible with the available tools, most of which have been devised for commercial purposes, development narratives should integrate criticism. For example, UN Global Pulse has begun to work with big corporations, such as Orange, in what they call 'data philanthropy' (UN Chronicle 2018a). Orange is to launch the Data for Development challenge to make their anonymized database available for researchers to develop applications for sustainable development (UN Global Pulse 2013). However, no critique seems to be part of this initiative. For example, Orange România SA –a provider of mobile telecommunications services— was fined in 2018, 'on the ground that copies of the identity documents of its customers had been obtained and stored without their express consent' (GDPR Hub 2019).

Despite the challenges outlined in this article, data science for SDGs has not disengaged itself from corporate *newspeak*. A Google search for 'data is the new oil' shows 'how entrenched this meme is among big data pundits' (Haupt 2016). 'Data and the machine learning algorithms that feed them are spoken about as a panacea for all ills' (Naughton 2018). Code is discussed 'in almost biblical terms', inspired by corporations, such as Google and Facebook, which 'have sold and defended their algorithms on the promise of objectivity' in decision-making (Smith 2018). Data are compared with electricity and oil (*The* Economist 2017). The Head of Paris21 Johannes Jütting wondered in 2013 whether big data were the 'new oil fuelling development' (Jütting 2013). In 2015, Emmanuel Letouzé, head Data-Pop Alliance, asked the same question (Letouzé 2015). During a conference in Madrid in 2016, Aditya Agrawal, from the Global Partnership for Sustainable Development Data, opened his speech, talking about data as 'the new oil'. United Nations Young Leader Rainier Mallol argued, 'Why data is the new oil' in an article for the World Government Summit in 2017 (Mallol 2017).

However, this is 'a ludicrous proposition', as data are nothing like oil (Haupt 2016). Not only is this metaphor unfortunate in times of climate change's huge impacts on development, but it is also inaccurate. Rather than finite, raw, natural resources, data are a 'cooked' product of cultural, economic, and political processes, and not inevitable or spontaneous (Gitelman 2013; Boellstorff 2013; Couldry and Yu 2018). A similar framing by large technology companies of consumer and other data as 'exhaust' – suggesting a by-product to be harmlessly collected, rather than a valuable product to be extracted and analyzed – has been

---

[1] https://www.ibm.com. and https://www.k2view.com/.

highlighted in wider discussions of 'surveillance capitalism' (Zuboff 2018).

It seems that, given the delicate task in its hands, the development sector –including UN and national agencies, NGOs, and alliances dealing with poverty eradication and development issues— could engender its rhetoric on data science. This is not to say that data science for SGDs should generate its own independent data infrastructure. The private sector has pioneered useful tools based on the data infrastructure; these tools can be and are being employed for development goals. Besides, there is a wealth of open-source tools that the developing world could start using, exploiting, and perfecting. However, the development sector could learn from activists and organizations that are appropriating the data infrastructure to bottom-up data, narratives, and practices, filling gaps, mobilizing people, and creating solutions and social change.

For example, the UN system is working with BBVA bank, Crimson Hexagon, Earth Networks, Nielsen, Orange, Planet, Plume Labs, Schneider Electric, and Waze (UN Chronicle 2018b). It has also signed an agreement with Twitter to access its data (UN Chronicle 2018b). In 2017, the association of mobile operators GSMA partnered with the UN on a strategy built around anonymizing big mobile data to work for the SDGs (GSMA 2017). There are currently 19 mobile operators committed to the GSMA programme 'on big data for social good', which they call BD4SG (GSMA 2017). However, the UN admits these deals carry some challenges:

> A significant challenge, however, is that outside of extreme cases, such as responses to humanitarian emergencies and the prevention of acts of terror, there is no regulatory playbook for data sharing for the public good. Undoubtedly, current efforts such as the European Union's General Data Protection Regulation are paving the way to using new technologies while mitigating the potential risks and harms of data use (UN Chronicle 2018b).

Again, creating its narratives and incorporating a critique of the technologies they use, as well as a mechanism to mitigate harm, could also be part of the mitigating effort. Fuchs has argued for a paradigm shift from 'digital positivism and administrative big data analytics towards critical digital and social media research' (Fuchs 2017). He argues that social science research should combine critical theory.

Discussion and Conclusions

Listening to the conversation about data for development and humanitarianism, one can hear contradictory messages: do we need more data or less? More digital technology or less?
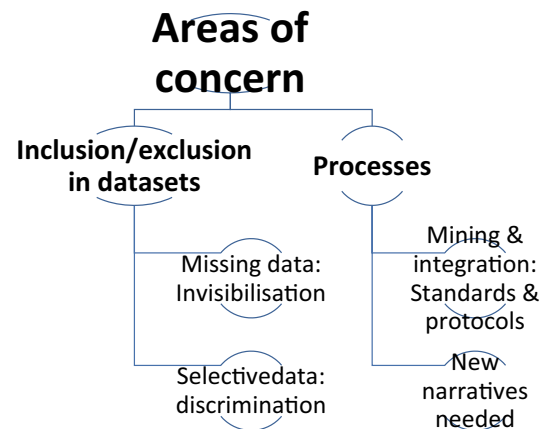
Fig. 2 Areas of concern in the integration of big data solutions for development Source Elaboration by the authors

However, the context and an exploration of circumstances on a case-by-case basis can provide some clues. The COVID-19 pandemic has resulted in millions of deaths globally and exposed profound digital divides, but it has also been a learning experience. While big data analytics and more are now central to the work of some humanitarian response organizations, optimism over the potential of an array of digital tools to monitor or treat the spread of the virus has faded. Many various apps and means of detecting symptoms have shown to be not only ineffective but have also raised the profile of debates around surveillance, the misuse of data, and a 'function creep' around such tools as they are used for other purposes (Lupton 2021: 16). It is in this wider context that the data for development and humanitarian response debates have continued (Fig. 2 summarizes the analysis).

There is a general recognition of digital divides among and within wealthy and poorer countries that affect specific individuals and communities. But these divisions are not straightforward and mean, for example, one cannot adopt the same policies and data gathering strategies in Eritrea, with one of the lowest levels of Internet adoption in the world (6.9% of the population), as in Kenya (40%) (Kemp 2020), although they are both developing countries ranked not far away in the Latest Human Development Index Ranking (Eritrea, in position 180, and Kenya, in position 145) (UNDP 2020). Digital divides are also prevalent within countries and across existing lines of marginalization: though it is estimated that women in Sub-Saharan Africa are on average 15% less likely to own a mobile phone than men, in the continent's largest refugee camp, that divide is 47% (GSMA 2019: 5). Such divides can be the product of deliberate policies that can compound existing dynamics: until last year, Rohingya refugees in camps in Bangladesh for example were banned from accessing the internet or buying mobile SIM cards (Kaurin 2021: 2).

In the context of increasing digitization of services across the world, these divides mean that where there is no autonomous access to such technologies, further marginalization from assistance and the fulfilments of rights can be compounded, as well as the emancipatory and democratizing promises of 'big data' coming to little as power inequities are reinforced, rather than challenged. This goes beyond mere access to the tools themselves: a lack of means of obtaining access to information – including a person's own data –to use it to inform decision-making has been referred to as a 'new digital divide' or 'big data divide' (Cinnamon 2019: 5). As we have seen, this extractive relationship, and the framing and terminologies of data collection as natural processes, benefit those that profit from these processes (Couldry and Yu 2018: 4). Ultimately, they reinforce and expand a colonial-style relationship, giving those who collate and analyze data the power to 'shape the world according to their own worldview' (Cinnamon 2019: 9).

Where there is access, a challenge is fairness to ensure groups are not singled out and discriminated against. Here the answers are not simple either, as even when protocols exist, projects should be explored on a case-by-case basis. One way to ensure that discrimination is avoided is to include people in the project design and analysis of the data from inception. Participatory initiatives are far from new initiatives for the aid sector. They have long been held as a means of overcoming unequal relationships and are not without their own dynamics of inequality that must be recognized and actively countered (Halkort 2019: 323). But in the humanitarian and development sector, where the power inequities between aid providers and recipients are already extremely large, at least some means of increased participation of the service users are vital to mitigate what are increasingly extractive and one-way digital relationships.

In the humanitarian sector, such an approach is still the exception rather than the default, but several more participatory initiatives rely, to varying degrees, on open data. Organizations such as WeRobotics have reacted to Western dominance of the humanitarian crisis-mapping industry through pushing a 'technology transfer' approach to affected communities, training, and hiring local drone pilots as part of autonomous national societies. Humanitarian Street Map follows a similar model of independent hubs in a networked model that is increasingly seen as good practice for more 'localized' humanitarian organizations more generally. Rather than the technologies themselves, successful instances of digital improving aid efforts have been attributed to human elements: pre-existing community networks, trust, proximity, and local translation (Kaurin 2021: 6).

These participatory approaches are often pushing against prevailing forces in data analytics and development thinking. The sector has a poor record of procuring tech services from the Global South, and an increased reliance on data analytics, machine learning, and more may simply limit the extent to which the sector's biggest actors incorporate the views of affected communities into programme design (Spencer 2021: 9). Digital tools in and of themselves deliver little by way of resolving social, political, or economic issues to which the SDGs allude. Without care, they can easily expose people to new harms around misinformation, a lack of consent to data extraction, and violating privacy to the extent to which much of the language around 'data as development' has been criticized as uncritically opening the world's poorest to harmful data analytics processes (Cinnamon 2019: 16).

Another set of challenges emerges when looking at the integration of data and comparison of datasets, needed to explore trends in data and develop fair policies. Examples, such as the glitches and data gaps found in the maps around femicides in Mexico (Amnesty International 2021), show that standardizing, completing, and integrating datasets in precarious circumstances is another challenge for development and human rights.

Ultimately, this study shows that, in attaining the sustainable development goals, *sustainability* should apply to both development and development data.

Declarations

References

Abu Elyounes, Doaa. 2021. Why the Resignation of the Dutch Government Is a Good Reminder of How Important It Is to Monitor and Regulate Algorithms. *Medium*, Berkman Klein Center Collection edition. https://medium.com/berkman-klein-center/why-the-resignation-of-the-dutch-government-is-a-good-reminder-of-how-important-it-is-to-monitor-2c599c1e0100.

Adeniran, Adedeji. 2021. Avoiding extremes in data governance: an Interview with Dr. Adedeji Adeniran of the Centre for the Study of the Economies of Africa. https://www.data4sdgs.org/news/avoiding-extremes-data-governance-interview-dr-adedeji-adeniran-centre-study-economies-africa.

Adenle, Ademola A., James D. Ford, John Morton, Stephen Twomlow, Keith Alverson, Andrea Cattaneo, Rafaello Cervigni, et al. 2017. Managing Climate Change Risks in Africa - A Global Perspective. *Ecological Economics* 141: 190–201. https://doi.org/10.1016/j.ecolecon.2017.06.004.

Adger, W. Neil., Saleemul Huq, Katrina Brown, Declan Conway, and Mike Hulme. 2003. Adaptation to Climate Change in the Developing World. *Progress in Development Studies* 3 (3): 179–195.

AI Now Institute. 2018. Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems. New York University. https://ainowinstitute.org/litigatingalgorithms.pdf.

Alston, Philip. 2019. World Stumbling Zombie-like into a Digital Welfare Dystopia, Warns UN Human Rights Expert. https://www.

ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25156.

Amnesty International. 2021. Mexico: Failings in Investigations of Feminicides in the State of Mexico Violate Women's Rights to Life, Physical Safety and Access to Justice. Amnesty International. https://www.amnesty.org/en/latest/news/2021/09/mexico-failings-investigations-feminicides-state-mexico-violate-womens-rights-life-physical-safety-access-justice/.

Andersen, Jonas Valbjørn, Aron Lindberg, Rikard Lindgren, and Lisen Selander. 2016. Algorithmic Agency in Information Systems: Research Opportunities for Data Analytics of Digital Traces. *In Koloa*. https://doi.org/10.1109/HICSS.2016.571.

Andrejevic, Mark. 2014. The Big Data Divide. *International Journal of Communication* 8: 1673–1689.

Awan, Nishat. 2016. Digital Narratives and Witnessing: The Ethics of Engaging with Places at a Distance. *GeoHumanities* 2 (2): 311–330.

Baack, Stefan. 2015. Datafication and Empowerment: How the Open Data Movement Re-Articulates Notions of Democracy, Participation, and Journalism. *Big Data & Society* 2 (2): 205395171559463.

Battaglini, Manuela. 2020. Historical Sentence of the Court of The Hague Striking down the Collection of Data and Profiling for Social Security Fraud (SyRI). *Transparent Internet* (blog). 2020. https://www.transparentinternet.com/data-ethics/syri-sentence-an-historical-sentence-against-discriminatory-profiling/.

Bell, D., M. Lycett, Alaa Marshan, and Asmat Monaghan. 2021. Exploring future challenges for big data in the humanitarian sector. *Journal of Business Research* 131: 453–468.

Berry, David M. 2011. The Computational Turn: Thinking about the Digital Humanities. *CULTURE MACHINE* 12. file:///C:/Users/m.gutierrez/Downloads/440–893–1-PB.pdf.

Bettilyon, Tyler Elliot. 2019. How Data Hoarding Is the New Threat to Privacy and Climate Change. *Medium*. https://onezero.medium.com/how-data-hoarding-is-the-new-threat-to-privacy-and-climate-change-1e5a21a49494.

Boellstorff, Tom. 2013. Making Big Data, in Theory. *First Monday*. https://doi.org/10.5210/fm.v18i10.4869.

Bradbury, Danny. 2018. Tool Scrubs Hidden Tracking Data from Printed Documents. *Naked Security*, https://nakedsecurity.sophos.com/2018/07/03/tool-scrubs-hidden-tracking-data-from-printed-documents/.

Braman, Sandra. 2009. *Change of State: Information, Policy, and Power*. Cambridge: The MIT Press.

Bryant, John. 2021. Digital mapping and inclusion in humanitarian response. London; ODI https://cdn.odi.org/media/documents/Digital_IP_Mapping_case_study_web_EDSoP6n.pdf

Capotosto, Jill. 2021. The mosaic effect: The revelation risks of combining humanitarian and social protection data. Humanitarian Law and Policy, ICRC Blogs 9 February, https://blogs.icrc.org/law-and-policy/2021/02/09/mosaic-effect-revelation-risks/

Chandran, Rina and Avi Asher-Schapiro. 2021 Rohingya lawsuit against Facebook a 'wake-up call Reuters / News24, 10 December. https://www.news24.com/news24/analysis/analysis-rina-chandran-avi-asher-schapiro-rohingya-lawsuit-against-facebook-a-wake-up-call-20211210

Cinnamon, Jonathan. 2019. Data inequalities and why they matter for development, Information Technology for Development https://www.tandfonline.com/doi/abs/https://doi.org/10.1080/02681102.2019.1650244

Civio. 2019. Que Se Nos Regule Mediante Código Fuente o Algoritmos Secretos Es Algo Que Jamás Debe Permitirse En Un Estado Social, Democrático y de Derecho. https://civio.es/novedades/2019/07/02/que-se-nos-regule-mediante-codigo-fuente-o-algoritmos-secretos-es-algo-que-jamas-debe-permitirse-en-un-estado-social-democratico-y-de-derecho/.

Consultancy.org. 2018. Market Size of Global Platform Economy Surpasses $7 Trillion Mark. https://www.consultancy.org/news/104/market-size-of-global-platform-economy-surpasses-7-trillion-mark.

Coppi, Giulio and Larissa Fast. 2019. Blockchain and distributed ledger technologies in the humanitarian sector. London: ODI https://cdn.odi.org/media/documents/12605.pdf

Couldry, Nick, and Yu. Jun. 2018. Deconstructing datafication's brave new world. *New Media and Society* 00: 1–19.

Crawford, Kate. 2013. The Hidden Biases in Big Data. *Harvard Business Review*. https://hbr.org/2013/04/the-hidden-biases-in-big-data.

Cukier, Kenneth. 2014. Big Data Is Better Data. Presented at the TED, June. http://www.ted.com/talks/kenneth_cukier_big_data_is_better_data?language=en.

Culnane, Chris, Benjamin Rubinstein, and Vanessa Teague. 2017. Health Data in an Open World. *Computers and Society*. https://arxiv.org/abs/1712.05627v1.

CyberGhost Research Team. 2021. 'An Activist's Guide to Online Privacy and Safety.' Privacy Hub. https://privacyhub.cyberghostvpn.com/privacyhub/an-activists-guide-to-online-privacy-and-safety/.

Data Protection Commission. 2018. Anonymisation and Pseudonymisation. Ireland. https://www.dataprotection.ie/en/guidance-landing/anonymisation-and-pseudonymisation.

Data4SDGs Toolbox. 2016. Guidelines for Developing Data Roadmaps for Sustainable Development. http://www.data4sdgs.org/sites/default/files/2017-09/Data%20Roadmaps%20for%20Sustainable%20Development%20Guidelines%20-%20Data4SDGs%20Toolbox.pdf.

de Montjoye, Yves-Alexandre., César. A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. Unique in the Crowd: The Privacy Bounds of Human Mobility. *Scientific Reports* 3 (March): 1376.

Dencik, Lina, Arne Hintz, and Jonathan Cable. 2016. Towards Data Justice? The Ambiguity of Anti-Surveillance Resistance in Political Activism. *Big Data & Society* 1 (12): 12. https://doi.org/10.1177/2053951716679678.

European Commission. 2017. *Article 29 Data Protection Working Party. WP 248 Rev.01*. https://ec.europa.eu/newsroom/document.cfm?doc_id=47711.

European Commission. 2018. Can I Be Subject to Automated Individual Decision-Making, Including Profiling? https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/my-rights/can-i-be-subject-automated-individual-decision-making-including-profiling_en.

European Commission. 2019. How Is Data on My Religious Beliefs/Sexual Orientation/Health/Political Views Protected? Policies, Information and Services. https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/how-my-personal-data-protected/how-data-my-religious-beliefs-sexual-orientation-health-political-views-protected_en.

Edionwe, Tolulope. 2017. The Fight against Racist Algorithms: Can We Teach Our Machines to Unlearn Racism? *The Outline*, https://theoutline.com/post/1571/the-fight-against-racist-algorithms?zd=2&zi=cfsot7bu.

Elwood, Sarah. 2007. Grassroots Groups as Stakeholders in Spatial Data Infrastructures: Challenges and Opportunities for Local Data Development and Sharing. *International Journal of Geographical Information Science* iFirst Article (January): 1–20.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Fuchs, Christian. 2017. From Digital Positivism and Administrative Big Data Analytics towards Critical Digital and Social Media Research! *European Journal of Communication* 32 (1): 37–49.

GDPR Hub. 2019. *CJEU - C-61/19 - Orange Romania SA v ANSP-DCP*. https://gdprhub.eu/CJEU_-_C-61/19_-_Orange_Romania_SA_v_ANSPDCP.

Gangadharan, Seeta Peña. 2012. Digital Inclusion and Data Profiling. *First Monday* 17 (5–7). http://firstmonday.org/article/view/3821/3199.

Gazi, Theodora. 2020. Data to the rescue: How humanitarian aid NGOs should collect information based on the GDPR, *Journal of International Humanitarian Action* 5.

Gitelman, Lisa, ed. 2013. *Raw Data Is an Oxymoron*. Cambridge, MA, London: The MIT Press.

Global Internet Forum to Counter Terrorism. 2010. About. GIFCT. https://www.gifct.org/about/.

Global Partnership for Sustainable Development Data. 2016. Data4SDGs Toolbox. http://www.data4sdgs.org/initiatives/data4sdgs-toolbox.

Goodman, Bryce, and Seth Flaxman. 2017. European Union Regulations on Algorithmic Decision Making and a 'Right to Explanation'. *AI MAGAZINE*.

Greenwood, Faine, Caitlin Howarth, Danielle Escudero Poole, Nathaniel A. Raymond, and Daniel P. Scarnecchia. 2017. The Signal Code: A Human Rights Approach to Information During Crisis. https://hhi.harvard.edu/publications/signal-code-human-rights-approach-information-during-crisis

GSMA. 2017. United Nations Foundation and GSMA Team Up to Support Data for Good. Newsroom. https://www.gsma.com/newsroom/press-release/united-nations-foundation-gsma-team-support-data-good/.

Gutierrez, Miren. 2019. Participation in a Datafied Environment: Questions about Data Literacy. *Comunicação e Sociedade* 36: 29–47.

Gutierrez, Miren. 2021. Algorithmic Gender Bias and Audiovisual Data: A Research Agenda. *International Journal of Communication* 15: 439–461.

Gutierrez, Miren, and Guillermo Gutierrez. 2019. Climate Finance: perspectives on Climate Finance from the Bottom Up. *Development* 62: 136–146. https://doi.org/10.1057/s41301-019-00204-5.

Gutierrez, Miren, and Marina Landa. 2021. From Available to Actionable Data: An Exploration of Expert and Reusers Views of Open Data. *Journal of Urban Technology* 36: 37–55. https://doi.org/10.1080/10630732.2021.1939626.

Gutierrez, Miren. 2018a. *Data Activism and Social Change*. Pivot. London: Palgrave Macmillan.

Gutierrez, Miren. 2018b. *Data Activism and Social Change*. Pivot. London: Palgrave Macmillan.

Halkort, Monika. 2019. Decolonizing Data Relations: On the Moral Economy of Data Sharing in Palestinian Refugee Camps. Canadian Journal of Communication, https://doi.org/10.22230/cjc.2019v44n3a3457

Haupt, Michael. 2016. 'Data Is the New Oil'—A Ludicrous Proposition. *Medium*, https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294.

Human Rights Watch (HRW). 2021. UN Shared Rohingya Data Without Informed Consent. 15 June https://www.hrw.org/news/2021/06/15/un-shared-rohingya-data-without-informed-consent

Igoe, Michael. 2018. David Beasley Takes on Global Hunger. *Devex*, https://www.devex.com/news/david-beasley-takes-on-global-hunger-92345.

Iliadis, Andrew. 2018. Algorithms, ontology, and social progress. *Global Media and Communication* 14 (2): 219–230.

Innerarity, Daniel. 2013. *The Democracy of Knowledge*. Bloomsbury Academic.

Innerarity, Daniel. 2016. Ricos y Pobres En Datos. *Globernance.Org* (blog). 22 February. http://www.danielinnerarity.es/opinion-preblog/ricos-y-pobres-en-datos/.

ITU. 2014. Measuring the Information Society Report. Geneva: International Telecommunication Union. https://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2014/MIS2014_without_Annex_4.pdf.

Jütting, Johannes. 2013. Is Big Data the New Oil Fuelling Development? Improving Lives through Better Statistics. Presented at the Paris21, Manila. http://nap.psa.gov.ph/ncs/12thncs/papers/PLENARY/PS%202/PS-2_2%20%20Is%20Big%20Data%20the%20New%20Oil%20Fuelling%20Development.pdf.

Kassam, N., E. Aghan, O. Aziz, H. Mbithe, K. Hameed, R. Shah, S. Surani, J. Orwa, and S. Somji. 2021. Factors Associated with Mortality Among Hospitalized Adults with COVID-19 Pneumonia at a Private Tertiary Hospital in Tanzania: A Retrospective Cohort Study. *International Journal of General Medicine* 14: 431–5440.

Kaurin, Dragana. 2021. Tech Localisation: Why the localisation of requires the localisation of technology. CDAC Background Paper http://www.cdacnetwork.org/tools-and-resources/i/20210930190855-5e47o

Kemp, Simon. 2020. Digital 2021: Global Overview Report. DataReportal. https://datareportal.com/reports/digital-2021-global-overview-report.

Kennedy, Helen. 2018. Living with Data: Aligning Data Studies and Data Activism Through a Focus on Everyday Experiences of Datafication. *Krisis* 1. https://krisis.eu/living-with-data/.

Klapper, Leora. 2019. Mobile Phones Are Key to Economic Development. Are Women Missing Out? *Brookings* (blog). https://www.brookings.edu/blog/future-development/2019/04/10/mobile-phones-are-key-to-economic-development-are-women-missing-out/.

Koffman, Tatiana. 2019. Blockchain - Africa Rising. *Forbes*. https://www.forbes.com/sites/tatianakoffman/2019/04/04/blockchain-africa-rising/#6fd065707711.

Kuner, Christopher. 2020. The GDPR and International Organisations. *American Journal of International Law* 114: 15–19.

Kwan, Campbell. 2019. Facebook, Microsoft, Twitter, YouTube Overhaul Counter-Terrorism Efforts for Christchurch Call. *NDNet*. https://www.zdnet.com/article/facebook-microsoft-twitter-youtube-overhaul-counter-terrorism-efforts-for-christchurch-call/.

Letouzé, Emmanuel. 2014. Big Data for Development: Facts and Figures. *Sci Dev Net*. https://www.scidev.net/global/data/feature/big-data-for-development-facts-and-figures.html.

Letouzé, Emmanuel. 2012. Big Data for Development: Opportunities and Challenges. *UN Global Pulse*. https://www.scidev.net/global/data/feature/big-data-for-development-facts-and-figures.html.

Letouzé, Emmanuel, and David Sangokoya. 2015. Leveraging Algorithms for Positive Disruption: On Data, Democracy, Society and Statistics. *Data-Pop Alliance*, 22 December.

Letouzé, Emmanuel. 2015. Big Data & Development: An Overview. DataPop Alliance. http://datapopalliance.org/wp-content/uploads/2015/12/Big-Data-Dev-Overview.pdf.

Lilja, Daniela, Sabrina Würmli, and Shamim Ahamed. 2020. How We Tried to Solve the COVID-19 Crisis in Two Days. https://www.helvetas.org/en/switzerland/how-you-can-help/follow-us/blog/highlights/developmenthack.

Lupton, Deborah. 2021. The Quantified Pandemic: Digitised Surveillance, Containment and Care in Response to the COVID-19 Crisis in S. Pink (ed.) *Everyday Automation: Experiencing and Anticipating Automated Decision Making*. Abingdon; Routledge

Mallol, Rainier. 2017. Why Data Is the New Oil. *Medium*, https://medium.com/@plus_socialgood/why-data-is-the-new-oil-2eb23c134243.

McNeill, Donald. 2021. Urban Geography 1: 'Big Tech' and the Reshaping of Urban Space. *Progress in Human Geography* 45 (5): 1311–1319.

Milusheva, Sveta, Daniel Björkegren, and Leonardo Viotti. 2021. Can We Trust Smartphone Mobility Estimates in Low-Income Countries? *World Bank Blogs* (blog). https://blogs.worldbank.

org/opendata/can-we-trust-smartphone-mobility-estimates-low-income-countries.

Mulder, Femke, Julie Ferguson, Peter Groenewegen, Kees Boersma, Jeroen Wolbers, and Jeroen Wolb. 2016. Questioning big data: crowdsourcing crisis data towards an inclusive humanitarian response, *Big Data and Society* July–December: 1–13.

Naughton, John. 2018. Magical Thinking about Machine Learning Won't Bring the Reality of AI Any Closer. *The Guardian* 5 August. https://www.theguardian.com/commentisfree/2018/aug/05/magical-thinking-about-machine-learning-will-not-bring-artificial-intelligence-any-closer.

Niklas, Jedrzej, and Seeta Peña Gangadharan. 2018. Between Antidiscrimination and Data: Understanding Human Rights Discourse on Automated Discrimination in Europe. London School of Economics and Political Science.

Oliver, Guy. 2014. NGOs against MONUSCO Drones for Humanitarian Work. *IRIN*, 23 July. http://www.irinnews.org/report/100391/ngos-against-monusco-drones-humanitarian-work.

Onuoha, Mimi. 2019. On Missing Data Sets. GitGub. Github.Com/MimiOnuoha. https://github.com/MimiOnuoha/missing-datasets.

Pandit, Harshvardhan J., Plamen Petkova, Declan O'Sullivan, and Dave Lewis. 2018. Investigating Conditional Data Value Under GDPR. Vienna: Procedia Computer Science. https://doi.org/10.5281/zenodo.3246486

Palmer, Lindsay. 2014. Ushahidi at the Google Interface: Critiquing the 'Geospatial Visualization of Testimony.' *Continuum* 28 (3): 342–356. https://doi.org/10.1080/10304312.2014.893989.

Parker, Ben. 2019. New UN Deal with Data Mining Firm Palantir Raises Protection Concerns. *IRIN*, https://www.irinnews.org/news/2019/02/05/un-palantir-deal-data-mining-protection-concerns-wfp.

Peng, Tony. 2017. LeCun vs Rahimi: Has Machine Learning Become Alchemy? *Synced*. https://syncedreview.com/2017/12/12/lecun-vs-rahimi-has-machine-learning-become-alchemy/.

Poell, Thomas, David B. Nieborg, and Jose van Dijck. 2019. Platformisation. *Concepts of the Digital Society*. https://doi.org/10.14763/2019.4.1425.

Rahimi, Ali. 2017. *Talk at the Conference on Neural Information Processing*. Long Beach, Ca. https://www.youtube.com/watch?v=Qi1Yry33TQE.

Rocher, Luc, Julien M. Hendrickx, and Yves-Alexandre. de Montjoye. 2019. Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models. *Nature Communications*. https://doi.org/10.1038/s41467-019-10933-3.

Roese, John. 2021. COVID-19 Exposed the Digital Divide. Here's How We Can Close It. World Economic Forum. https://www.weforum.org/agenda/2021/01/covid-digital-divide-learning-education/.

Schradie, Jen. 2015. 5 Reasons Why Online Big Data Is Bad Data for Researching Social Movements. *Mobilizing Ideas*, 4 February 4. https://mobilizingideas.wordpress.com/2015/04/02/5-reasons-why-online-big-data-is-bad-data-for-researching-social-movements/.

Smith, Andrew. 2018. Franken-Algorithms: The Deadly Consequences of Unpredictable Code. *The Guardian* 30 August. https://www.theguardian.com/technology/2018/aug/29/coding-algorithms-frankenalgos-program-danger.

Sørensen, Jannick Kirk, and Hilde Van Den Bulck. 2010. Public Service Media Online, Advertising and the Third-Party User Data Business: A Trade versus Trust Dilemma? *Convergence: the International Journal of Research into New Media Technologies* 26 (2): 421–447.

Spencer, Sarah W. 2021 Humanitarian AI: The hype, the hope and the future. Humanitarian Practice Network https://odihpn.org/resources/humanitarian-artificial-intelligence-the-hype-the-hope-and-the-future/

Sun, Ruonan, and Gregor B. Keating. 2015. Information Technology Platforms: Conceptualisation and a Review of Emerging Research in IS Research. Australasian Conference on Information Systems https://eprints.qut.edu.au/131200/1/__qut.edu.au_Documents_StaffHome_StaffGroupB%24_bakerj2_Desktop_Keating.pdf

Sylvester, Gerard. 2019. E-Agriculture in Action: Blockchain for Agriculture. Bangkok: Food and Agriculture Organisation and International Telecommunication Union. http://www.fao.org/3/CA2906EN/ca2906en.pdf.

The Economist. 2017. The World's Most Valuable Resource Is No Longer Oil, but Data. *The Economist*, 5 June. https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource.

Thomson Reuters Foundation. 2021. Digital Rights and Health Reporting Hub. https://www.trust.org/media-development/opportunities/?sfid=a053z00000vE4f0AAC&sfProgId=a153z00001FGD6YAAX&areaOfFocus=Governance%20and%20Human%20Rights.

Trevisan, Filippo. 2013. Social Engines and Social Science: A Revolution in the Making. Presented at the Economic and Social Research Council - Google Forum UK.

Tufekci, Zeynep. 2014a. Engineering the Public: Internet, Surveillance and Computational Politics. *First Monday* 19(7). http://firstmonday.org/article/view/4901/4097.

Tufekci, Zeynep. 2014b. Engineering the Public: Internet, Surveillance and Computational Politics. *First Monday* 19(7). http://firstmonday.org/article/view/4901/4097.

UN Chronicle. 2018a. A Decade of Leveraging Big Data for Sustainable Development. https://unchronicle.un.org/article/decade-leveraging-big-data-sustainable-development.

UN Chronicle. 2018b. A Decade of Leveraging Big Data for Sustainable Development. https://unchronicle.un.org/article/decade-leveraging-big-data-sustainable-development.

UN Global Pulse. 2013. Data for Development (D4d) Challenge at Net Mob 2013. https://www.unglobalpulse.org/D4D-NetMob.

UN Global Pulse. 2016. Risks, Harms and Benefits Assessment. United Nations. https://www.unglobalpulse.org/privacy/tools.

UNDESA. 2014. Securing Reliable Data for Development. DESA News. https://www.un.org/en/development/desa/newsletter/desanews/feature/2014/03/index.html#10113.

UNDP. 2020. Latest Human Development Index Ranking. United Nations Development Programme. http://hdr.undp.org/en/content/latest-human-development-index-ranking.

United Nations Development Group. 2017. Data Privacy, Ethics and Protection Guidance Note On Big Data For Achievement Of The 2030 Agenda. United Nations. https://undg.org/wp-content/uploads/2017/11/UNDG_BigData_final_web.pdf.

United Nations. 2021. As COVID-19 Exposes Global Disparities, Closing Digital Gap Key for Achieving Sustained Equitable Growth, Speakers Say as Social Development Commission Begins Annual Session. https://www.un.org/press/en/2021/soc4890.doc.htm.

van Dijck, Jose. 2014. Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology. *Surveillance & Society* 12(2): 197–208. https://doi.org/10.24908/ss.v12i2.4776.

Vinuesa, Ricardo, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark, and Francesco Fuso Nerini. 2020. The Role of Artificial Intelligence in Achieving The Sustainable Development Goals. *Nature Communications* 11 (233): 1–10. https://doi.org/10.1038/s41467-019-14108-y.

Wilton Park. 2019. Digital Dignity in armed conflict: a roadmap for principled humanitarian action in the age of digital transformation. https://www.wiltonpark.org.uk/wp-content/uploads/2020/09/WP1698-Report.pdf

Wes, Matt. 2017. Looking to Comply with GDPR? Here's a Primer on Anonymization and Pseudonymization. *International Association of Privacy Professionals* (blog). https://iapp.org/news/a/looking-to-comply-with-gdpr-heres-a-primer-on-anonymization-and-pseudonymization/.

Whittaker, Meredith, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Myers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. 2018. AI Now Report 2018. AI Now Institute, New York University. https://ainowinstitute.org/AI_Now_2018_Report.pdf.

World Food Programme. 2019. Palantir and WFP Partner to Help Transform Global Humanitarian Delivery. 2019. https://www1.wfp.org/news/palantir-and-wfp-partner-help-transform-global-humanitarian-delivery.

Wronkiewicz, Mark. 2018. Realistic Expectations for Applied Machine Learning. *Medium*, https://medium.com/devseed/realistic-expectations-for-applied-machine-learning-a8171250db28.

Yang, Kaifeng, and Sanjay K. Pandey. 2011. Further Dissecting the Black Box of Citizen Participation: When Does Citizen Involvement Lead to Good Outcomes? *Public Administration Review* 71 (6): 880–892. https://doi.org/10.1111/j.1540-6210.2011.02417.x.

Yao, Keping. 2018. Executive Development Course: Digital Government for Transformation Towards Sustainable and Resilient Societies – the Singapore Experience. Presented at the Laws and Regulations & Data Governance, UNDP Global Center for Public Service Excellence, Singapore. http://workspace.unpan.org/sites/Internet/Documents/UNPAN98156.pdf.

Zarsky, Tal. 2016. The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision-making. *Science, Technology and Human Values*. 41 (1): 118–132.

Zelenkauskaite, Asta, and Erik P. Bucy. 2016. A Scholarly Divide: Social Media, Big Data and Unattainable Scholarship. *First Monday* 21 (5). http://firstmonday.org/ojs/index.php/fm/article/view/6358/5511.

Zibuschka, Jan, Sebastian Kurowski, Heiko Roßnagel, Christian H. Schunck, and Christian Zimmermann. 2019. Anonymization Is Dead – Long Live Privacy. Garmisch-Partenkirchen.

Zuboff, Shoshana. 2018. *The Age of Surveillance Capitalism*. London: Profile Books.