Please cite the Published Version

Additional Information: This is an open access article published in Sexualities, by SAGE Publications.

Data like any other? Sexual and reproductive health, Big Data and the Sustainable Development Goals

Natalie Hammond ⓘ

Department of Social Care and Social Work, Manchester Metropolitan University, Manchester, UK

Angelo Moretti

Department of Methodology and Statistics, Faculty of Behavioural Sciences, Utrecht University, Utrecht, Netherlands

Abstract

This article examines the possibilities and pitfalls of using Big Data to address sexual and reproductive health concerns as related to the Sustainable Development Goals (SDGs), paying particular attention to contextual difference in development settings. The global datafication of sexual and reproductive life has taken place at great speed. However, evidential deficiencies and a lack of critical engagement of the specific issues around working with sexual and reproductive health Big Data in development contexts is apparent. Informed by critical data studies, and framed by a political economy perspective which calls attention to power structures, we seek to deepen our understanding of the role and challenges that Big Data around sexual and reproductive health in the Low and Middle-Income Countries can play in addressing the SDGs. First, we explore the ways in which sexual datafication processes produce Big Data. We then consider how such Big Data could directly contribute to addressing the SDGs beyond simply monitoring and evaluating. Next, we unpick how the sensitive and stigmatised nature of sexual and reproductive health can have ramifications in data-driven contexts where significant power asymmetries exist. By doing so, we provide a more nuanced articulation of the challenges of datafication by contextualising the stigma around sexual and reproductive in a datafied context. We argue that whilst Big Data in relation to sexual and reproductive

Corresponding author:
Natalie Hammond, Department of Social Care and Social Work, Manchester Metropolitan University, Bonsall Street, Manchester M15 6GX, UK.
Email: n.hammond@mmu.ac.uk

health shows potential to support the SDGs, there are specificities that must be considered to ensure that the push for data-driven approaches does no harm.

Keywords
Datafication, Big Data, Sustainable Development Goals, sexual and reproductive health

Introduction

Datafication processes, whereby complex human feelings, relationships and actions are transformed into digital data which often involves quantification (Lupton, 2016), and the resultant Big Data that is generated, provide opportunities for addressing the Sustainable Development Goals (SDGs) (Hassani et al., 2021; Wu et al., 2020). The datafication of sexual and reproductive life has taken place at great speed, and data is now being collected in a variety of contexts and forms such as via apps and Internet searches and from social media, electronic health records and online transactions. This provides a wide range of 'alternative' data sources that can be used to complement traditional sexual and reproductive health datasets. However, the role of datafication and Big Data in addressing sexual and reproductive health concerns in relation to the SDGs has been less well considered compared to other topics.

The importance of data has increased globally and data is often the key input into systems that are used to explain, manage, regulate and predict the world we live in; however, data are not neutral (Kitchin and Lauriault, 2014). Yet discourses around Big Data are contradictory with some suggesting much potential for Big Data to address key global challenges, whereas others argue critical concerns around data quality, privacy, bias and other issues must not be side-lined by such hype (Car et al., 2019; Kitchin, 2016; Scott, 2021). We contend that the specificities of data related to sexual and reproductive health, differ from other data for a range of complex reasons as will be explored below. Additionally, these specificities in relation to Low and Middle-Income Countries (LMIC) particularly, have been rendered invisible. In this paper, we are among the first to argue that the nuances of sexual and reproductive health data in LMIC require critical consideration to ensure that data-driven approaches are beneficial, appropriately resourced and do no harm. By doing so, we borrow from and extend Blumenstock's call for attention to context:

> In the rush to find technological solutions to complex global problems there's a danger of researchers and others being distracted by the technology and loosing track of the key hardships and constraints that are *unique to each local context*. Designing data-enabled applications that work in the real world will require a slower approach that pays much more attention to the people behind the numbers (2018: 70) (our emphasis added)

As Blumenstock notes, technology should not detract from social considerations and people must be brought centre stage. We seek to add a further layer to this, whereby the

context of the data itself must be considered, as there are significant consequences and risks when working with data around sensitive topics as we will demonstrate below.

The central question addressed in this paper is, what are the opportunities and challenges of working with Big Data in relation to sexual and reproductive health (SRH) in LMIC? We adopt a critical data studies approach which argues that data are not neutral objective or 'raw', but are instead contextual and situated (Dalton and Thatcher, 2014). The paper is organised as follows. First, we explore the ways in which sexual datafication processes are generating Big Data that could contribute to improving sexual and reproductive health around the world. Second, we consider how this Big Data could contribute towards addressing the SDGs focusing specifically on LMIC. Following this, borrowing from political economy approaches and calling attention to power differentials, we critically consider the challenges of working with SRH Big Data where we begin by detailing issues around infrastructure, skills and resources in LMIC where sexual and reproductive health may not be considered a development priority. To follow on, we present issues around inequalities, bias, discrimination, privacy and data quality where we argue that Big Data sets may exclude certain groups leading to biased conclusions for programme and policy development. Finally, we draw attention towards how the increased visibility that vast volumes of data bring, could compromise the identity and safety of marginalised communities. Thus crucial that significant efforts are made to prevent harms that technological developments may technological developments are not used to cause harm.

Bid data for development

A data revolution being driven by the increased volume of real-time data is taking place across LMIC as well as in more industrialised nations. Data are the basic inputs into processes that create information and knowledge, with knowledge providing the basis for understanding and explaining the world. Such understanding can be used to direct actions, inform policy development and exert influence over others (Kitchin, 2021). Consequently, whoever has access to high quality, large volumes of data, has a competitive and power advantage over those who do not. Understanding that knowledge, which creates influence and power is dependent on data, enables the attraction to Big Data to become clear. The use of Big Data in development contexts is now well considered in a range of domains such as transport, energy systems, health, aid distribution and disaster contexts (Agrewal and Prabakaran, 2020; Iliashenko et al., 2021; Heeks and Shekhar, 2019; Qadir et al., 2016; Sarker et al., 2020; Zhang et al., 2018). Big Data is an umbrella term referring to vast quantities of digital data that is continually being generated (UN Global United Nations Global Pulse, 2013). Big Data for development is thought to differ from the way the term Big Data is commonly used, however the reasons behind this are beyond the scope of this paper (see Ali et al., 2016). For this paper, we draw on the UN Global Pulse (2012) definition where Big Data for Development generally shares some, or all, of the following features:

(1) Digitally generated – that is, the data are created digitally (as opposed to being digitised manually) and can be stored using a series of ones and zeros, and thus can be manipulated by computers.

(2) Passively produced – a by-product of our daily lives or interaction with digital services.

(3) Automatically collected – that is, there is a system in place that extracts and stores the relevant data as it is generated.

(4) Geographically or temporally trackable – for example, mobile phone location data or call duration time.

(5) Continuously analysed – that is, information is relevant to human well-being and development and can be analysed in real-time. Here, 'real-time' does not always mean immediately and could be understood as information which is produced and made available in a relatively short and relevant period of time, and information which is made available within a timeframe that allows action to be taken in response, that is, creating a feedback loop. Importantly, it is the intrinsic time dimensionality of the data, and that of the feedback loop that jointly define its characteristic as real-time. (pg 15).

Such data can be split into user content generated through active engagement, and device generated data which can be passively collected (van Heerden, 2020). User generated content includes data generated from utilising social media including information from users posts as well as other data such as location and network information; from smartphone messaging apps; and through e-commerce and search engine data. Passively generated data, also known as exhaust data, are produced as a by-product of human computer interactions such as around phone usage, or that from bodily worn devices which record individuals activities or bodily processes. These types of data possess a shared characteristic; they are 'organic' meaning that they are by-products of processes and were not collected or sampled with the explicit intention of drawing conclusions, unlike traditional data collection instruments such as censuses or surveys (Laney, 2001).

What can Big Data do for the sustainable development goals?. The Sustainable Development Goals are a collection of 17 interlinked global goals, which aim to eradicate poverty, establish socioeconomic inclusion, improve health and protect the environment. Issues around SRH cut across multiple SDGs, particularly Goal 3: Ensure healthy lives and promote well-being for all at all ages and Goal 5: Achieve gender equality and empower all women and girls (UN, 2016). Big Data have the potential to provide timely information on a large scale, adding value by generating knowledge that can be further used to inform interventions from planning to implementation, as well as evaluations of development programmes and monitoring SDGs indicators (Letouzé, 2015; Lopes and Bailure, 2018; MacFeely, 2021; Silber, 2018). It is argued that Big Data can generate insights around health, well-being and quality of life that are missed by traditional data sources for example, Big Data can address gaps where some population groups are excluded from traditional data sources (Silber, 2018). Additionally, Big Data is said to 'represent

information about people's *behaviour* instead of information about their *beliefs*' (Letouzé, 2015: 4, our emphasis added). Social desirability bias, where respondents provide data in a way that they think will be viewed favourably, is a concern in SRH research and evidence gathering, due to the stigmatised and taboo nature of the topic. For example, respondents may deliberately answer questions inaccurately, either by underreporting stigmatised activities, or, by over reporting normative ones (Kelly et al., 2013; Rao et al., 2017). For example, Cullen (2020), using data from Nigeria and Rwanda argues that standard survey methods may significantly underestimate the prevalence of intimate partner violence (IPV). Additionally, the method used generates different rates with the most common method, face-to-face interviewing, showing the lowest IPV rates, and anonymous methods documenting the highest. Accordingly, efforts to address the difficulties of gathering accurate, complete and reliable SRH data are welcome and the attraction to Big Data is evident.

Whilst there is much fanfare about the promissory potential of Big Data, it is important to consider the digital landscape in LMIC to examine the feasibility and scope for collecting Big Data around any topic, not just in relation to SRH. There were approximately 5.27 billion unique mobile phone users (75% of these using a smartphone), about 4.72 billion Internet users (more than 60% of the world's total population) and approximately 4.33 billion active social media users (more than 55% of the global population) in April 2021 (Kemp, 2021). These numbers are increasing, with data suggesting that 332 million new users came online over a 12 month period, with the number of social media users increasing by 13.7% over the same period (Kemp, 2021). However, the distribution of Internet users is uneven, with developing and least developed countries as well as those in rural areas having reduced digital access (ITU, 2020). In LMIC, mobile phones are the primary means of internet access and nearly half of women access the internet this way (Rowntree, 2019). Thus, mobile phone data or data generated through mobile devices constitutes a valuable data source in development contexts, since it is the only digital technology used by most people in low-income groups (Kshetri, 2014). Additionally, many LMIC are beginning to digitise systems, for example, the District Health Information Software 2 (DHIS2) system is creating vast swathes of digital data for processing and analysis. That said, lack of affordability and the high cost of connectivity relative to income prevents access, and digital services in many least developed countries (LDCs) remain prohibitively expensive (ITU, 2021). Additionally, infrastructure issues around electricity, signal coverage and bandwidth negatively impact access to, and the use of, digital systems in LMIC (Houngbonon et al., 2021). Other factors such as low digital literacy, limited content in local languages as well as social gender-based norms where women are less likely to be online compared to men, play an important role in explaining limited digital integration (Bastion and Mukku, 2020; James,2019). Finally, a lack of in country capacity for data processing, poor data cultures and skills shortages for analysing Big Data, present challenges for integrating Big Data solutions into LMIC (Kalema and Mokgadi, 2017; Kshetri, 2014; Young et al., 2021). Therefore, whilst Big Data-driven approaches in LMIC are feasible, there are several generic access challenges which may hinder their uptake and success.

Sexual datafication, Big Data and SRH

The datafication of sexual and reproductive life has rapidly occurred due to the increased digital mediatisation of all domains of daily life. Datafication refers to the conversion of qualitative aspects of life into quantified data (Ruckenstein and Schüll, 2017). By exploring the ways in which sexual and reproductive life is digitally mediated, we can begin to see the vast opportunities for Big Data pertaining to SRH to be collected either, actively via user generation, or passively as an exhaust from other activities. For example, online spaces enable those on the fringes of society such as those belonging to minority sexual groups, those with certain fantasies or sex workers and clients to network, share experiences and advice and seek information (Hawkins and Watson, 2017; McDermott et al., 2015; Milrod and Monto, 2020; Noack-Lundberg et al., 2020; Carter et al., 2021; Randall and McKee, 2017). Commercial sex in its many forms is advertised, organised, performed and paid for via the digital realm (Hammond, 2015; Jones, 2015; Kingston et al., 2020; Sanders, et al., 2018; Sanders et al., 2020). Digital technologies such as online spaces and text messaging enable protective practices to be engaged in by sex workers (Bernier et al., 2021). The digital sphere however also opens opportunities for sexual abuse, exploitation and harassment, paedophilia and trafficking (Holt et al., 2010; Kloesss et al., 2017; Machimbarrena et al., 2018; Mandau, 2020; McGlyn, 2017; Ringrose et al., 2021).

Online retailers make up the largest market share of the sex toy market at 60%, and in some developing countries, online purchasing presents the only option due to legal concerns (Grand View Research, 2021). Pornhub, the world's largest provider of online porn, reports that 130 million people a day visited Pornhub and in 2020 mobile devices made up 84% of all Pornhub's traffic worldwide; 80% of that from smartphones, with tablet and laptop traffic seeing reductions (Pornhub, 2021). Whilst the USA, Japan and the UK are the top three traffic providers on Pornhub, several LMIC are on the top 20 list including Mexico, Brazil, Columbia and the Philippines (Pornhub, 2021). A variety of digital dating opportunities in the form of apps, online adverts or matchmaking websites catering for general interests as well as speciality preferences where users input their likes, dislikes and backgrounds have become popular (Reynolds, 2015). People use social media to share information via text and other media, such as images or videos about sexual and reproductive health from public health campaigns, to documenting birth stories and announcing pregnancy loss (Anbalibi and Forte, 2018; Gabarron and Wynn, 2016; Jones et al., 2019; Sanders, 2019). Additionally, people use online spaces, digital assistants and chatbots for locating general sex-based information as well as health information around contraception, abortion and accessing sexual and reproductive healthcare services (Courtenay and Baraister, 2021; Jerman et al., 2018; Mitchell et al., 2014; Nadarzynski et al., 2021; Patterson et al., 2019; Wilson, et al., 2017). Fertility support for both men and women via online forums or social media is also reported (Grunberg et al., 2018; Stenström and Pargman, 2021).

There are now multiple systems for booking and managing health appointments, as well as the use of digital online consultation's, the use of which accelerated during the COVID-19 pandemic (Kempton et al., 2020; Nadarzynski et al., 2017; Zhao et al., 2017). Health

management information systems (HMIS) are designed to manage healthcare data including systems that collect, store, manage and transmit a patient's electronic medical records (EMR) or systems supporting healthcare policy decisions. DHIS2, an open source HMIS used in 73 LMIC, covers approximately 2.4 billion people. DHIS2 has specific data work packages that support HIV and reproductive, maternal, newborn, child and adolescent health, providing vast quantities of data potentially covering approximately a third of the world's population. Prescription, insurance and pharmacy services all have online systems providing further information and valuable data to those with access (Aldughayfiq and Sampalli; Geissler J, 2021; Goundrey-Smith, 2018). The purchase and provision of condoms, the pill, HIV and other STI testing kits and home pregnancy tests can all be digitally mediated through an assemblage of state and private providers. (Ahmed-Little et al., 2016; Pai et al., 2021). Further self-care options include tracking apps around menstrual health and fertility (Lupton, 2015). Whilst not exhaustive, this list provides some insight into the volume and breadth of digital data captured globally pertaining to sexual and reproductive health. These transactions and interactions generate substantial volumes of both user generated and passively collected data which, when processed and analysed could reveal insights about people's sexual practices and behaviours, sexual and reproductive health and bodies. This extensive volume of data once processed, could provide opportunities for a range of stakeholders (e.g. commercial organisations, healthcare providers, researchers, NGO's and government or multinational organisations) to address SRH issues and work towards addressing the SDGs. In some cases, this work has begun as will be explored below.

The role of Big Data in sexual and reproductive health has been less well considered compared to other health domains, where it has been extensively argued that Big Data, such as hospital records, patients' medical records and digital data produced by devices part of the 'internet of things' can support the healthcare system in developed and LMIC (Dash et al., 2019; Wyber et al., 2015). However, there is an emergent body of work exploring the use of Big Data in HIV (Qiao et al., 2021). For example, research has explored using Internet search data (Young and Zhang 2018) the analysis of social media content (Cai et al., 2020, Stevens et al., 2020; van Heerden, 2020; Young 2015; Young et al., 2021), using electronic health records (Yang et al., 2021), the application of machine learning techniques to datasets (Weissman et al., 2021), web scrapping (Rennie et al., 2020) and using mobile phone data to understand HIV flows (Valdano et al., 2021). Other areas beyond HIV include using social media data to explore gender-based violence (Carlyle et al., 2019; XueJia et al., 2019), analysing data from a wide range of apps related to SRH such as menstruation, fertility and pregnancy (Barassi, 2017; Hamper, 2020; Lupton, 2015; Starling et al., 2018; Tatsumi et al., 2020), and the use of electronic health records (Simons and Kohn, 2019). Big Data has been used to predict post-partum depression and a range of reproductive and gynaecological molecular and cellular characterisations, physiological and physio-pathological insights and clinical outcomes (Khamisy-Farah et al., 2021; Moreira et al., 2019).

The discussion above is not intended to be a definitive guide of all data trails pertaining to SRH and further work mapping and categorising such data could be beneficial in working to address the challenges explored below. However, this snapshot evidences an increasing interest in the way that Big Datasets and associated techniques can be applied

to address issues around sexual and reproductive health. Nevertheless, there remain two deficiencies. First, the work is dispersed, and each study focuses on an individual topic be that HIV, post-partum disorders or GBV; there is no literature bringing together this body of work and addressing the collective challenges and concerns of working with data around sexual and reproductive health. Second, much work around Big Data and SRH is based in Higher Income Countries (see Van Herrden and Young, 2020); thus, the specificities and challenges faced in LMIC have been less well considered in relation to SRH. The following section will explore in what ways Big Data pertaining to SRH can work towards achieving the Sustainable Development Goals paying particular attention to the SRH dynamics in LMIC contexts.

Opportunities

There has been much hype around Big Data (see Wyber et al., 2015). It has been argued that Big Data promises a 'data deluge', from data-scarce to data-rich studies of 'detailed, interrelated, timely and low cost data – that can provide much more sophisticated, wider scale, finer grained understandings of societies and the world we live in' (Kitchin, 2013: 263). There is little disagreement that Big Data has the potential to support the SDGs; the scale of this benefit is however contested due to the challenges this presents. Big Data can assist in working towards the SDGs in two ways – first, by providing supplementary data to support monitoring targets related to SRH, and second, by providing data and evidence to enable better programming, services and initiatives to support SRH and outcomes as they relate to the SDGs, we discuss both of these below.

Measuring and monitoring. Big Data can be used to support the measurement and monitoring of the SDGs indicators related to SRH by combining Big Data sets from alternative data sources with traditional data sets to obtain newer insights. For example, if multiple data sources related to the same entity (e.g. individuals, households, communities, groups, etc.) are combined, information on behaviours, patterns and relationships on different dimensions of a phenomenon can be generated (Abreu Lopes, and Ballur, 2018; Silber, 2018). Since one of the important principles of the SDGs is to 'leave no one behind', data disaggregation plays an important role (Martinez, 2017). Disaggregation can be done by gender (e.g. in goal 5), ethnicity (e.g. in goal 10), income group or other relevant classifications. For example, in terms of SRH, the target could be finding the number of women diagnosed with a specific STD and belonging to a particular ethnic group living below the poverty line, providing crucial disaggregated information for monitoring and evaluation. SDGs indicators can be classified into three tiers (IAEG-SDGs, 2020). In Tier I, the methods to produce the indicators exist and the data is available. However, many SDGs indicators remain unavailable even at a national level, with Tier II and Tier III indicators being more problematic. The methodology for computing Tier II indicators exists, and is internationally established, however, the data is not regularly produced by countries. As a result, there are areas or regions for which the data are not fully available. Tier III indicators lack a methodology or international standards for their computation, a topic which was the focus of the 51st session of the UN

Statistical Commission which took place in 2020. Big Data approaches can support the development of indicators in the last two tiers, thus addressing persistent data gaps.

Traditional data sources for the measurement and evaluation of the SDGs, such as via large-scale national sample surveys on sexual reproductive health may be present, however, some groups in the population can be excluded. Consequently, Big Data can also fill gaps where information is not available. For example, mobile phone spending can be used as a proxy variable for income level in areas or groups in the population where this data is not available via other data sources. Traditional data sources may not be able to provide timely information since data may not be available every year, for example. Hence, information between the data collection periods may be needed for monitoring and evaluating the SDGs. Understanding temporal dynamics is crucial to monitoring SDGs indicators over time and improving quality of life. In this context, Big Data can be used to predict indicators, as well as develop and evaluate theories of change. It becomes possible to investigate whether the target phenomena are influenced by individual or systemic factors. For example, these can be helpful to predict change in social norms, which cannot be directly observed in traditional indicators. In addition, relationships found in Big Data can predict health or social indicators in geographical areas where traditional surveys have not been carried out, but correlated geospatial data is available (Abreu Lopes, and Ballur, 2018). Monitoring however, is insufficient to improve the SDGs, 'the measurement approach alone does not cover the whole spectrum of ways in and channels through which Big Data as an entirely new ecosystem could impact—contribute to or hamper—human progress as called for and measured by the SDGs' (Letouzé, 2015: np). Thus, whilst insights derived from Big Data can be used to measure some SDGsroindicators, the true potential for Big Data lies in its ability to assist progress towards achieving specific targets and thus contributing directly to the achievement of the SDGs (Perera-Gomez and Lokanathan et al., 2017), as will be explored below.

Directly contributing to the SDGs. This section will tease out some of the benefits specifically in relation to SRH that go beyond monitoring SDG indicators. Big Data can contribute to SRH programming, services and initiatives in a number of ways. Big Data could contribute to outcomes measured by the SDGs via non-policy actions, that is, by people using 'insights and suggestions derived from Big Data, such as Google Maps estimates, algorithmic recommendations of when to see a doctor, etc. These are largely unrelated to policies but remain in the realm of "applications"—ways in which Big Data helps "do stuff, concrete tasks, more effectively"' (Letouzé, 2015: np). For example, in the United States, algorithms run against Electronic Health Records have been used to identify patients at increased risk of HIV and alerting healthcare providers about patients who may benefit from PrEP with the aim of improving PrEP prescribing and thus preventing new HIV infections (Krakower et al., 2019). By preventing new HIV infection rates this could contribute towards achieving the SDG target 3.3.1, Number of new HIV infections per 1000 uninfected population, by sex, age and key populations. By providing real-time insights into the population SRH, this could enable targeted interventions for vulnerable groups to be developed and adapted quickly whilst in operation to achieve maximum benefit and reduce costs. Social marketing public health campaigns are often

used to increase testing for STI's. Digital campaigns can draw on data from web-based ad click through's alongside connecting clickthrough's to outcomes, for example, web-based clickthrough's to HIV/STI tests ordered online (Gilbert et al., 2019). Analysing real-time data in this context enables STI programme planners to understand the effectiveness of digital campaigns and to address deficiencies in targeted communications, improving the impact and efficiency of such approaches, thus reducing costs. mHealth, the use of mobile technologies and multimedia to fulfil health goals and provide support to healthcare delivery tasks (Nurmi, 2013), provides opportunities as passive and user generated data collected from these systems could be analysed to target specific groups. Interestingly, SRH mHealth initiatives in LMIC may help overcome barriers to accessing SRH, particularly those that are rooted in the social contexts surrounding sexuality such as regarding provider prejudice, discrimination, stigmatisation, fear of refusal, lack of privacy and confidentiality (Biddlecom et al., 2007). Barriers in accessing services in traditional ways create data gaps around certain groups, however, data from mHealth systems could plug such gaps enabling greater insights and more targeted programming. It is beyond the scope of this paper to cover every opportunity that Big Data presents for supporting the SDGs via a direct contribution; however, the narrative above and earlier in the paper demonstrates such potential. An exercise thoroughly mapping different domains where Big Data could contribute to SRH-related SDGs targets would be a useful follow-on endeavour to push this field forward.

Challenges

Despite this potential, we should remain cautious of the limitations of Big Data and below we call attention to the challenges around working with SRH Big Data. Critique around Big Data and health has focused on generic issues around self-selection bias, data quality, duplication of respondents and coverage issues (Abreu and Ballur, 2018; Hilbert, 2016; Maaroof, 2015). Furthermore, problems related to privacy and confidentiality have also been considered (Shlomo and Goldstain, 2015). However, through our discussion, we argue that SRH Big Data is not data like any other, in fact, there are a range of specific issues and challenges that arise due to the taboo and stigmatised nature of the topic that require reflection and considered actions to mitigate potential harms.

Political economy approaches can help explore such issues. Taking Collinson's definition, a political economy approach is 'concerned with the interaction of political and economic processes within a society: the distribution of power and wealth between different groups and individuals, and the processes that create, sustain and transform these relationships over time' (2003: 3). Thus, the analysis below calls attention towards power structures, highlighting potential winners and losers in the rapid datafication of SRH. Taylor and Broeders (2015) argue that the use of new communication technologies by those in LMIC is resulting in a shift in power from the state to corporations who gather, process and analyse this burgeoning digital data. As the volume of data increases, this leads to increased visibility for populations who were previously less surveyed. Re-sultantly, there is an increase in power to monitor and surveil for unregulated actors, giving corporations a new type of power and greater influence over the lives of

individuals. With this power comes the potential for misuse and the exclusion of smaller state actors with local knowledge and a contextual understanding of the vast volume of data. They and others (Hayes, 2017), warn a lack of accountability may risk experimentation by big tech who may be attracted to LMIC as they offer the opportunity to test their solutions in a real world context. Additionally, Big Data analytics may in fact fail to address the structural conditions which frame challenges in LMIC (Talyor and Broeders, 2015) such as HIV, poverty, inequalities, and gender-based violence. Distributed governance creates a power to represent data subjects whereby individuals lose the ability to control how they are represented as data flows around increasingly large and complex systems of stakeholders across geographical boundaries. Within this complex system of power imbalances, data itself becomes the driver for the collection and processing of data, and function creep occurs (see Hayes, 2017). Corporations thus accumulate increasing volumes of data enhancing their power to profile and render visible for counting, sorting, monitoring and intervening with, those who were previously unseen or those who wish to remain at the periphery of surveillance (Taylor and Broeders, 2015).

Capacity and resources. There are two interrelated issues in relation to capacity and resources. In some instances, in LMIC care related to SRH is chronically underfunded with significant unmet need remaining (Pathak and Tariq, 2018). Whilst HIV has seen a relatively large amount of funding, this fails to meet the demand and SRH is often not a funding priority (Schäferhoff et al., 2019). Stigma related to SRH arises at multiple levels, including governmental, societal and individual. Such stigma, rooted in, and perpetuated by, patriarchal desires to control women's decision-making and bodies and moral ideas around 'good' and 'bad' sexual behaviour and desires, has the potential to influence policy, funding and programming. Despite the growth of donor aid from 2002 to 2017 in some east African countries, resulting in a period of improved indicators (Kibira et al., 2021), UNAIDS (2020) reports that increases in resources for HIV in LMIC stalled in 2017, with funding decreasing by 7% between 2017 and 2019. Furthermore, financial resources from donor aid are not evenly distributed with HIV receiving the most funding support and other aspects such as maternal health, abortion and family planning receiving a much smaller allocation in comparison (Schäferhoff et al., 2019). Harnessing the potential from SRH Big Data requires increasing investment by donors and governments, alongside more effective approaches that strengthen foundational data systems and governance frameworks and support local knowledge and capacity development which may meet challenges due to the stigma of the topic. In countries where laws prevent access to abortion or criminalise homosexuality, financial (and social) support for Big Data initiatives that recognise and address the needs of women and other minority groups may not be forthcoming. For example, the Global Gag Rule, issued by the Trump administration sought to curtail access to abortion. They sought to ban foreign NGOs that received US Government family planning assistance from using funds, from any source to provide abortion services, counselling, or referrals, or to advocate for the liberalisation of their country's abortion laws, including for the first time, HIV funding through the President's Emergency Plan for AIDS Relief (PEPFAR) (Priyanka, 2019). Whilst the Global Gag Rule has been rescinded, policies and funding requirements could impact on

the ability of using Big Data initiatives in SRH contexts. The growing involvement of a complex web of stakeholders in Big Data analytics increases those who have power over citizens, with many tech firms coming from higher income countries. Some of these may have a particular ideology to embed creating risks. For example, reports from the US reveal the use of mobile data (normally used in marketing of consumer goods or services) alongside mobile geo-fencing, to target women attending abortion clinics with anti-abortion advertisements (Coutts, 2016). This creates risks for women at a time of vulnerability who may be seeking sensitive health care. Such surveillance enables the sharing of names and addresses of women seeking abortion care, and those who provide it, with anti-choice groups (Coutts, 2016). How long before such tactics, driven by big tech companies from higher income states, become embedded in LMIC? Additionally, SRH will have to work against other donor priorities such as climate change, COVID-19 and its impacts, and increased support for economically productive sectors. Additionally, the British reduction in Official Development Assistance (ODA) has the potential to impact funding for SRH. Thus, Big Data initiatives including leveraging capacity or developing infrastructure specifically around SRH in LMIC, where there are competing priorities and many of the SDGs remain unmet, alongside the context whereby the rights of women and other marginalised groups are rarely prioritised, may meet social, political and financial resistance limiting their effectiveness.

Privacy and risks. Issues of privacy and safety around Big Data have been gaining recognition (Maaroof, 2015). In parallel, issues around data protection in LMIC have increasingly been recognised. For example, the rapid withdrawal of Allied forces from Afghanistan saw the swift resurgence of the Taliban. With this came fears around the compromised safety of data that could be used to help identify Afghans who had supported coalition forces, with grave consequences (Hu, 2021). The theft of over a half a million records concerning missing people and their families, detainees and other people receiving services from the Red Cross and Red Crescent Movement drew worldwide attention and condemnation (ICRC, 2022), demonstrating that no data is considered out of bounds. It is not only data breaches that risk privacy with reports that data collected about refugees may be used for other purposes such as counter terrorism or migration management (see Hayes, 2017). Such instances emphasise the risks of increased visibility that come with increasing datafication. Institutional frameworks to protect privacy may not be in lieu. This has important consequences, especially in a development context. Citizens may lack understanding of the multitude of ways that they become entangled with data as they live their daily lives and how this is appropriated by others to create advantages (Smith, 2018) and even when the datafication of services is recognised, distrust remains (Steedman et al., 2020). Furthermore, individuals might be unaware of consenting to data collection, and access to welfare may be contingent on agreeing to digital data collection and the resultant surveillance (Cukier and Mayer-Schoenberger, 2013; Holloway et al., 2021).

Consideration of legal frameworks and ethics must be in place to protect data sharing processes (Maaroof, 2015). As the International Development world becomes more data-driven, the generation of profiles and the use of complex targeting and eligibility

assessments to identify and provide better provisions for service users, necessitates increased data collection around individuals, groups or spaces (Hayes, 2017). Thus, the volume of data being collected increases visibility and enhances opportunities for datasets to be linked, presenting the risk of identification (see Shlomo and Goldstein, 2015). Visibility, identification and other privacy related issues in the field of SRH add an additional layer of complexity. Sexual and reproductive health is considered a significantly private domain of health and wider life. Some HIV positive people report being shunned by local communities or experiencing high levels of violence, homosexuality or involvement in sex work also brings stigma, violence and criminal sanctions or worse, and in Somali and Nigeria the death penalty stands in some regions for homosexuality (Jjuuko and Tabengwa, 2018). Details about fertility, or pregnancy related concerns are private and abortion remains criminalised in several localities (Jain 2019; Larrea et al., 2021). Thus, the use of Big Data in relation to SRH issues presents very specific and critical concerns around privacy. In Kenya, concerns were raised around the use of biometrics in HIV research, highlighting that function creep, whereby data gathered for health purposes could be used by the police to target key populations for arrest (KELIN, 2018). Thus, it is essential that the correct protections are put in place to protect Big Data related to SRH, and that protective processes are applied at all stages from collecting and storing of data, to processing and decision-making across all stakeholders. Additionally, wider work around digital literacies is vital to ensure that citizens understand how their data will be collected and used, as well as their data rights when engaging with all digital systems but especially in SRH contexts. Community involvement in Big Data initiatives is essential to safely develop systems. Furthermore, with the risk of increased visibility, questions remain about the ways in which those from marginalised communities may seek to protect themselves and disengage from digital systems, or seek to minimise their digital trails. The impact such unintended consequences have on the ability to receive welfare, health care and other forms of support requires urgent attention.

Bias and discrimination. Bias and discrimination in alternative data sources and algorithms have received particular interest. Banks (2018) has written in depth about how Big Data systems have developed in sophistication, and are used as forces for control, manipulation, and punishment, constricting poor and working-class people's opportunities, resulting in discrimination. Issues around race, gender and disability (Obermeyer et al., 2019; Packin, 2021) are reported, Bemjamin (2019) demonstrates that technology is developed within a racist context and generates information that, when (mis)used, exacerbates inequities for those already marginalised. Bias algorithms learn from bias data, thus, consideration around the ways that data may be biased is a fundamental issue to consider in relation to SRH. Bias can enter at many points; access to technologies that generate Big Data can be curtailed due to economic reasons, infrastructure, and other social divisions such as gender or disability. Resistance at local levels or lack of funding from donor organisations to support capacity building and infrastructure as well as live projects as explained above, may mean the Global North dominates in the Big Data projects that are conducted. Colonial history offers important insights into how external labelling, analysis, measurement, and systems can result in inhumane and callous effects

(Mawere and Van Stam, 2020). The Western domination of digital platforms and data processing create systems and algorithms that are removed from their context and are designed and implemented by those who understand the technology, but lack awareness about the cultural context. Thus, when using data to drive decision-making, planners and decision makers should consider who has access to these technologies, who is absent from the data and the impacts of such exclusions (UN Women, 2018). It is essential that local actors and social scientists who can provide contextual knowledge play a pivotal role in teams which a currently dominated by data scientists (Taylor and Broeders, 2015).

Digital data availability is heterogeneous across borders; for example, countries characterised by larger mobile phone and Internet penetration rates will generate data that is more directly produced by people. In contrast, in countries where there are large aid communities, data will be more programme-related. There are also differences across socio-demographic and geographical characteristics and differences within borders themselves (Maaroof, 2015; Hoi-Wai, 2014). Hoi-Wai (2014) stresses that it is fundamental to investigate the biases generated when Big Data influences policies. Particularly, attention should be given to those countries that produce a smaller amount of data or have less capacity in Big Data to avoid adding issues across place based digital divides. The digital divide can operate across other lines particularly those related to other forms of exclusion. Those belonging to minority groups, for example, LGBT + communities and sex workers may experience digital exclusion. For example, there may be reticence to post on social media about sexual orientation, or individuals may be excluded from healthcare facilities due to stigma resulting in their electronic health records being excluded from datasets which are used for decision-making purposes (Ganesh et al., 2016; Kim et al., 2018). This leads to bias data due to the under representation of those groups. Furthermore, those that are free to post on social media about their LGBT + or sex work status do not represent all views or positions and may be in a privileged position compared to those who seek to maintain a hidden identity, or those who experience other intersectional stigmas and inequalities around class, race or disability, for example. Others may seek to deliberately self-exclude for other reasons. For example, Heeks (2021) reports that refugees and migrants may be unwilling to use digital systems for fear of repercussions. These communities are vulnerable to sexual violence and face a range of challenges related to sexual and reproductive health (Amiri et al., 2020; Krause, 2020). Despite limited evidence around digital self-exclusion, it becomes apparent that self-exclusion from digital systems may lead to planning and programming decisions being made using data where the most vulnerable are absent, creating further exclusions which may contribute to worsening SRH outcomes. Thus, whilst some members of marginalised communities may benefit from digital technologies and become included in Big Data datasets to aid appropriate programming, those who are often the *most* marginalised of the marginalised continue to remain excluded from datasets and thus interventions, widening inequalities.

Big data has been the object of social critique (see e.g. Boyd and Crawford, 2012). Some of these debates relate to LGBT + lives. For instance, we refer to the Jen Jack Gieseking's call for a 'queer feminist approach to the scale of big data'. Here, it is argued that the characteristics of big data undermine the importance of communities that are

'small' due to histories of discrimination and violence (Gieseking, 2018). Other work, for example, McGlotten (2016) details the challenges of big data related to non-white LGBT + people. The binary construction of data is a diffused practice in data collection strategies both in Big Data and traditional survey contexts. This leads to a simplification of how sexuality might be identified (see Guyan, 2022). Data collected via Big Data, such as Facebook, may offer users a range of gender options, which work towards inclusivity. However, allowing users to choose only one identity option, is insufficient to account for the multiple, and also overlapping 'experiences of self' that may characterise queer identity (Ruberg and Ruelos 2020).

Despite claims that online data reflects people's beliefs more than data collected via survey's etc., online spaces are heavily curated (Tiggemann and Anderberg, 2021) thus queries about the accuracy of data collected should remain at the fore. Exhaust data, such as, GPS data may not be curated in the way that online social media posts are. For example, some individuals may seek to deliberately hide or present a different identity due to stigma, violence or fear of criminal sanction. Tech-related violence impacts the data generated and limits access and participation and prevents freedom of online expression (Shephard, 2016) creating inherent biases in data. Increased visibility that comes with the inclusion in datasets presents challenges for those from minority sexual groups with some members maintaining two social media accounts, one straight and one queer ((Maya Indira Ganesh et al., 2016; see also Shephard, 2016). Thus, datasets generated from social media data should be used with caution and the inherent biases around access and the way in which the digital data has been curated should remain at the fore.

Conclusion

In this paper, we have highlighted the opportunities and challenges of working with Big Data in the context of sexual and reproductive health to address the SDG. Whilst there are clear opportunities for Big Data to contribute to the SDGs beyond simply monitoring and measuring, the field of SRH presents nuances that must be considered if Big Data can be utilised to safely realise such promissory opportunities. Big Data solutions and approaches cannot be understood purely from the technical domain and it is inadequate to apply techniques and processes across multiple areas without considering the specificities of the topic in local contexts. A political economy approach highlights the power asymmetries, demonstrating the risks to vulnerable communities. Thus, we need collaborative and multi-disciplinary approaches that challenge such asymmetries where technical insights are combined with local knowledge and contextual familiarity. As we have argued, Big Data in the context of SRH, is not data like any other and while this paper is not intended to be a conclusive document for the safe, ethical and effective use of Big Data in SRH, in fact this is just the beginning, it is intended to stimulate dialogue and provoke thought among those seeking to draw on Big Data in SRH contexts. As Big Data as a field itself matures and grows in development contexts, more will become known about the usefulness and impacts of such approaches. However, in the meantime planners, policy makers, programme managers, technology providers and researchers seeking to use SRH Big Data in development contexts should be aware of the ways that SRH data

differs to other data. Thus, working in partnership with communities, they should seek to address such issues in their work to help improve well-being alongside mitigating for negative consequences. We believe that failing to take heed of the issues outlined above would be negligent, unethical and could lead to catastrophic consequences for the most marginalised.

Declaration of conflicting interests

Funding

ORCID iD

Natalie Hammond https://orcid.org/0000-0002-3340-2992

References

Abreu Lopes C and Ballur S (2018) Gender equality and big data: UN women innovation facility. Available at: https://unglobalpulse.org/wp-content/uploads/2018/03/Gender-equality-and-big-data-en-2018.pdf

Agrawal R and Prabakaran S (2020) Big data in digital healthcare: lessons learnt and recommendations for general practice. *Heredity* 124: 525–534.

Ahmed-Little Y, Bothra V, Cordwell D, et al. (2016) Attitudes towards HIV testing via home-sampling kits ordered online (RUClear pilots 2011–12). *Journal of Public Health (Oxford, England)* 38(3): 585–590.

Ali A, Qadir J, Rasool R, et al. (2016) Big data for development: applications and techniques. *Big Data Analytics* 1: 2–24.

Aldughayfiq B and Sampalli S (2021) Digital health in physicians and pharmacists office: a comparative study of e-prescription systems architecture and digital security in eight countries. *OMICS: A Journal of Integrative Biology* 25 (2): 102–122.

Amiri M, El-Mowafi IM, Chahien T, et al. (2020) An overview of the sexual and reproductive health status and service delivery among Syrian refugees in Jordan, nine years since the crisis: a systematic literature review. *Reproductive Health* 17: 166. DOI: 10.1186/s12978-020-01005-7.

Andalibi N and Forte A (2018) Announcing pregnancy loss on Facebook: a decision-making framework for stigmatized disclosures on identified social network sites. *Proceedings Of the 2018 CHI Conference On Human Factors In Computing Systems* 158: 1–14. Jan-14 2018.

Barassi V (2017) BabyVeillance? Expecting parents, online surveillance and the cultural specificity of pregnancy apps. *Social Media and Society* 3(2): 05631E+15.

Bastion GD and Mukku S 2020. Data and the Global South: Key issues for inclusive digital development. Available at: https://us.boell.org/sites/default/files/2021-01/20201216-HB-broschure-dataandglobalsouth-A4-01.pdf [Last accessed 02/03/23]

Bemjamin R (2019) *Race After Technology*. Cambridge: Polity.

Bernier T, Shah A, Ross LE, et al. (2021) The use of information and communication technologies by sex workers to manage occupational health and safety: scoping review. *Journal of Medical Internet Research* 23(6): e26085.

Biddlecom AE, Singh S, Munthali A, et al. (2007) Adolescents' views of and preferences for sexual and reproductive health services in Burkina Faso, Ghana, Malawi and Uganda. *African Journal of Reproductive Health* 11(3): 99–110.

Boyd D and Crawford K (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society* 15(5): 662–679.

Blumenstock J (2018) Don't forget people in the use of big data for development. *Nature* 561: 170–172.

Cai M, Shah N, Li J, et al. (2020) Identification and characterization of tweets related to the 2015 Indiana HIV outbreak: a retrospective infoveillance study. *PloS One* 15(8):e0235150.

Car J, Sheikh A, Wicks P, et al. (2019) Beyond the hype of Big Data and artificial intelligence: building foundations for knowledge and wisdom. *BMC Medicine* 17(1): 143–145.

Carlyle KE, Guidry JPD, Dougherty SA, et al. (2019) Intimate partner violence on instagram: visualizing a public health approach to prevention. *Health Education and Behavior : The Official Publication of the Society for Public Health Education* 46(2): 90S–96S.

Carter P, Gee M, McIlhone H, et al. (2021) Comparing manual and computational approaches to theme identification in online forums: a case study of a sex work special interest community. *Methods in Psychology* 5: 100065.

Collinson S (ed), (2003) *Power, livelihoods and conflict: case studies in political economy analysis for humanitarian action. Humanitarian Policy Group*. London: Overseas Development Institute. Available at: http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/289.pdf [Last accessed 02/06/2022].

Courtenay T and Baraitser P (2021) Online contraceptive discussion forums: a qualitative study to explore information provision. *BMJ Sexual and Reproductive Health* 47(3): e5.

Coutts S (2016) Anti-Choice Groups Use Smartphone Surveillance to Target 'Abortion-Minded Women' During Clinic Visits. Available at: https://rewirenewsgroup.com/article/2016/05/25/anti-choice-groups-deploy-smartphone-surveillance-target-abortion-minded-women-clinic-visits/ [Last accessed 06/06/2022].

Cukier KN and Mayer-Schoenberger V (2013) The rise of Big Data. *Foreign Affairs* 92 (3): 28–40.

Cullen C (2020) Method matters: underreporting of intimate partner violence in Nigeria and Rwanda. *World Bank Policy Research Working Paper (9274)*.

Dalton CT and Thatcher J (2014) What does a critical data studies look like, and why do we care? Available at: https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care [Last accessed 07/10/21].

Dash S, Shakyawar SK, Sharma M, et al. (2019) Big data in healthcare: management, analysis and future prospects. *Journal of Big Data* 6(53): 54.

Gabarron E and Wynn R (2016) Use of social media for sexual health promotion: a scoping review. *Global Health Action* 9(1): 32193.

Ganesh MI, Deutch J and Schulte J (2016) *Privacy, Anonymity, Visibility: Dilemmas in Tech Use by Marginalised Communities*. Brighton: IDS.

Geissler J (2021) Sustainable upscaling: the role of digitalization in providing health care and health insurance coverage in developing countries. In: Herberger T and Dötsch J (eds). *Digitalization, Digital Transformation and Sustainability in the Global Economy. Springer Proceedings in Business and Economics*. Cham: Springer, pp. 53–69.

Gieseking JJ (2018) Size matters to lesbians, too: Queer feminist interventions into the scale of big data. *The Professional Geographer* 70(1): 150–156.

Gilbert M, Salway T, Haag D, et al. (2019) Assessing the impact of a social marketing campaign on program outcomes for users of an internet-based testing service for sexually transmitted and blood-borne infections: observational study. *Journal of Medical Internet Research* 21(1): e11291.

Goundrey-Smith S (2018) The connected community pharmacy: benefits for healthcare and implications for health policy. *Frontiers in Pharmacology* 9: 1352.

Grand View Research (2021) Sex toys market size, share and trends analysis report by type (male, female), by distribution channel (e-commerce, specialty stores, mass merchandizers), by region and segment forecasts, 2021–2028. Available online at: https://www.grandviewresearch.com/industry-analysis/sex-toys-market [Last accessed 06/06/2022].

Grunberg PH, Dennis CL, Da Costa D, et al. (2018) Infertility patients' need and preferences for online peer support. *Reproductive Biomedicine and Society Online* 6: 80–89.

Guyan K (2022) *Queer Data. Using Gender, Sex and Sexuality Data for Action*. London: Bloomsbury.

Hamper J (2020) 'Catching ovulation': exploring women's use of fertility tracking apps as a reproductive technology. *Body and Society* 26(3): 3–30.

Hammond N (2015) Men who pay for sex and the sex work movement? Client responses to stigma and increased regulation of commercial sex policy. *Social Policy and Society* 14(1): 93–102.

Hassani H, Huang X, MacFeely S, et al. (2021) Big data and the United Nations sustainable development goals (UN SDGs) at a glance. *Big Data and Cognitive Computing* 5(3): 28–28.

Hawkins B and Watson RJ (2017) LGBT cyberspaces: a need for a holistic investigation. *Children's Geographies* 15(1): 122–128.

Hayes B (2017) Migration and data protection: Doing no harm in an age of mass displacement, mass surveillance and "big data"doing no harm in an age of mass displacement, mass surveillance and "big data". *International Review of the Red Cross* 99(904):179–209.

Heeks R (2021) The rise of digital self-exclusion.Available online at: https://ict4dblog.wordpress.com/2021/08/03/the-rise-of-digital-self-exclusion/ [Last accessed 29/09/2021].

Heeks R and Shekhar S (2019) Datafication, development and marginalised urban communities: an applied data justice framework. *Information, Communication and Society* 22(7): 992–1011.

Hilbert M (2016) Big data for development: a review of promises and challenges. *Development Policy Review* 34(1): 135–174.

Hoi-Wai JC (2014) Big data for development in China. *UNDP China*.

Holloway K, Al Masri R and Abu Yahi A (2021) *Digital Identity, Biometrics and Inclusion in Humanitarian Responses to Refugee Crises*. London: ODI-HPG. Available at: https://cdn.odi.org/media/documents/Digital_IP_Biometrics_case_study_web.pdf [Last accessed 15/06/2022].

Holt TJ, Blevins KR and Burkert N (2010) Considering the pedophile subculture online. *Sexual Abuse: A Journal of Research and Treatment* 22(1): 3–24.

Houngbonon GV, Le Quentrec E and Rubrichi S (2021) Access to electricity and digital inclusion: evidence from mobile call detail records. *Humanities and Social Sciences Communications* 8(1): 170–211.

Hu M (2021) The Taliban reportedly have control of US biometric devices a lesson in life-and-death consequences of data privacy. Available at: https://theconversation.com/the-taliban-reportedly-have-control-of-us-biometric-devices-a-lesson-in-life-and-death-consequences-of-data-privacy-166465 [Last accessed 06/06/2022].

IAEG-SDGs (2020) Tier classification for global SDG indicators (as of 28 December 2020). *Unstats*. Available at: https://unstats.un.org/sdgs/files/TierClassificationofSDGIndicators_28Dec2020_web.pdf [Last accessed 09/09/21].

ICRC (2022) Hacking the data of the world's most vulnerable is an outrage. Available at: https://www.icrc.org/en/document/hacking-data-outrage [Last accessed 13/06/2022].

Iliashenko O, Iliashenko V and Lukyanchenko E (2021) Big data in transport modelling and planning. *Transportation Research Procedia* 54: 900–908.

ITU (2020) Measuring digital development facts and figures 2020. Available at: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf.[Last accessed 09/09/21].

ITU (2021) The affordability of ICT services. Available at: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/prices2020/ITU_A4AI_Price_Briefing_2020.pdf.[Last accessed 09/09/21].

Jain D (2019) Time to rethink criminalisation of abortion? towards gender justice approach. *NUJS Law Review.* 12(1),21–42.

James J (2021) Confronting the scarcity of digital skills among the poor in developing countries. *Development Policy Review* 39(2): 324-339.

Jerman J, Onda T and Jones RK (2018) What are people looking for when they Google "self-abortion"? *Contraception* 97(6): 510–514.

Jjuuko A and Tabengwa M (2018) Expanded criminalisation of consensual same-sex relations in Africa contextualising recent developments. In: Nikol N, Jjuko A, Lusimbo R, et al. (eds), *Envisioning Global LGBT Human Rights:(Neo) Colonialism, Neoliberalism, Resistance and Hope*. London: University of London, pp. 63–96.

Jones K, Williams J, Sipsma H, et al. (2019) Adolescent and emerging adults' evaluation of a Facebook site providing sexual health education. *Public Health Nursing* 36(1): 11–17.

Jones A (2015) Sex work in a digital era. *Sociology Compass* 9(7): 558–570.

Mathias Kalema B and Mokgadi M (2017) Developing countries organizations' readiness for Big Data analytics. *Problems and Perspectives in Management* 15: 260–270.

Kelly CA, Soler-Hampejsek E, Mensch BS, et al. (2013) Social desirability bias in sexual behavior reporting: evidence from an interview mode experiment in rural malawi. *International Perspectives on Sexual and Reproductive Health* 39(1): 14–21.

Kemp S (2021) Digital 2021 April statshot report-datareportal-global digital insights. https://datareportal.com/reports/digital-2021-april-global-statshot [Last accessed 10.10.21].

Kempton M, Rees J and Edwards A (2020) I just called to say: results of a patient satisfaction survey evaluating the use of telemedicine for consultations within the Oxfordshire integrated sexual health service. *International Journal of STD and AIDS*. 31(Suppl 12): 81–82.

Khamisy-Farah R, Furstenau LB, Kong JD, et al. (2021) Gynecology meets big data in the disruptive innovation medical era: state-of-art and future prospects. *International Journal of Environmental Research and Public Health* 18(10): 5058.

Kibira D, Asiimwe C, Muwonge M, et al. (2021) Donor commitments and disbursements for sexual and reproductive health aid in Kenya, Tanzania, Uganda and Zambia. *Frontiers in Public Health* 9: 645499.

Kim HY, Grosso A, Ky-Zerbo O, et al. (2018) Stigma as a barrier to health care utilization among female sex workers and men who have sex with men in Burkina Faso. *Annals of Epidemiology* 28(1): 13–19.

Kingston S, Hammond N and Redman S (2020) *Women Who Buy Sex: Converging Sexualities?* London: Routledge.

Kitchin R (2013) Big data and human geography: opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.

Kitchin R (2016) Big data. *International Encyclopedia of Geography: People, the Earth, Environment and Technology*. Wiley-Blackwell. pp. 1–3.

Kitchin R (2021) *The Data Revolution: A Critical Analysis of Big Data, Open Data and Data Infrastructures*. London: Sage.

Kitchin R and Lauriault T (2014) Towards critical data studies: charting and unpacking data assemblages and their work. Available at: https://papers.ssrn.com/sol3/papers.cfm?Abstract_id=2474112 [Last accessed 10.10.21].

KELIN, The Kenya Key Populations Consortium (2018) Everyone said no: biometrics, HIV and human rights: a Kenya case study. Avaiable at: www.kelinkenya.org/wp-content/uploads/2018/07/"Everyone-said-no".pdf [Last accessed 15/06/2022].

Kloess J, Seymour-Smithe S, Hamilton-Giachritsis C, et al. (2017) A qualitative analysis of offenders' modus operandi in sexually exploitative interactions with children online. *Sexual Abuse*. 29(6): 563–591.

Krakower DS, Gruber S, Hsu K, et al. (2019) Development and validation of an automated HIV prediction algorithm to identify candidates for pre-exposure prophylaxis: a modelling study. *The Lancet HIV* 6(10):e696–e704.

Krause U (2020) Violence against women in camps? exploring links between refugee camp conditions and the prevalence of violence. In: Crepaz K, Becker U and Wacker E (eds), *Health in Diversity-Diversity in Health*. Wiesbaden: Springer.

Kshetri N (2014) The emerging role of Big Data in key development issues: opportunities, challenges, and concerns. *Big Data and Society* 1(2): 205395171456422.

Laney D (2001) *3-D Data Management: Controlling Data Volume, Velocity and Variety. META group research note, Feb 6th*. Stamford: Gartner. Available online: http://gtnr.it/1bKflKH [Last accessed 01.10.21].

Larrea S, Assis MP and Mendoza CO (2021) "Hospitals have some procedures that seem dehumanising to me": experiences of abortion-related obstetric violence in Brazil, chile and ecuador. *Agenda* 35 (3): 54–68.

Letouzé E (2015) Thoughts on Big Data and the SDGs. Available at: https://sdgs.un.org/sites/default/files/documents/7798BigData%2520-%2520Data-Pop%2520Alliance%2520-%2520Emmanuel%2520Letouze.pdf [Last accessed 12.09.21].

Lupton D (2016) *The Quantified Self*. Cambridge: Wiley.

Lupton D (2015) Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. *Culture, Health and Sexuality* 17(4):440–453.

Maaroof A (2015) Big data and the 2030 agenda for sustainable development. Available at: https://www.unescap.org/sites/default/files/1_BigData2030Agenda_stock-takingreport_25.01.16.pdf [Last accessed 06.06.21].

MacFeely S (2019) The Big (data) bang: opportunities and challenges for compiling SDG indicators. *Global Policy* 10(1): 121–133.

Machimbarrena JM, Calvete E, Fernández-González L, et al. (2018) Internet risks: an overview of victimization in cyberbullying, cyber dating abuse, sexting, online grooming and problematic internet use. *International Journal of Environmental Research and Public Health* 15(11): 2471–2485.

Martinez A. A snapchat on analytical tools for disaggregating SDG data. International Conference on Sustainable Development Goals Statistics, Manila, 2017.

Mandau MBH (2020) "Snaps", "screenshots", and self-blame: a qualitative study of image-based sexual abuse victimization among adolescent Danish girls. *Journal of Children and Media* 15(3): 431–447.

Mawere M and van Stam G. Data sovereignty: a perspective from zimbabwe. data sovereignty: a perspective from Zimbabwe. In 12th ACM Conference on Web Science Companion, Southampton, 2020. pp. 13–19.

Maya Indira G, Deutch J and Schulte J (2016) Privacy, anonymity, visibility: dilemmas in tech use by marginalised communities. *Brighton: IDS*.

McDermott E, Roen K and Piela A (2015) Explaining self-harm: youth cybertalk and marginalized sexualities and genders. *Youth and Society* 47(6): 873–889.

McGlotten S (2016) Black Data. In: Johnson PE (ed), *No Tea, No Shade: New Writings in Black Queer Studies*. Durham: Duke University Press, pp. 262–286.

McGlynn C, Rackley E and Houghton R (2017) Beyond 'revenge porn': the continuum of image-based sexual abuse. *Feminist Legal Studies* 25(1): 25–46.

Milrod C and Monto M (2020) Prostitution and sex work in an online context. In: Holt T (ed), *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. Cham: Palgrave Macmillan, pp. 1177–1201.

Mitchell KJ, Ybarra ML, Korchmaros JD, et al. (2014) Accessing sexual health information online: use, motivations and consequences for youth with different sexual orientations. *Health Education Research* 29(1): 147–157.

Moreira MW, Rodrigues JJ, Kumar N, et al. (2019) Postpartum depression prediction through pregnancy data analysis for emotion-aware smart systems. *Information Fusion* 47: 23–31.

Nadarzynski T, Puentes V, Pawlak I, et al. (2021) Barriers and facilitators to engagement with artificial intelligence (AI)-based chatbots for sexual and reproductive health advice: a qualitative analysis. *Sexual Health* 18(5): 385–393.

Nadarzynski T, Morrison L, Bayley J, et al. (2017) The role of digital interventions in sexual health. *Sexually Transmitted Infections* 93: 234–235.

Noack-Lundberg K, Liamputtong P, Marjadi B, Ussher J, Perz J, Schmied V, Dune T and Brook E (2020) Sexual violence and safety: The narratives of transwomen in online forums. *Culture, Health & Sexuality* 22(6): 646-659.

Nurmi J (2013) Sexual and reproductive mHealth. *Better Access to Health Care through Mobile Phones*. Geneva: Geneva Foundation for Medical Education and Research.

Obermeyer Z, Powers B, Vogeli C, et al. (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464): 447–453.

Packin NG (2021) Disability Discrimination using Artificial Intelligence Systems and Social Scoring. *IJCL* 8(2): 487–512.

Pai N, Esmail A, Saha Chaudhuri P, et al. (2021) Impact of a personalised, digital, HIV self-testing app-based program on linkages and new infections in the township populations of South Africa. *BMJ Global Health* 6(9): e006032.

Pathak N and Tariq S (2018) Underfunded and fragmented-a storm is brewing for sexual and reproductive health services. *Nature Reviews. Urology* 15(8): 472–473.

Patterson SP, Hilton S, Flowers P, et al. (2019) What are the barriers and challenges faced by adolescents when searching for sexual health information on the internet? Implications for policy and practice from a qualitative study. *Sexually Transmitted Infections* 95(6): 462–467.

Carter P, Gee M, McIlhone H, et al. (2021) Comparing manual and computational approaches to theme identification in online forums: A case study of a sex work special interest community. *Methods in Psychology* 5, p. 100065.

Perera-Gomez T and Lokanathan S (2017) Leveraging big data to support measurement of the sustainable development goals. Available at SSRN: https://ssrn.com/abstract=3058530 [Last accessed 10/06/2021].

Pornhub (2021) The pornhub tech review. Available at: https://www.pornhub.com/insights/tech-review [Last accessed 02/06/2021].

Priyanka P (2019) The devastating impact of Trump's global gag rule. *Lancet (London, England)* 393(10189): 2359.

Qadir J, Ali A, ur Rasool R, et al. (2016) Crisis analytics: big data-driven crisis response. *Journal of International Humanitarian Action* 1(1): 12–21.

Qiao S, Li X, Olatosi B, et al. (2021) Utilizing Big Data analytics and electronic health record data in HIV prevention, treatment, and care research: a literature review. *AIDS Care*: 1–21.

Ruberg B and Ruelos S (2020) Data for queer lives: data for queer lives: how LGBTQ gender and sexuality identities challenge norms of demographics. *Big Data and Society.* 7(1), 205395172093328.

Randall R and McKee A (2017) Becoming BDSM in an online environment. In: Nixon P and Dusterhof K (eds), *Sex in the Digital Age*. London: Routledge, pp. 168–178.

Rao A, Tobin K, Davey-Rothwell M, et al. (2017) Social desirability bias and prevalence of sexual hiv risk behaviors among people who use drugs in baltimore, maryland: implications for identifying individuals prone to underreporting sexual risk behaviors. *AIDS and Behavior* 21(7): 2207–2214.

Rennie S, Buchbinder M, Juengst E, et al. (2020) Scraping the web for public health gains: ethical considerations from a 'big data' research project on HIV and Incarceration. *Public Health Ethics* 13(1): 111–121.

Reynolds C (2015) "I am super straight and i prefer you be too". *Journal of Communication Inquiry* 39(3): 213–231.

Ringrose J, Regehr K and Whitehead S (2021) Teen girls' experiences negotiating the ubiquitous dick pic: sexual double standards and the normalization of image based sexual harassment. *Sex Roles* 85(9): 558–576.

Rowntree O (2019) GSMA connected women: The mobile gender gap report 2019. http://GSMA-The-Mobile-Gender-Gap-Report-2019.pdf [Last accessed 10.06.2021].

Ruckenstein M and Schüll ND (2017) The datafication of health. *Annual Review of Anthropology* 46: 261–278.

Sanders J (2019) Sharing special birth stories. an explorative study of online childbirth narratives. *Women and Birth: Journal of the Australian College of Midwives* 32(6): e560–e566.

Sanders T, Brents B and Wakefield C (2020) *Paying for Sex in a Digital Age : US and UK Perspectives*. London: Routledge.

Sanders T, Scoular J, Campbell R, et al. (2018) *Internet Sex Work. Beyond The Gaze*. Cham: Springer International Publishing.

Sarker M, Wu M, Chanthamith B, et al. (2020) Resilience through big data: natural disaster vulnerability context. international conference on management science and engineering management. *ICMSEM 2020: Proceedings of the Fourteenth International Conference on Management Science and Engineering Management* 1: 105–118.

Schäferhoff M, van Hoog S, Martinez S, et al. (2019) Funding for sexual and reproductive health and rights in low-and middle-income countries: threats, outlook and opportunities. *The Partnership for Maternal, Newborn and Child Health*.

Scott IA (2019) Hope, hype and harms of Big Data. *Internal Medicine Journal* 49(1): 126–129.

Shlomo N and Goldstein H (2015) Editorial: big data in social research. *Journal of the Royal Statistical Society. Section A 178. Part 4*: 787–790.

Silber J (2018) Advances in data integration and Small Area Estimation for equitable and sustainable development. *Discussion of Session EO675 on Advances in Data Integration and SAE for Equitable and Sustainable Development 11th International Conference of the ERCIM WG on Computational and Methodological Statistics(CMStatistics 2018)*. Italy: University of Pisa.-December 14-16.

Simons HR and Kohn JE (2019) Examining temporal trends in documentation of pregnancy intentions in family planning health centers using electronic health records. *Maternal and Child Health Journal* 23(1): 47–53.

Starling MS, Kandel Z, Haile L and Simmons RG. (2018) User profile and preferences in fertility apps for preventing pregnancy: an exploratory pilot study. *Mhealth* 30;4:21.

Stenström K and Pargman TC (2021) Existential vulnerability and transition: Struggling with involuntary childlessness on Instagram. *Nordicom Review* 42(s4): 168–184.

Stevens R, Bonett S, Bannon J, et al. (2020) Association between HIV-related tweets and HIV incidence in the United States: infodemiology study. *Journal of Medical Internet Research* 22(6): e17196.

Shephard N (2016) Big Data and Sexual Surveillance. Available at: https://www.apc.org/en/pubs/big-data-and-sexual-surveillance [Last accessed 24.11.21].

Smith GJ (2018) 'Data doxa: The affective consequences of data practices'. *Big Data and Society* 5 (1): 1–15. DOI: 10.1177/2053951717751551.

Steedman R, Kennedy H and Jones R (2020) Complex ecologies of trust in data practices and data-driven systems. *Information, Communication and Society* 23(6): 817–832.

Taylor L and Broeders D (2015) In the name of development: power, profit and the datafication of the global South. *Geoforum* 64: 229–237.pp.

Tiggemann M and Anderberg I (2020) Social media is not real: the effect of 'Instagram vs reality'images on women's social comparison and body image. *New Media and Society* 22(12): 2183–2199.

Tatsumi T, Sampei M, Saito K, et al. (2020) Age-dependent and seasonal changes in menstrual cycle length and body temperature based on big data. *Obstetrics and Gynecology* 136(4): 666–674.

UNAIDS (2020) HIV financing gap widening https://www.unaids.org/en/resources/presscentre/featurestories/2020/november/20201116_hiv-financing-gap-widening#:~:text=IncreasesinresourcesforHIV,constant2016United%20Statesdollars). [Last accessed 02/03/23]

UN ESCAP (2018) Innovative big data approaches for capturing and analyzing data to monitor and achieve the SDGs. Available online: https://www.unescap.org/sites/default/d8files/knowledge-products/InnovativeBigDataApproachesforCapturingandAnalyzingDatatoMonitorandAchievetheSDGs.pdf [Last accessed 27.09.2021].

United Nations Global Pulse (2013) Big data for development: a primer. Available online at: https://www.unglobalpulse.org/wp-content/uploads/2013/06/Primer-2013_FINAL-FOR-PRINT.pdf

UN Global Pulse (2012) Big data for development: challenges and opportunities. https://www.unglobalpulse.org/wp-content/uploads/2012/05/BigDataforDevelopment-UNGlobalPulseMay2012.pdf [Last accessed 10.08.21].

UN (2016) final list of proposed sustainable development goal indicators https://sustainabledevelopment.un.org/content/documents/11803Official-List-of-Proposed-SDG-Indicators.pdf [Last accessed 06.10.22].

Valdano E, Okano JT, Colizza V, et al. (2021) Using mobile phone data to reveal risk flow networks underlying the HIV epidemic in Namibia. *Nature Communications* 12(1): 2837–2910.

van Heerden A and Young S (2020) Use of social media big data as a novel HIV surveillance tool in South Africa. *PloS One* 15: e0239304.

Weissman S, Yang X, Zhang J, et al. (2021) Using a machine learning approach to explore predictors of healthcare visits as missed opportunities for HIV diagnosis. *AIDS.* 35: s7–s18.

Wilson E, Free C, Morris TP, et al. (2017) Internet-accessed sexually transmitted infection (e-STI) testing and results service: a randomised, single-blind, controlled trial. *PLoS Medicine* 14(12): e1002479.

Wu B, Tian F, Zhang M, et al. (2020) Cloud services with big data provide a solution for monitoring and tracking sustainable development goals. *Geography and Sustainability* 1(1): 25–32.

Wyber R, Vaillancourt S, Perry W, et al. (2015) Big data in global health: improving health in low- and middle-income countries. *Bulletin of the World Health Organization* 93: 203–208.

Xue J, Chen J and Gelles R (2019) Using data mining techniques to examine domestic violence topics on twitter. *Violence and Gender* 6(2): 105–114.

Yang X, Zhang J, Chen S, et al. (2021) Utilizing electronic health record data to understand comorbidity burden among people living with HIV: a machine learning approach. *AIDS*, 35: S39–S51.

Young JC, Lynch R, Boakye-Achampong S, et al. (2021) Volunteer geographic information in the global south: barriers to local implementation of mapping projects across Africa. *GeoJournal* 86(5): 2227–2243.

Young SD (2015) A "big data" approach to HIV epidemiology and prevention. *Preventive Medicine* 70: 17–18.

Young SD and Zhang Q (2018) Using search engine big data for predicting new HIV diagnoses. *PloS One* 13: e0199527.

Young JC, Lynch R, Boakye-Achampong S, et al. (2021) Volunteer geographic information in the Global South: barriers to local implementation of mapping projects across Africa. *GeoJournal* 86(5): 2227–2243.

Zhang Y, Huang T and Bompard EF (2018) Big data analytics in smart grids: a review. *Energy Informatics* 1(1): 8–24.

Zhao P, Yoo I, Lavoie J, et al. (2017) 'Web-Based Medical Appointment Systems: A Systematic Review'.' *Journal of Medical Internet Research* 19 (4):e134.

Natalie Hammond, Manchester Metropolitan University. Natalie is a sociologist and senior lecturer in the Department of Social Care and Social Work. Her research focuses on gender, sexual and reproductive health and well-being and digital technologies.

Angelo Moretti is an Assistant Professor in Statistics at Utrecht University in the Department of Methodology and Statistics. Angelo is a survey statistician and an elected member of the International Statistical Institute (ISI). He has conducted research in small area estimation under a wide range of approaches, such as multivariate generalised mixed-models and survey calibration. His research also focuses on mean squared error estimation based on bootstrap approaches, and data integration methods (statistical matching and probabilistic record linkage). He is also interested in applications related to social exclusion, crime and public attitudes indicators.