



Datafication, Power and Control in Development: A Historical Perspective on the Perils and Longevity of Data

Katarzyna Cieslik^{ID} and Dániel Margócsy

Abstract: The collection, processing, storage and circulation of data are fundamental element of contemporary societies. While the positivistic literature on ‘data revolution’ finds it essential for improving development delivery, critical data studies stress the threats of datafication. In this article, we demonstrate that datafication has been happening continuously through history, driven by political and economic pressures. We use historical examples to show how resource and personal data were extracted, accumulated and commodified by colonial empires, national governments and trade organizations, and argue that similar extractive processes are a present-day threat in the Global South. We argue that the decoupling of earlier and current datafication processes obscures the underlying, complex power dynamics of datafication. Our historical perspective shows how, once aggregated, data may become imperishable and can be appropriated for problematic purposes in the long run by both public and private entities. Using historical case studies, we challenge the current regulatory approaches that view data as a commodity and frame it instead as a mobile, non-perishable, yet ideally inalienable right of people.

Key words: Datafication, big data, data revolution, digital colonialism, Data4Dev

I. Datafication: Data Opportunity or Data Problem?

The collection, processing, storage and circulation of data are central elements of a large number of sectors of contemporary societies. This process of ‘datafication’ (Cukier and Mayer-Schoenberger, 2013) has also become central to international development (Cinnamon, 2020; Etzo and Collender, 2010; Mann, 2018). Specifically, the ‘ICT revolution’, currently on the way in the Global South, is believed to have created unparalleled opportunities for big data to improve efficiency, transparency and the

accountability of development projects, ranging from agricultural and service provisioning to humanitarian interventions (IEAG, 2014). While a number of scholars praise big data especially in relation to economic planning and policy-making (Kitchin, 2014; Kleine and Unwin, 2009), others point to its many limitations as well as the threats it poses to personal freedom and democracy (Arora, 2016; Crawford and Schultz, 2014; O’Neil, 2013; Sadowski, 2019). Research has already demonstrated that, in their rush to adopt new technologies, development and humanitarian agencies deploy solutions that enable data

surveillance and that have ended up supporting systems that pose serious threats to individuals' human rights (de Corbion et al., 2018; Hosein and Nyst, 2013). Such threats are particularly pronounced in developing country contexts as they 'hit harder where people, laws, and human rights are the most fragile' (Milan and Trere, 2019).

The common denominator for both the proponents and the critics of 'datafication' is a fair degree of technological determinism: the underlying assumption that the phenomenon is both recent and original, fuelled by the contemporary advances in information and communication technologies (Kleine and Unwin, 2009; Taylor and Broeders, 2015). Within this discourse, ICT-enabled universal connectivity is thought to be the primary driver of datafication (Mejias and Couldry, 2019). In this article, we propose an alternative approach. We use historical evidence from the early modern period to show that 'datafication', as defined by most authors, is not a novel process. We argue that datafication is not a one-time result of a technological invention. Throughout history, datafication happened again and again because of political and economic pressures, and it relied on a variety of paper, electronic and digital technologies (Asif, 2019). Moreover, since datafication processes occur in the context of asymmetrical power relations, they can result in an extraction of value that some scholars call 'data colonialism' (Arora, 2018; Mejias and Couldry, 2019; Thatcher et al., 2016). Our article focuses on how datafication can contribute to processes that can lead to the unequal distribution of wealth and resources and can also facilitate the surveillance of individuals or populations.

Although critical data studies researchers have established that many individuals and groups across the world are increasingly reluctant to share their data (Bronson and Knezevic, 2016; Davidson, 2018; Wiseman et al., 2019), the data-for-development discourse still prevails in the context of the Global South. The discipline of development

studies excels in a long-standing critique of the quantification and measurement imperative within the sector (Chambers, 1997), but the controversies surrounding the cyclical nature of data production, that is, the complex processes through which data are collected, processed, stored, owned, used and reused, have only recently captured the researchers' attention (Iazzolino and Mann, 2019; Mann, 2018). As a result, the argument may have shifted from an assumed neutrality of big data towards a strong belief that it is possible to inscribe it with the 'right' kind of values (Hilbert, 2016). In this synthetic review, we rely on studies from the history of data and information to examine how the complex and long life cycle of data poses tangible risks to those who provide data, with a special focus on colonial settings that are particularly relevant for development studies. While we acknowledge that there are significant differences between the abuses of information 500 years ago and the dangers of algorithmic datafication in the twenty-first century, there is also an unacknowledged continuity that scholars of development studies should not ignore (Mann and Hilbert, 2020). We argue that, even in cases where data production was not driven solely by governments or multi-national corporations, there was a real and recurrent danger of data being appropriated and processed for the purposes of economic and political control.

Our contribution is as follows. First, we qualify the universally accepted view that datafication is a recent phenomenon, driven by the contemporary advances in connectivity, arguing that, in fact, twenty-first-century datafication shares many features of what had already happened through history. Second, by framing current and past cases of datafication as a deliberate political project of development, we call into question the conventional, top-down understanding of governmental surveillance that has proliferated since the seminal writings of Scott (1998, see also Merry et al., 2015).

We examine how colonial and post-colonial subjects, often under conditions of constraint, have also contributed to the co-construction of practices of data collection and data processing (Brendecke, 2016; D'Onofrio, 2016). Our cases demonstrate that even bottom-up processes can be harnessed for the purposes of extractive political and economic control. Third, we argue that the current scholarship has not acknowledged sufficiently how data have a long life cycle, how they can change hands repeatedly and how they can be repurposed across decades or centuries (but see Leonelli and Tempini, 2020). Fourth, on the basis of these arguments, we call into question the current regulatory proposals that advocate returning ownership of data to individuals or making data publicly available in open-access frameworks (Sieber and Johnson, 2015). We show that such solutions fail to take into account the argument that data are often co-constructed and that they have a life cycle. We argue that all legislative systems that consider data as intellectual property that can be traded and sold, including those that propose communal ownership or commons-type solutions to data, are prone to exploitation by those in power in the long term. Such solutions are also always dependent on ever-changing, international and local political systems (Mann, 2018). We propose to extend the framework of the data justice movement by considering data the inalienable right of people, which can be shared only in a temporally limited and reversible process (Cinnamon, 2020; Heeks and Renken, 2018; Qureshi, 2020; Taylor, 2017).

The article is structured as follows: the next section provides an overview of the existing definitions of datafication. The following section builds this to argue that researchers should balance a largely optimistic discourse of 'data revolution' with discussions that focus instead on the dangers of information extraction and surveillance inherent in the increased use of data, a process some have called data colonialism (Coleman, 2018; Kwet, 2019; Thatcher et al., 2016). We then turn to the historical analysis of early modern cases

of colonial data processing, relying on the rich outpouring of literature on the history of information, archives and knowledge (Daston, 2017). We find that, despite the apparent differences, the current and past cases of datafication share commonalities and, over time, lend themselves to the same threats of appropriation and control. We then discuss current proposals to reform existing regulatory approaches.

II. Defining Datafication

Often used interchangeably, the terms 'big data' and 'datafication' have become highly popular in recent years. The current definitions of these terms often explain datafication as the increased ability to quickly process large amounts of information, often with the help of deep learning algorithms (Ylijoki and Porras, 2016). Analysts have argued that the sheer size of big data allows one to make accurate prognoses without scientific modelling (Anderson, 2008; Cukier and Mayer-Schoenberger, 2013). As sample size grows, deep learning putatively allows computers to recognize patterns without hypotheses or models, in a manner that is not transparent to users and programmers.

Yet, even when acknowledging these radical advances in programming cultures, many elements of the current datafication process pre-date the current information technology revolution. Cukier and Mayer-Schoenberger's influential work explained that 'to datafy a phenomenon is to put it in quantified form so that it can be tabulated and analyzed' (Cukier and Mayer-Schoenberger, 2013: 78), that is, to reduce information into elements that can be processed with 'computer memory, powerful processors, smart algorithms, clever software, and math' (Cukier and Mayer-Schoenberger, 2013: 29). Other scholars similarly define datafication simply as the process of 'dematerialization' that converts natural phenomena into symbolic material that can be indexed and searched (Lycett, 2013; Mejias and Couldry, 2019).

In the context of economic organization, Fourcade and Healy (2017) also claimed that datafication is a phenomenon that enables mass commodification through abstraction and mathematical processing.

Apart from an emphasis on computers, all these definitions refer to processes that are well documented to have been practiced for several hundred years. For example, Bruno Latour long explained the rise of Western science since 1500 in terms that strongly resemble the above definitions of 'datafication'. Latour's concept of 'inscription' is strikingly similar to today's 'data': the production of inscriptions is the process of turning social and natural phenomena into mathematical formulas and images that are 'mobile, flat, reproducible' and can be 'reshuffled and recombined' to produce new scientific knowledge (Latour, 1986: 21). Like the process of datafication, Latour's 'inscriptions' rely on the manipulation of abstracted phenomena with the help of mathematical and geometrical processes such as algorithms. In agreement with Latour, other scholars have also emphasized that the symbolic and numerical representation and manipulation of people and knowledge have been standard elements of the process of building colonial empires (Appadurai, 1993). Following this literature, we define the common element between twenty-first-century datafication and earlier efforts of information extraction as follows: the reduction of phenomena into abstract entities that can be exchanged and shared with others easily and the organization of these entities into a database that can be processed through mathematical and other types of analysis. We emphasize one additional, common feature that is often forgotten: data, once collected and stored, can be used and reused for new purposes across surprisingly long periods of time. For this reason, we stress the processual and cyclical nature of datafication, emphasizing that it extends beyond collection and storage and that it includes continuous processing and reprocessing by those in power, be they national

governments, transnational corporations or academics.

We employ a definition of data that is broad enough to include phenomena spanning centuries because an emphasis on the novelty of twenty-first-century datafication processes obscures the long-term effects of datafication that become visible only if viewed from a historical perspective. As the next section shows, the mainstream scholarly and policy discourse on development tends to emphasize the here-and-now gains of data processes, such as the increased transparency, accountability and efficiency of development interventions. It also tends to downplay or ignore concerns about what happens to data in the long run and how it can be used and reused for a variety of purposes after the original development intervention has been accomplished.

III. Data for Development Discourse

For the context of studying development, the shared commonalities between the twenty-first-century datafication and the earlier processes of information extraction serve as a warning call about narratives that present data for development (Data4Dev) as an unquestionable good. This highly problematic discourse is fairly widespread in the current world of scholarship and policy-making. As expressed by the UN High Level Panel of Eminent Persons:

We need a data revolution. Too often, development efforts have been hampered by a lack of the most basic data about the social and economic circumstances in which people live. Substantial improvements in national and subnational statistical systems including local and subnational levels and the availability, quality and timeliness of baseline data, disaggregated by sex, age, region and other variables, will be needed. (United Nations, 2013: 3)

Though never explicitly defined, the 'data revolution' is supposed to entail a technological push, 'open data' schemes, capacity building in national statistics agencies, and more

large-scale surveys (Demombynes and Sandefur, 2014). The Open Data Partnership (a multilateral initiative to promote open data to strengthen governance) is one example, with almost 80 countries pledging commitment to open government data (OGD) and global data flows across borders.¹ The open Data4Dev project is a similar project, supported generously by Canada's International Development Research Centre (IDRC), the World Bank and United Kingdom's Department for International Development (DFID).² The degree of 'openness', per country or per region, is now also being measured by using Open Data Index and Open Data Barometer.

Within the Global South, there is an especially strong emphasis on data revolution in the context of Africa. In 2018, the Africa Data Revolution Report urged governments to set up and institutionalize OGD to promote economic growth and foster innovation. Similarly, the Mo Ibrahim Foundation (MIF) Report stresses the 'data gap' in Africa, pointing to issues of data structure and quality, coordination and insufficient collection frequency as core factors hindering progress towards the African Union's Agenda 2063 (MIF, 2019). Within this rhetoric, it is the lack or unreliability of development statistics that are at fault for the relative failure of the development project (Jerven, 2013). Against this background, big data has been praised for enabling more efficient agriculture (Carletto et al., 2015; Kamilaris et al., 2017), tailored healthcare systems (Amankwah-Amoah, 2016) and real-time environmental monitoring (Cieslik et al., 2018).

Fuelled by the same discourse, in India, Aadhaar—the world's largest biometric data system, created in 2009 to facilitate the administration of welfare benefits—links 'databases of bank accounts, mobile phones, income tax returns, payment apps, email IDs and so on, even if such a linking is not mandated by the law' (Mertia, 2020: 11). Administered by the Unique Identification Authority of India, Aadhaar has been praised as a 'robust

and inclusive identification system', a 'pillar of sustainable development, particularly when leveraged by new technologies that greatly increase their accessibility, precision and usefulness' (Gelb and Metz, 2018: 3). A recent edited volume by Khera (2019) presents a selection of essays telling an alternative story: the one that Aadhaar was never really about welfare, but about citizen profiling, government surveillance and commercial data mining. Despite these concerns, a vocal 2015–2018 court case verdict of the Supreme Court of India upheld the use of Aadhaar (Khera, 2019), and registration in the system is now required to access a number of public goods.³

The strong belief that data are knowledge and that knowledge is progress unites the public and private sectors. Though often meant to facilitate development response, government-donated and anonymized open data can be easily repurposed for business or political clients (Burns, 2015). As Mann (2018) have shown, an increasing number of projects now involve extracting data from various organizations in the Global South for expert analysis in Western countries and, under the guise of humanitarian and development assistance, these data become the source of revenue, knowledge and power for Western companies. Echoing these concerns, Taylor and Broeders (2015) showed how datafication has reinvented public–private partnerships in low- and middle-income countries, ushering in an era of data corporation hegemony in the development sector (see also Kwet, 2017). Exploitation of data not only for profit but also for political uses (e.g., predictive analytics) is what Coleman (2018) calls a new 'scramble for Africa' currently underway on the continent.

The relative regulatory void lends special urgency to researching data processes in the Global South as data exploitation, profiling and surveillance frequently occur in contexts of insufficient or ineffective legal frameworks (Privacy International, 2020). In 2010, the Economic Community of West African States (ECOWAS) adopted a Supplementary

Act on Personal Data Protection and, in 2013, the Southern African Development Community (SADC) published a Model Data Protection Act, but neither of these are legally binding (Makulilo, 2016), making Africa a 'testing ground for technologies produced elsewhere' (Privacy International, 2020). The African Union Commission (AUC) adopted the Malabo Convention on Cyber Security and Personal Data Protection already in 2014 with the objective to create a 'safe digital environment for citizens' but to date, only 24 African countries have adopted national personal data protection policies. In May 2018, the AUC and the Internet Society launched the Personal Data Protection Guidelines for Africa, advocating for special consideration for personal privacy, trust, safety and responsible use, but these are yet to be debated by the national governments. According to Sutherland,

with rare exceptions, governments and parliaments have failed to propose, scrutinize and enact the legislation and institutional arrangements essential for cybersecurity and data protection, leaving data and systems exposed to commercial misuse and to significant risks from criminals, foreign powers, hacktivists and terrorists. (Sutherland, 2018: 10)

In Asia, the development of data protection regulation can at best be described as uneven. In India, the Personal Data Protection Bill is currently under review by a Joint Parliamentary Committee, but progress has been slow due to the COVID-19 pandemic. In addition, the bill has already sparked controversy as it gives the government permission to access business intelligence and the intellectual property of companies for largely unspecific 'development' purposes (Mathur, 2020). The situation is similar in Latin America, where, with the exception of Mexico and Colombia, there are no legal restrictions of data use by the public sector (Rodriguez and Alimonti, 2020). In Africa, some governments have managed to introduce certain elements of

regulation (South Africa, Kenya, Nigeria, Togo and Uganda) (Greenleaf and Cottier, 2020; Kshetri, 2019). In most cases, however, the public-sector stakeholders are exempt from data regulations (also in Africa) and they may claim unlimited access to any data as long as they act 'in the public interest' (see e.g. the case of Nigeria, Olowogboyega, 2020). Last but not least, across the Global South, data law enforcement remains a challenge: that is, according to Abdulrauf (2020), many African institutions simply 'lack the teeth' to make data companies comply with the newly introduced provisions.

While data policy may not seem to be a priority in comparison with other, more dire, development challenges in the Global South, the fact that digital registration is now required in many countries to access basic services (e.g., healthcare) as well as to execute electoral rights (biometric voting) lends new urgency to both research and regulation (Breckenridge, 2014). We believe that case studies from the history of colonization provide further warnings about the cyclical nature and longevity of data that the current scholarship in critical data studies tends to downplay (e.g., Biruk, 2018). We argue that, in order to understand the risks of datafication, we need to study how data have been used and reused across long time periods and how the ownership and uses of the data have become dissociated from the initial context of production (Leonelli and Tempini, 2020). We need to understand the complex temporal and spatial organization of data landscapes because 'to understand "big data" and whatever comes next, we must resist this urge to let it stand apart from history and pass silently into our everyday lives' (Dalton and Thatcher, 2014). As the following sections reveal, iconic case studies from history provide important lessons on how the data individuals share about themselves may be used against them without their knowledge. We rely on historical case studies partly because critical scholarship on twenty-first-century datafication in development has

only recently begun to appear, and partly because history is the discipline best placed to study developments across a long time span. Nonetheless, our argument is firmly about how to conceptualize development and Data4Dev in the twenty-first century and how to regulate the uses of data in the context of the Global South. For this reason, in the sections that follow, we use examples from history to evaluate three existing data privacy regulation proposals, coming from critical data studies: data as a commodity, data as a public good and data as an inalienable right.

IV. Data Processes Through History

The Data4Dev discourse rarely discusses extensively that the use of information collection for the purposes of governing development has a long history. It has even been argued that centralized states could emerge in Antiquity because of the emergence of writing enabled the storage of data related to taxation (Goody and Watt, 1963). Throughout the millennia, a variety of data collection and storage techniques were invented for the purposes of political administration, including cartographic surveys, censuses, libraries, museums or index card systems (Krajewski, 2011; Schiebinger and Swan, 2005), in a variety of states across the globe (Dennis, 2015; Guha, 2003; Habib, 2013; Peabody, 2001). In the wake of the printing revolution, sixteenth-century European writers already considered the rapid accumulation of data a major new problem and devised paper technologies to manage this information overload (Blair, 2010). The Spanish colonization of Latin America was a major step in this development of complex information management systems and still has scientific and political relevance for today's concerns with datafication (Brendecke, 2016; Portuondo, 2009). Colonial processes of information management are essential for understanding the problematic nature of datafication because they instantiate how the long-distance government of subjugated and subaltern populations is reliant upon

the development of efficient systems of data production. Throughout the centuries, colonization relied on the violent and inhumane appropriation of land and resources by Western powers, and then on the continuous extraction of resources and wealth from these colonies across a prolonged time period through direct or indirect rule. Our article focuses primarily on the extraction and processing of information for the long-term management of colonial empires, which relied on the co-production of data by explicitly or implicitly coerced indigenous informants. Though co-produced, such data were nonetheless often exploited against local populations as they were re-purposed for novel uses during its long shelf-life. A look at these colonial processes of information management reveals some curious and concerning parallels with current data-gathering processes in the Global South. We offer three case studies. We first study the colonial government of Latin America, arguably the first major attempt at colonization by European powers. Second, we turn to the early Royal Society's proposals for relying on data for colonial and population management, which highlights that innovative mathematical and statistical solutions have been used for data processing for a long time. Third, we analyse the production of data in Enlightenment colonial governments to offer a contrasting narrative to James Scott's influential *Seeing like a State*, which argued that governments of this time period produced data in a top-down manner. In Table 1, we explain how each of these three cases represents a case of datafication (as per our definition, see Section II) and we provide current-day examples of parallel processes.

Information Processing in Colonial Latin America

An emphasis on empirical knowledge and extensive data collection lay at the centre of the Spanish Empire's colonial project in Latin America from the beginning, which gathered navigational charts, cartographic data and botanical illustrations to establish political

power and spur economic development (Bleichmar, 2012; Canizares-Esguerra, 2006; Harley, 1989; Kirsch, 2014). Spanish colonial data collection included information on colonial subjects, the mapping of the natural resources of the continent, as well as systematic mathematical data on latitude, longitude

and tidal heights based on observation with instruments, mostly kept unpublished and under lock at the imperial archives for future use by the state (Portuondo, 2009). Importantly, even coercive societies, such as the sixteenth-century Spanish colonial empire, relied on the (frequently forced) collaboration of many

Table 1. Historical and Current Datafication

Datafication	<i>Historical Examples</i>			<i>Current Examples</i>
	Spanish colonial conquest 1500–1800	Political arithmetic in Britain	Colonial empires c. 1800	Current datafication
Types of data	Paper, maps and navigational data	Paper, maps and tabulated data	Maps, pre-colonial archives, surveys and census	Digital content data (social media, forums and websites), sensing data (from remote sensors and satellites), data exhaust (passively generated) and public data (surveys and databases maps or census data)
Data production	Scientists, sailors, priests and neighbours	Irish surveyors, ‘searcher’ women and local informants	Local populations	All users of digital devices, in particular users of specialized apps and the social media (e.g., banking apps and farming apps) and institutions
Data harvesting capture	Spanish colonial government, church authorities and scientists	The English state and commercial enterprises	Colonial governments and institutions; scientists	International organizations, governments, private companies and research institutions
Data aggregation	Centralized archives	Archives, maps and mathematical tables	Colonial administrators and local elites	Data companies, governments and research institutes through cloud and physical repositories
Data processing	Text-based	Political arithmetic and mathematics	Map-based and mathematical	Digital and algorithmic
New forms of exploitation	Inquisition, colonial rule, And colonial reform	Ireland’s colonization, policy towards the poor and maritime colonial empire	Colonial state structures	Marketization of data, sale and re-sale

Source: The authors.

people in the colonies. Jorge Canizares-Esguerra has argued that sixteenth- and seventeenth-centuries Spanish administrators and historians were heavily reliant and accepting of native and creole informants, writers and historians to learn more about the history of pre-Columbian Latin America and the details of the Columbian conquest (Canizares-Esguerra, 2002). Standardized historical, mathematical and cartographic information was, therefore, produced on location with a large number of observers and experts involved, which was then shipped back to the Iberian Peninsula where the state bureaucracy sorted, analysed, and stored this knowledge and used it to advise the governing body of the Council of Indies on policy matters. Scientific knowledge was collected together with information on the inhabitants of Latin America in the hopes of implementing better and stronger political and economic control over the colonies. Political and economic control did not necessarily result in improvement at the local level, as shown by recent studies of the cinchona, a highly popular medicinal plant to fight malaria. As Crawford (2016) has documented, Spanish colonial authorities appropriated knowledge about the medicinal qualities of this plant from indigenous populations and relied on local expertise to select those variants that were most effective in fighting disease. By the late eighteenth century, cinchona was one of the most important exports from Peru and New Granada, leading to the establishment of large-scale plantations concentrated in the hands of a few merchants. It also resulted in the dispossession and impoverishment of much of the local population (Gänger, 2020).

Understanding the contributions of local populations to the Spanish Empire's centralizing efforts is key towards understanding the problematic concept of consent that twenty-first-century datafication projects rely upon to justify their acquisition and use of data from local informants. Colonial history serves as a useful reminder of how the fiction of rational

actors agreeing to legal transactions of data while keeping their long-term interests in mind is not applicable in situations where the actors are operating in conditions of poverty or where they live in a coercive society (Fanon, 2008). As Bredecke (2016) has shown in an incisive analysis of how colonial power operates, the Spanish colonial government was heavily influenced by the practices of inquisition, an institution based on the extensive collection of data based on coercion and surveillance. The Spanish inquisition and the empire's inquisitorial practices worked because local people volunteered information about themselves and, equally importantly, about each other, in exchange for a variety of perceived favours. For example, inquisitorial trials against suspected Muslims in Latin America were usually based on reports by neighbours and other witnesses, or, as in the case of a certain Maria Ruiz, by the victim herself (Qamber, 2006). A former Muslim, Ruiz turned herself in to the Inquisition because of her fear of Christian hell, providing detailed information on her earlier life to her inquisitors. In coercive societies, there were good and rational reasons to perform transactions with private data that one would not have released in more ideal circumstances. To illustrate how scientific data were exchanged under such circumstances, Schiebinger (2004) has recounted how the eighteenth-century French traveling naturalist Nicolas Joseph Thierry de Menonville worked as a spy to steal economically useful plants, such as the cochenille, from Spanish Mexico to diversify French agricultural output. While Thierry de Menonville paid local cultivators in Mexico for sharing their economically useful plants, he also explained that, if his local providers had all refused to trade with him, he would have stolen the cochenille in an act of war, anyway. Transactions performed in the shadow of piracy and war were consensual contracts only in a highly limited sense of the law.

The data contained in the archives of the Spanish Empire enabled the long-term

rule of the colonies across several centuries, allowing administrators to rely on extensive data sets to make decisions about governing and the fate of individuals. The idea of the archive was that information did not decay and could be preserved across decades and centuries. Unlike the project-based scientists of today, who do not often consider how the data they have collected will be deployed after their project ends, colonial administrators were very much aware of the longevity of information. When Spanish botanists began to catalogue Latin American nature in the late eighteenth century, part of their project involved the study of manuscripts produced 200 years earlier because they believed that old colonial knowledge could be used for the new purposes of Enlightenment agricultural reform (Bleichmar, 2015). It is no accident that it was during this age of Enlightenment reform that the *Archivo General de Indias* was established, a comprehensive archive that brought together documents and data stored across the Empire, in order to facilitate the processing and management of knowledge about the Americas (Slade, 2011).

Political Arithmetic and the Early Royal Society in England

The twenty-first-century datafication is often touted for its reliance on innovative mathematics and algorithms, so it is important to emphasize that novel mathematical and data management techniques were already characteristic of early modern information processing. The spectacular rise of mathematics and statistics in the seventeenth century was intricately connected to colonial projects of data management. The early Royal Society is an iconic example of the rise of mathematics, and its fellows were enthusiastic about using numbers to manage the empire. These fellows aimed at ordering society with the help of numbers and tables just as they hoped to use the same tools to manipulate nature (Buck, 1978, Shapin and Schaffer, 1985), and they played a key role in colonial projects in Ireland, the

Caribbean and slave ports across the globe. The history of statistics in this period offers more detailed evidence on how the contributions of local informants resulted in data that could be used and reused across decades and centuries.

Like the Spanish Empire, the Royal Society was also reliant on acquiring information from local sources through duplicitous means. In his early years, the luminary Isaac Newton (1669: f.4r), future president of the Royal Society, wrote explicitly about the necessity of dissimulation for traveling observers among foreigners, arguing that they should let their

discours bee more in Quaerys & doubtings than peremptory assertions or disputings, it being the designe of Travellers to learne not teach; besides it will persuade your acquaintance that you have the greate esteem of them & soe make them more ready to communicate what they know to you.

As Newton explained, one could collect information through such acts of dissimulation, by ‘seeming to approve & commend what they like’, about the ‘wealth and state affaires of nations’, the fortifications of foreign countries, or the cost of living there. Once treated with respect, Newton claimed, people would willingly part with confidential information that could be used against themselves. Newton’s *Principia mathematica* relied on precisely such sources to gather tidal data from colonial slaving and trading ports, such as Tonkin (Hanoi), for his mathematical analysis of tidal patterns, which held the promise of improving the navigational abilities and imperial designs of the Navy (Schaffer, 2009). Importantly, Newton’s analysis did not signal the end point for using his sets of tidal data, which came to be reused again and again. After the publication of the *Principia*, Newton’s associate Edmond Halley undertook extensive maritime travel to collect further data points and used these together with the earlier data to improve upon Newton’s calculations and to publish updated charts for the use of English mariners (Reidy, 2008).

In the same period, data also played a crucial role in the emerging disciplines

of surveying, demography and political economy. The first detailed land survey was the Down Survey of Ireland by William Petty, conducted in the wake of the island's colonization by Oliver Cromwell in the early 1650s. Like all surveys, the Down survey and the map it resulted in were based on Petty's familiarity with the complex instrumentations of measurement and required extensive mathematical and geometrical expertise. They allowed the English state to embark upon the massive restructuring of land ownership and the dispossession of the Irish in the process. Petty also played a major role in applying statistical methods to the study of population by contributing to John Graunt's *Natural and Political Observations ... on the Bills of Mortality* of 1662, the pioneering work of demographics. The *Natural and Political Observations* relied on extensive amounts of population data (McCormick, 2009:134). Although the *Natural and Political Observations* provided important insights into the health of Londoners, its primary aim was to use these data to make explicit policy recommendations, such as the provision of regular government income to beggars, to control the behaviour of the poor. The information management of colonies and reforms in population management went hand in hand.

A major innovation of the *Natural and Political Observations* was to repurpose old data with the help of new mathematical tools in order to develop government policy. The numbers it relied on were not novel. They came from the weekly bills of mortality, printed regularly from 1602 onwards, which provided statistical information on the number of dead, parish by parish. These bills of mortality have been considered one of the earliest newspapers, and in that role, they have been lauded for generating public discourse and local health organization in moments of crisis (Heitman, 2020). Yet, they relied on the collaboration of poor and vulnerable women who collected the data by forced agreement. Impoverished widows supported by the alms of the parish, called

'searchers', were responsible for determining and reporting the causes of death by visiting the houses of the dead and inspecting them up close, a task not without risks during the time of the plague. Searchers risked not only death but also social isolation for their contact with the sick and were also accused of witchcraft on occasion. While these impoverished widows were remunerated for their reports and had the theoretical possibility to refuse becoming a searcher, they stood to lose their alms if they actually did so, ensuring that they would not do so in practice (Munkhoff, 1999).

Even in the original context, the numbers provided by searchers were not innocent. As the *Natural and Political Observations* noted, the weekly bills of mortality were first produced 'so the Rich might judg' of the necessity of their removal, and Trades-men might conjecture what doings they were like to have in their respective dealings' (Graunt, 1676: Preface). With the urban poor providing the data, the wealthier strata of society could decide more efficiently whether to remove to rural sites of safety. Yet, the bills of mortality could become potentially efficient tools of government only when they were reprocessed with the help of mathematics by Graunt and Petty, which offered the promise of developing complex policies for population control, decades after their were produced. And the repurposing did not stop with the publication of the *Natural and Political Observations*. When better mortality records became available from Wrocław in the early 1680s, Halley immediately set out to use these data to calculate the appropriate rate of return for life annuities, a major source of income for the state, even if it again took several decades before the actual calculations of annuities began to incorporate Halley's results (Deringer, 2018, Slack, 2004).

Colonial Governance in the Eighteenth and Ninetieth Centuries

Observation and data continued to be crucial for the development of the British, Dutch and other colonial empires in the eighteenth and

nineteenth centuries. As Scott (1998) has shown, this was the period when scientific forestry emerged in the hopes of developing agricultural practices based on the predictive calculation of long-term timber yield not only in Europe but also in colonies such as the Dutch East Indies (Knaap, 1987). Scott has argued that such projects were at once dangerous and prone to failure because they imposed an abstracted, imperial vision on nature and people from above. Yet, imperial projects of data collection and processing were not always performed in strictly top-down processes, and they could be productively harnessed for the purposes of colonial control. This was especially the case in eighteenth-century India where the British government established indirect rule that relied on the co-option of local political and legal power structures. In such political circumstances, data were often constructed with the help of local populations. As Raj has shown, for instance, British cartography was born in India where European cartographers, such as James Rennell, cooperated with Indian experts and traditions in the process of surveying the subcontinent, which was mapped in more detail than the British Isles (Raj, 2017). Reliance on local informants was also essential for many colonial administrators in need of long-term, historical data to understand the political and economic situation of India. As Dirks has shown, effective government required mining the historical chronologies, genealogical records and financial registers of previous local Indian rulers, making it essential for colonial officials to ensure the cooperation of existing local elites (Dirks, 1993; Wagoner, 2003). Data originally collected for effective local rule were taken over decades or centuries later by the British Empire for effective colonial rule.

Importantly, the appropriation of local data could go hand in hand with the appropriation of local practices of data processing, as well. In the making of his famous maps, which correlated temperature, precipitation and other variables across the globe with the help

of innovative visualization techniques, the German polymath Alexander von Humboldt relied to a large extent on the previous efforts of Latin American colonial scientists, which he appropriated without acknowledgment during his lengthy stays in cities such as Quito and Lima (Canizares-Esguerra, 2006). Creole scholars openly shared data and the techniques of processing and visualizing these data with Humboldt, which the German scholar presented, years later, as the result of his own explorations and discoveries, enhancing his credibility in European metropolis at the cost of creole elites.

The Lessons of History

By highlighting how the infrastructure of datafication was built during the colonial period, our cases have made explicit how datafication relies on complex social networks and political ecologies (Bouk, 2017; Dencik, 2020). Datafication is not simply a technological novelty: it is also the result of particular and hierarchical social interactions. Our analysis of the beneficiaries of and contributors to the datafication processes has stressed the negotiated and participatory nature of data, in line with the recent literature on quantification and statistics (D'Onofrio, 2016), challenging top-down interpretations of quantification. As we have shown, the Latin American colonial empire worked with data provided by local informants, the survey of Ireland proceeded with the help of locally recruited surveyors, the mortality data in London were collected by local searchers, and the mapping and government of India relied on the extensive expertise and archives of local states. In all these cases, the collaboration of local experts and the co-option of previous techniques of administration were heavily shaped by inequalities in political power, yet they reveal the inadequacy of the simple, binary oppositions between the oppressors and the oppressed that Scott has posited. Applying Brendecke's more general insights on colonial power to the issue of data management, we have shown that colonial administration

could sometimes rely on the opportunistic cooperation of local, indigenous elites or on the forced collaboration of the enslaved and the poor, but this did not make it any more innocuous in the process. Consent to data transfer manufactured in highly hierarchical situations resulted in unequal benefits for providers and consumers of data.

Second, these case studies have highlighted the long life cycle of data and how it can be deployed for originally unforeseen purposes years, decades or centuries after their creation. In Latin America, botanical information collected in the sixteenth century was used and reused for the next 200 years, at least. In seventeenth-century England, population mortality data were analysed with the help of new statistical methods more than 50 years after it was first deployed. In colonial India, local archival, financial and historical data produced by local rulers were deployed for new purposes after colonization by Britain. These case studies highlighted a further problematic aspect of data produced or co-produced by local informants and then acquired in politically charged situations or under coercive pressure by colonial powers. Since data had (and still have) a long shelf-life and they could be deployed for novel purposes over the long term, local informants participated in the initial

data transfer without full cognizance of the potential future value of data. The strength of colonial empires lay, in part, in their ability to amass large amounts of data that they could exploit and process for novel purposes across a long time period.

Our double emphasis on the co-construction and longevity of data offers a new model of the life cycle of datafication in the long term. Figure 1 offers a visual summary of the model.

The first stage of data generation can involve top-down or bottom-up processes and may involve voluntary and involuntary participants, as in the case of creole elites or Irish colonial subjects. Data aggregation can be carried out by individuals, such as Humboldt, multinational companies or the state. Aggregation is crucial for making data useful for applications in a variety of contexts, and the ownership of the distributed and aggregated data may not remain the same as that of individual's data. Locals often played an essential role in collecting data, while aggregation and processing could be performed at colonial and metropolitan centres. Data were then stored in increasingly centralized colonial archives, which made it easy to recall and recycle data for novel purposes in the hands of the imperial government. It was malleable enough to manipulate and

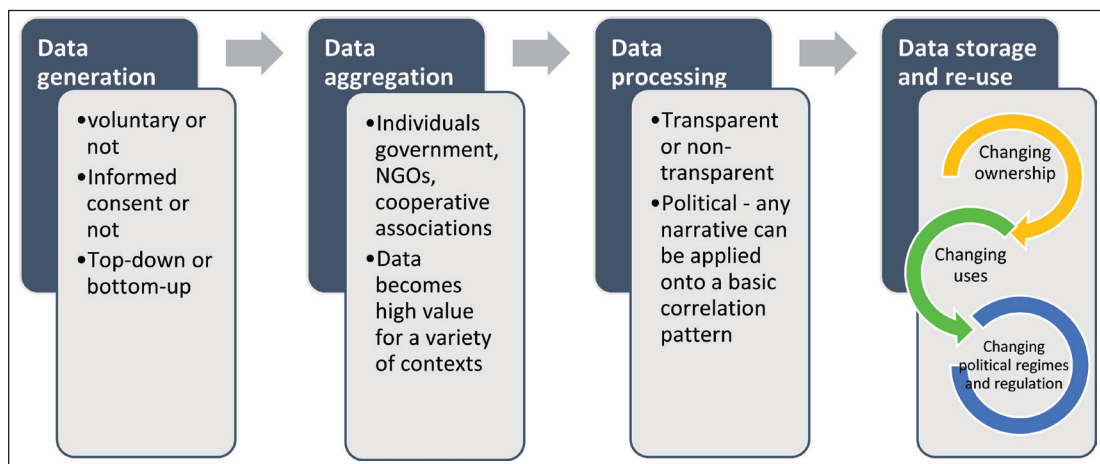


Figure 1. Data Processes—A Historical Perspective

Source: The authors.

recycle for new purposes in new contexts, yet durable enough to survive the ravages of time. It is this neglected problem of repurposing that our next section discusses in detail, engaging in discussion with the literature on data justice.

V. Recommendations from Critical Data Studies and the Historical Evidence

Historical studies offer important lessons for recent proposals in development studies that discuss the potential policy responses to the increasing datafication of twenty-first-century society. In the following section, we discuss three different approaches towards datafication from critical data studies, with a special focus on scholarship in this vein on datafication processes in the Global South. We evaluate how effective the proposed policies may be at curbing potential abuses of co-constructed data that have a long shelf-life. We focus only on the scholarship that considers data a powerful tool, while acknowledging the existence of arguments that datafication often produces faulty knowledge that cannot lead to efficient societal control (e.g., boyd and Crawford, 2012; Jerven, 2013).

Data as a Commodity

The critical data studies literature often describes datafication in the Global South as a form of capitalist extraction or 'surveillance capitalism', and sometimes termed it 'data colonialism' (Aitken, 2017; Sadowski, 2019; Thatcher et al., 2016; Zuboff, 2015). It argues that, although legal in the strict sense of the term, it is still highly problematic how global corporations acquire data as a commodity in unfavourable contractual transactions either in exchange for using a service or by explicitly paying for it (Elvy, 2017: 1407), relying on end-user license agreements (EULAs). When it comes to explicit policy recommendations, some authors in this field have suggested that potential measures could include the one-time return of data ownership and management to individuals or to particular

national governments, or to develop personal data economy companies that provide a service to individuals to ask for the return of their data (Mejias, 2020). Yet, proposals to return data to individuals or to the state suffer from some weaknesses. They assume that, once data are returned, individuals and the state will keep them forever without being coerced into new and equally problematic transactions. As our case studies have revealed, individuals tend to consent to data transactions because they are under the condition of restraint and are not fully aware of the consequences in the ensuing years or decades. This is especially relevant for many developing countries where citizens may also lack the financial resources to refuse to enter such exchanges. As our case study from the Spanish colonial empire has shown, imperial subjects sometimes engaged in exchanges of personal information for temporary benefits from the state, even though this information could and would later be turned against them. In other cases, local informants and scholars offered data to colonial scientists for free, either because they did not realize how these data would be used by a colonial administrator or because colonial actors misled them about their intentions. This is also true for present-day vulnerable populations, such as the urban poor in India volunteering information to access welfare benefits, or the smallholder farmers in Rwanda consenting to data extraction by a mobile application offering agricultural extension advice. The idea of returning data to national governments is equally problematic as these governments can have active interests in the manipulation of their citizens (Susskind, 2018). As our case study has shown, in eighteenth-century indirect rule India, local governments did provide data to English colonizers. In the context of the Global South, moreover, there has been an extensive debate about the state's handling of data privacy and the dangers of repression in countries such as Brazil, India or China, raising the spectre that initiatives for development and for government

control override concerns over the freedom of citizens (Mahrenbach et al., 2018; Singh, 2021). As discussed in Section III, the core elements of state-level data protection are often missing (e.g., the right of choice and consent, the right to access and correct, and the right to redress) and only 28% of countries on the African continent have procedures to ensure data is anonymized prior to publication (UNECA, 2018: 27). A one-time return of data either to individuals or states is, therefore, not an appropriate solution when individuals are acting under coercive situations or when there is a danger that the state itself exercises a system of surveillance.

Data as a Public Good

A second, diametrically opposed solution to the problem of datafication has been even more popular within the development studies context. Instead of returning data to individuals and the state, scholars and politicians have advocated to abolish the proprietary nature of data and to make it open access and available to everyone, usually within a public goods framework (Gurstein, 2011; Janssen et al., 2012; Kitchin, 2014). As shown in Section III, some countries in the Global South have already opted for making most of government-owned data open access in order to facilitate coming up with new solutions for fast economic development (see, e.g., the repositories of <https://dataportal.opendataforafrica.org/> by the African Development Bank, or <https://africaopendata.org/> by Code for Africa, Amazon Web Services and the World Bank). Yet, this approach fails to recognize that commons solutions tend to work only under well-defined political conditions (Mann, 2018; Ostrom, 1990) and that, once data enter the public domain, it can be exploited for various novel and unforeseen political and economic purposes. As Goldgar (1995) has shown for eighteenth-century European science, the norm of openness does not result in the equal distribution of credit or in the erasing of hierarchies. Similarly, in twentieth-century

Africa, governments sometimes relied on the same publicly available statistical data sets to make economic and planning decisions both in favour of and against imperial agendas (Morgan, 2009; Newbigin, 2020; Serra, 2014). In some cases, such data could help achieve impressive results but, as Young (2014) has pointed out, they could also be mobilized towards justifying policies that exacerbated economic inequalities across regions. Within the present-day context, there are similar concerns that open-access data, shared originally to spur innovation in development, could be captured by those few Western organizations that have the computing power and algorithms to exploit it, just as London mortality data could be exploited for calculating annuities with the mathematical expertise of the Royal Society. Indeed, according to critical data studies scholar Manovich (2011), there are three classes of people in the realm of big data: 'those who create data, those who have the means to collect it, and those who have expertise to analyse it'. As a result, open-access data policies are not sufficient by themselves to alter the inequalities inherent in existing political hierarchies.

Data as an Inalienable Right

The third critical policy alternative, associated with a larger set of emerging proposals on data justice, suggests that the ownership, use and processing of data should be limited through a variety of legal interventions, framing datafication as an issue of rights and not of intellectual property (Cinnamon, 2020; Heeks and Renken, 2018; Qureshi, 2020; Taylor, 2017). It claims data should be considered an inalienable right of people, and this legal status can serve as a guarantor against capture by powerful economic and political powers. In this literature, scholars have proposed the legal establishment of the right of data access (i.e., making data openly available to everyone), the right of data representation (i.e., the right of marginalized groups to be included in data sets), as well as the right of data ownership

and privacy, including the right to data erasure (i.e., to request data to be deleted). As this literature acknowledges, these rights can clash with each other, such as the right of data access and the right to privacy, and it is unclear how one can resolve these concerns. Yet, the data justice movement's more radical proposal has distinct benefits as compared to the previous two proposals. It allows citizens to continuously assess and re-evaluate their choices about how their data are used and it does not presume that citizens and states will always make the right decisions about data once it is returned to them.

Our historical case studies lead us to suggest an extension of the proposals of the data justice movement by pointing out the potential benefits of imposing temporal limits on the transfer of data from those who provide data to states and companies. Echoing the data justice movement's claims, we propose that data should not necessarily be considered intellectual property that one can only part with for eternity. Instead, it could be taken to be the inalienable right of people who can license it to particular, well-defined projects only for a limited amount of time. Such a proposal is fundamentally different from the claims to return data ownership. It does not make data return conditional on request and it does not turn the return of data into a one-time event. Such solutions have already been proposed in biomedical research. In some recent clinical trials, subjects have been able to determine and limit how their data and specimens could be used in future medical research, including the option to have the data erased after a certain period of time or that such data could only be used in some, but not all research projects (Master et al., 2015, cf. Starkbaum and Felt, 2019). It may be worthwhile to consider adopting such policies and debating what advantages and disadvantages the strict temporal limitation of data transfer may bring forward in the context of development.

The imposition of temporal limits on data transfer is advantageous because, as we have

seen, longevity is one of the key issues present in datafication, present in all of the historical case studies discussed in the previous section. The value and meaning of data change across time, and those who own data can harvest new knowledge and gain financial profit through the development of novel processing methods. Those early seventeenth-century searchers who collected mortality data for London did not envisage that this information could be used decades later to justify and promote particular policies for dealing with the poor. Those who shared local knowledge of cinchona in Peru did not necessarily realize that it would become one of the best-selling drugs of the eighteenth-century Spanish Empire. Arguably, if data and information sharing agreements have a temporal limit (or a 'sunset clause'), they allow data providers to negotiate more profitable agreements again and again in the future. Even if consent about data use is engineered and obtained in unequal situations, a potential for change and renegotiation occurs when the license expires and needs to be renegotiated. And while temporal licensing cannot completely erase the dangers of data capture in case of radical political change, it does limit the damage by ensuring that, at least until political change happens, the providers of data receive the chance to regularly and periodically determine what to do with their data.

Our proposal to consider data an inalienable right that can be licensed only for a limited time has focused on the context of development in the Global South, but its lessons may also be applied to similar political and economic situations across the globe. Datafication, after all, is a problematic process both in high-income and low- and middle-income countries. We do not claim that our policy suggestions will resolve all issues related to datafication. For example, corporations have been making the argument to establish new types of data, for example, 'urban data', that are not personal and private, suggesting that there is, therefore, no need to consider the issues of privacy or

Table 2. Regulation APPROACHES to Data

<i>Approaches to Data</i>	<i>Data as a Commodity</i>	<i>Data as a public Good</i>	<i>Data as an Inalienable Right</i>	<i>Our Proposition</i>
Definition	Data are collected and circulated as a commodity by governments and firms	Data are non-rival resource (public good) that can best benefit society when everyone has access to it	Data are intrinsically tied to people, but it can be exploited by those in power	Data are mobile and non-perishable common good that are often exploited by governments and firms
Legal/regulation approach	Intellectual property rights and EULAs	Open data schemes	Equitable and just sharing of data by public regulation	Data as a special, personal right that should only be alienated temporarily
References	Sadowski (2019), Aitken (2017), Thatcher et al. (2016), Fourcade and Healy (2017)	Kai et al. (2019), Janssen et al (2012), Kitchin (2014), Gurstein (2011)	Cinnamon (2020), Qureshi (2020), Taylor (2017), Arora (2016)	Master et al. (2015)

Source: The authors.

the compensation of original data providers (Goodman and Powles, 2019: 472). Yet, in so far as the concept of private, personal data remain relevant in the datafication discourse, our proposal does allow for some amelioration of the current regime of data ownership and extraction. It makes corporate capture more difficult by offering the chance to providers for reconsidering their options, and it opens the opportunity for further reform when the general political and economic situation allows citizens to make a stance for effective change.

Table 2 presents the comparison of the three critical data studies approaches and shows how our proposition extends and expands the ‘data as an inalienable right’ approach (shaded column on the right).

VI. Conclusion

In this article, we have argued that a long-term historical perspective is necessary to properly analyse the promises and potential of big data in relation to sustainable development. Building on critical data studies, we showed that existing approaches mistakenly frame

datafication as a novel phenomenon, driven by the recent advances in information and communication technologies. European states have relied on large-scale data sets for managing their colonies at least since 1500, and they never failed to provide utopian narratives that claimed these efforts would bring tangible benefits to the populations affected. History has seen again and again how governments and agencies resort to the large-scale collection of data (including data on the private lives of people) and their mathematical analysis to drive development. Arguably, the persistent decoupling of colonial datafication and recent datafication processes is both deliberate and political. Positioning current datafication as a new and original phenomenon comes with a promise of a positive, technology-driven social change. Regrettably, this line of thinking also dominates the current policy approaches, advocating for open data schemes within the public policy domain.

Based on the historical findings and discussions in this article, we urge scholars and practitioners to approach such proposals

with caution. Our cases illustrated that data are virtually imperishable. Once collected, it acquires a life on its own. Against this background, our historical perspective also provides a caveat against proposed solutions that want to control the power of multinational corporations by returning ownership of data to national governments. We propose to see data as the special property of people that can only be alienated temporarily, for strictly defined purposes within a strictly defined time frame, similar to data use in medical studies. Such a strong limit seriously curtails what governments and corporations can do with aggregated data and may block both positive and negative developments without discrimination. Yet if one is seriously concerned about the potential abuses of data, such an approach should be put out on the table and it should be considered alongside other proposals that lay emphasis on other aspects of datafication. And if our historical case studies have served a purpose, it was to show that one should be seriously concerned.

Acknowledgements

The authors are grateful to the two anonymous reviewers for providing all the useful comments and suggestions. They would like to express their thanks to the participants of the Knowledge, Technology, Innovation Seminar at the University of Wageningen and to Dr Federico D'Onofrio for the inspiring discussions of the article content. Dr Katarzyna Cieslik acknowledges the generous support of the Philomathia Social Science Foundation.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.


Funding

The authors received no financial support for the research, authorship and/or publication of this article.

Author Affiliations

Katarzyna Cieslik is the corresponding author (kc566@cam.ac.uk) and is with the Department of Geography, University of Cambridge, UK. Daniel Margócsy is with the Department of History and Philosophy of Science, University of Cambridge, UK.

ORCID iD

Katarzyna Cieslik  <https://orcid.org/0000-0002-8240-0117>

Notes

1. See <https://www.opengovpartnership.org/>
2. See <https://www.od4d.net/>
3. See <https://caravanmagazine.in/reportage/aadhaar-mixing-public-risk-private-profit>

References

- Abdulrauf, A.L.** 2021: Giving 'teeth' to the African Union towards advancing compliance with data privacy norms. *Information & Communications Technology Law* 30(2): 87–107.
- Aitken, R.** 2017: All data is credit data: constituting the unbanked. *Competition and Change* 21: 274–300.
- Amankwah-Amoah.** 2016: Emerging economies, emerging challenges: mobilising and capturing value from big data. *Technological Forecasting and Social Change* 110: 167–74.
- Anderson, C.** 2008: The end of theory: the data deluge makes the scientific method obsolete. *Wired*. Available at: <https://www.wired.com/2008/06/pb-theory/> (accessed 25 January 2022).
- Appadurai, A.** 1993: Number in the colonial imagination. In: Breckenridge, C.A. and van der Veer, P. editors, *Orientalism and the Postcolonial Predicament: Perspective on South Asia*. University of Pennsylvania Press, pp. 314–39.
- Arora, P.** 2016: The bottom of the data pyramid: big data and the Global South. *International Journal of Communication* 10: 1681–99.
- Arora, P.** 2018: Decolonizing Privacy Studies. *Television & New Media* 20(4): 366–78.
- Asif, M.A.** 2019: Technologies of power—from area studies to data sciences. *Spheres* 5. Available at: <https://spheres-journal.org/contribution/technologies-of-power-from-area-studies-to-data-sciences/> (accessed 25 January 2022).
- Biruk, C.** 2018: *Cooking Data: Culture and Politics in an African Research World*. Duke University Press.
- Blair, A.** 2010: *Too Much to Know: Managing Scholarly Information before the Modern Age*. Yale University Press.
- Bleichmar, D.** 2012: *Visible Empire: Botanical Expeditions and Visual Culture in the Hispanic Enlightenment*. University of Chicago Press.
- Bleichmar, D.** 2015: The imperial visual archive: images, evidence, and knowledge in the early modern Hispanic world. *Colonial Latin American Review* 24, 236–66.
- Bouk, D.** 2017: The history and political economy of personal data over the last two centuries in three acts. *Osiris* 32(1): 85–106.

- 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): 662–79.
- Breckenridge, K.** 2014: *Biometric State: The Global Politics of Identification and Surveillance in South Africa, 1850 to Present*. Cambridge University Press.
- Brendecke, A.** 2016: *The Empirical Empire: Spanish Colonial Rule and the Politics of Knowledge*. De Gruyter.
- Bronson, K. and Knezevic, I.** 2016: Big Data in food and agriculture. *Big Data & Society*. DOI: 10.1177/2053951716648174.
- Buck, P.** 1978: Seventeenth-century political arithmetic: civil strife and vital statistics. *Isis* 68: 67–84.
- Burns, R.** 2015: Rethinking big data in digital humanitarianism: Practices, epistemologies, and social relations. *GeoJournal* 80(4): 477–90.
- Cañizares-Esguerra, J.** 2002: *How to Write the History of the New World: Histories, Epistemologies, and Identities in the Eighteenth-Century Atlantic World*. Stanford University Press.
- Canizares-Esguerra, J.** 2006: *Nature, Empire, and Nation: Explorations in the History of Science in the Iberian World*. Stanford University Press.
- Carletto, C., Jolliffe, D. and Banerjee, R.** 2015: From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *The Journal of Development Studies* 51(2): 133–48.
- Chambers R.** 1997: *Whose Reality Counts: Putting the First Last*. Intermediate Technology Publications.
- Cieslik, K., Leeuwis, C., Dewulf, A., Feindt, P., Lie, R., Werners, S. van Wessel, M. and Struik, P.** 2018: Addressing socio-ecological development challenges in the digital age: environmental virtual observatories for connective action. *Journal of the Royal Netherlands Society for Agricultural Sciences (NJAS)* 86–87: 2–11.
- Cinnamon, J.** 2020: Data inequalities and why they matter for development. *Information Technology for Development* 26: 214–33.
- Coleman, D.** 2018: Digital colonialism: the 21st-century scramble for Africa through the extraction and control of user data and the limitations of data protection laws. *Michigan Journal of Race & Law* 24 (2): 417–40.
- Crawford, K. and Schultz, J.** 2014: Big data and due process: toward a framework to redress predictive privacy harms. *Boston College Law Review* 55: 93–128.
- Crawford, M. J.** 2016: *The Andean Wonder Drug: Cinchona Bark and Imperial Science in the Spanish Atlantic, 1630–1800*. University of Pittsburgh Press.
- Cukier, K. and Mayer-Schoenberger, V.** 2013: The rise of big data: How it's changing the way we think about the world. *Foreign Affairs* 92: 28–36.
- D'Onofrio, F.** 2016: *Observing Agriculture in Early Twentieth-Century Italy: Agricultural Economists and Statistics*. Routledge.
- Dalton C. and Thatcher J.** 2014: What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'big data'. *Society & Space*. Available at: <https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care> (accessed 25 January 2022).
- Daston, L.** 2017: *Science in the Archives: Pasts, Presents, Futures*. University of Chicago Press.
- Davidson, J.** 2018: Bayer, Monsanto and Big Data: who will control our food system in the era of digital agriculture and mega-mergers? *Medium: Friends of the Earth*. Available at: https://medium.com/@foe_us/bayer-monsanto-and-big-data-who-will-control-our-food-system-in-the-era-of-digital-agriculture-aae80d991e4d (accessed 25 January 2022).
- de Corbion, A.P., Hosein, G., Fisher, T., Geraghty, E., Quintanilla, J., Marelli, M. and Pelucchi, S.** 2018: The humanitarian metadata problem: 'doing no harm' in the digital era. *Privacy International and ICRC*. Available at: <https://privacyinternational.org> (accessed 25 January 2022).
- Demombynes, G. and Sandefur, J.** 2014: Costing a data revolution. CGD Working Paper 383. Center for Global Development. Available at: <http://www.cgdev.org/publication/costing-data-revolution-working-paper-383> (accessed 25 January 2022).
- Dencik, L.** 2020: Situating Practices in Datafication – from Above and Below. In: Stephansen, H.C. and Trerè, E., editors, *Citizen Media and Practice*. Routledge, pp. 243–55.
- Dennis, J.** 2015: *Writing, Publishing, and Reading Local Gazetteers in Imperial China, 1100–1700*. Harvard University Press.
- Deringer, W.** 2018: *Calculated Values: Finance, Politics, and the Quantitative Age*. Harvard University Press.
- Dirks, N.** 1993: Colonial histories and native informants: The biography of an archive. In: Breckenridge, C. and Appadurai, A., editors, *Post Colonialism and the Predicament of History*. Oxford University Press.
- Elvy, S.A.** 2017: Paying for privacy and the personal data economy. *Columbia Law Review* 117: 1369–460.
- Etzo, S. and Collender, G.** 2010: The mobile phone 'revolution' in Africa: rhetoric or reality? *African Affairs* 109: 659–68.
- Fanon, F.** 2008: *Black Skin, White Masks*. Grove Books.
- Fourcade, M. and Healy, K.** 2017: Seeing like a market. *Socio-Economic Review* 15(1): 9–29.
- Gänger, S.** 2020: *A Singular Remedy Cinchona Across the Atlantic World, 1751–1820*. Cambridge University Press.

- Gelb, A.** and **Metz, A.D.** 2018: *Identification Revolution: Can Digital ID be Harnessed for Development?* Brookings Institution Press.
- Goldgar, A.** 1995: *Impolite Learning: Conduct and Community in the Republic of Letters, 1680–1750*. Yale University Press.
- Goodman, E.P.** and **Powles, J.** 2019: Urbanism under Google: lessons from sidewalk Toronto. *Fordham Law Review* 88(2): 457–98.
- Goody, J.** and **Watt, I.** 1963: The consequences of literacy. *Comparative Studies in Society and History* 1: 304–45.
- Graunt, J.** 1676: *Natural and Political Observations*. John Martyn.
- Greenleaf, G.** and **Cottier, B.** 2020: 2020 Ends a Decade of 62 New Data Privacy Laws. *163 Privacy Laws & Business International Report* 24–26, Available at: <https://ssrn.com/abstract=3572611> (accessed 25 January 2022).
- Guha, S.** 2003: The politics of identity and enumeration in India c. 1600–1990. *Comparative Studies in Society and History* 45: 148–67.
- Gurstein, M.B.** 2011: Open data: Empowering the empowered or effective data use for everyone. *First Monday* 16. DOI: 10.5210/fm.v16i2.3316.
- Habib, I.** 2013: *The Agrarian System of Mughal India, 1556–1707*. Oxford University Press.
- Harley, J.B.** 1989: Deconstructing the map. *Cartographica: The International Journal for Geographic Information and Geovisualization* 26(2): 1–20.
- Heeks, R.** and **Renken, J.** 2018: Data justice for development: What would it mean? *Information Development* 34: 90–102.
- Heitman, K.** 2020: Authority, autonomy and the first London Bills of Mortality. *Centaurus* 62: 275–84.
- Hilbert, M.** 2016: Big data for development: a review of promises and challenges. *Development Policy Review* 34: 135–74.
- Hosein, G.** and **Nyst, C.** 2013: Aiding surveillance: an exploration of how development and humanitarian aid initiatives are enabling surveillance in developing countries. Available at: <https://ssrn.com/abstract=2326229> (accessed 25 January 2022). DOI: 10.2139/ssrn.2326229
- Iazzolino, G.** and **Mann, L.** 2019: Harvesting data: who benefits from platformization of agricultural finance in Kenya? *Developing Economics*. Available at: <https://developingeconomics.org/2019/03/29/harvesting-data-who-benefits-from-platformization-of-agricultural-finance-in-kenya/> (accessed 25 January 2022).
- IEAG.** 2014: A world that counts mobilising the data revolution for sustainable development. Available at: <https://www.undatarevolution.org/report/> (accessed 25 January 2022).
- Janssen, M., Charalabidis, Y.** and **Zuiderwijk, A.** 2012: Benefits, adoption barriers and myths of open data and open government. *Information Systems Management* 29, 258–68.
- Jerven, M.** (2013) *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*. Cornell University Press.
- Kamilaris, A., Kartakoullis, A.** and **Prenafeta Boldú, F.** 2017: A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture* 143: 23–37.
- Khera R.** 2019: Introduction. In: Khera, E., editor. *Dissent on Aadhaar: Big Data Meets Big Brother*. Orient BlackSwan.
- Kirsch, S.** 2014: Insular territories: US colonial science, geopolitics, and the (re)mapping of the Philippines. *The Geographical Journal* 182. DOI: 10.1111/geoj.12072.
- Kitchin, R.** 2014: *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE Publications.
- Kleine, D.** and **Unwin, T.** 2009: Technological revolution, evolution and new dependencies: What's new about ICT4D. *Third World Quarterly* 30:1045–67.
- Knaap, G.** 1987: *Memories van Overgave van Gouverneurs van Ambon in de zeventiende eeuw* [Retrospective reports of the Governors of Ambon in the seventeenth century]. RGP.
- Krajewski, M.** 2011: *Paper Machines: About Cards & Catalogs, 1548–1929*. MIT Press.
- Kshetri, N.** 2019: Cybercrime and cybersecurity in Africa. *Journal of Global Information Technology Management* 22(2): 77–81.
- Kwet, M.** 2019: Digital colonialism: US empire and the new imperialism in the Global South. *Race and Class* 60(4): 3–26.
- Latour, B.** 1986: Visualisation and cognition: drawing things together. *Knowledge and Society* 6: 1–40.
- Leonelli, S.** and **Tempini, N.** 2020: *Data Journeys in the Sciences*. Springer.
- Lycett, M.** 2013: 'Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems* 22: 381–6.
- Mahrenbach, L., Mayer, K.** and **Pfeffer, J.** 2018: Policy Visions of big data: views from the Global South. *Third World Quarterly* 39: 1861–82.
- Makulilo, A.B.** 2016: A person is a person through other persons. A critical analysis of privacy and culture in Africa. *Beijing Law Review* 7: 192–204.
- Mann, L.** 2018: Left to other peoples' devices? A political economy perspective on the big data revolution in development. *Development and Change* 49: 3–36.
- Mann, L.** and **Hilbert M.** 2020: AI4D: artificial intelligence for development. *International Journal of Communication* 14: 4385–405.
- Manovich, L.** 2011: Trending: the promises and the challenges of big social data. In: Gold, M.K., editor. *Debates in the Digital Humanities*. The University of Minnesota Press.

- Master, Z., Campo-Engelstein, L. and Caulfield, T.** 2015: Scientists' perspectives on consent in the context of biobanking research. *European Journal of Human Genetics* 23: 569–74.
- Mathur, A.** 2020: Will India's new data protection law serve as a government surveillance tool? The Global Legal Post. Available at: <https://www.globallegalpost.com/news/will-india39s-new-data-protection-law-serve-as-a-government-surveillance-tool-4261027> (accessed 25 January 2022).
- McCormick, T.** 2009: *William Petty and the Ambitions of Political Arithmetic*. Oxford University Press.
- Mejias, U.A. and Couldry, N.** 2019: Datafication. *Internet Policy Review* 8: 1–10.
- Mejias, U. A.** 2020, 8 September: *To fight data colonialism, we need a Non-Aligned Tech Movement. This is one way to ensure technology does not go against the interest of society*. Al Jazeera Opinion. Available at: <https://www.aljazeera.com/opinions/2020/9/8/to-fight-data-colonialism-we-need-a-non-aligned-tech-movement>
- Merry, S.E., Davis, K.E. and Kingsbury, B.,** editors. 2015: *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*. Cambridge University Press.
- Mertia, S.** 2020: Introductions: relationalities abound. In Mertia, S., editor. *Lives of Data: Essays on Computational Cultures from India*. Institute of Networked Cultures. <https://networkcultures.org/wp-content/uploads/2020/12/LivesofData.pdf> (accessed 25 January 2022).
- Milan, S. and Treere, E.** 2019: Big data from the south(s): beyond data universalism. *Television and New Media* 20: 319–35.
- Mo Ibrahim Foundation (MIF).** 2019: *Agendas 2063 & 2030: is Africa on track?* Africa Governance Report. Available at: <https://mo.ibrahim.foundation/iiag/gr-2019-key-findings> (accessed 25 January 2022).
- Morgan, M.S.** 2009: Seeking parts, looking for wholes. *History of Observation in Economics Working Paper* 1. Available at: <http://eprints.lse.ac.uk/id/eprint/32423> (accessed 25 January 2022).
- Munkhoff, R.** 1999: Searchers of the dead: authority, marginality, and the interpretation of plague in England, 1574–1665. *Gender and History* 11: 1–29.
- Newbigin, E.** 2020: Accounting for the nation, marginalizing the empire: taxable capacity and colonial rule in the early twentieth century. *History of Political Economy* 52, 455–72.
- Newton, I.** 1669: Letter to Francis Aston, 18 May. CUL MS Add 9597/2/18.4.
- Olowogboyega, O.** 2020: *A deep dive into the proposed guidelines for e-hailing companies in Lagos*. Techcabal, 11 August. Available at: <https://techcabal.com/2020/08/11/guidelines-e-hailing-lagos/> (accessed 25 January 2022).
- O'Neil, C.** 2013: *On Being a Data Skeptic*. O'Reilly.
- Ostrom, E.** 1990: *Governing the COMMONS: The Evolution of Institutions for Collective Action*. Cambridge University Press.
- Peabody, N.** 2001: Cents, Sense, census: human inventories in late precolonial and early colonial India. *Comparative Studies in Society and History* 43: 819–50.
- Portuondo, M.** 2009: *Secret Science: Spanish Cosmography and the New World*. Johns Hopkins University Press.
- Privacy International.** 2020: 2020 is a crucial year to fight for data protection in Africa. Available at: <https://privacyinternational.org/long-read/3390/2020-crucial-year-fight-data-protection-africa> (accessed 25 January 2022).
- Qamber, R.** 2006: Inquisition proceedings against Muslims in 16th-century Latin America. *Islamic Studies* 45: 21–57.
- Qureshi, S.** 2020: Why data matters for development? Exploring data justice, micro-entrepreneurship, mobile money and financial inclusion. *Information Technology for Development* 26: 201–13.
- Raj, K.** 2017: Networks of knowledge, or spaces of circulation? The birth of British cartography in colonial South Asia in the late eighteenth century. *Global Intellectual History* 2(1): 49–66.
- Reidy, M.S.** 2008: *Tides of history: Ocean Science and Her Majesty's Navy*. University of Chicago Press.
- Rodriguez, K. and Alimonti, V.** 2020: *A Look-Back and Ahead on Data Protection in Latin America and Spain*. Electronic Frontier Foundation. <https://www.eff.org/deeplinks/2020/09/look-back-and-ahead-data-protection-latin-america-and-spain> (accessed 25 January 2022).
- Sadowski, J.** 2019: When data is capital: datafication, accumulation, and extraction. *Big Data & Society* 7: 1–12.
- Schaffer, S.** 2009: Newton on the beach: the information order of principia mathematica. *History of Science* 47: 243–76.
- Schiebinger, L.** 2004: *Plants and Empire: Colonial Bioprospecting in the Atlantic World*. Harvard University Press.
- Shapin, S. and Schaffer, S.** 1985: *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton University Press.
- Schiebinger, L. and Swan, C.,** editors. 2005: *Colonial Botany: Science, Commerce, and Politics in the Early Modern World*. University of Pennsylvania Press.
- Scott, J.C.** 1998: *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.
- Serra, G.** 2014: An uneven statistical topography: the political economy of household budget surveys in late colonial Ghana, 1951–1957. *Canadian Journal of Development Studies* 35(2014): 9–27.

- Sieber, R.E.** and **Johnson, P.A.** 2015: Civic open data at a crossroads: dominant models and current challenges. *Government Information Quarterly* 32: 308–15.
- Singh, P.** 2021: Aadhaar and data privacy: biometric identification and anxieties of recognition in India. *Information, Communication & Society* 24(7): 978–93.
- Slack, P.** 2004: Government and information in seventeenth-century England. *Past and Present* 184: 33–68.
- Slade, F.D.** 2011: An imperial knowledge space for Bourbon Spain: Juan Bautista Muñoz and the founding of the Archivo General de Indias. *Colonial Latin American Review* 20(2): 195–212.
- Starkbaum, J.** and **Felt, U.** 2019: Negotiating the reuse of health-data: research, big data, and the European general data protection regulation. *Big Data & Society*. Available at <https://doi.org/10.1177/2053951719862594> (accessed 25 January 2022).
- Susskind, J.** 2018: *Future Politics*. Oxford University Press.
- Sutherland, E.** 2018: Digital privacy in Africa: cybersecurity, data protection & surveillance. Available at: <https://ssrn.com/abstract=3201310> or <http://dx.doi.org/10.2139/ssrn.3201310> (accessed 25 January 2022).
- Taylor, L.** 2017: What is data justice? The case for connecting digital rights and freedoms globally. *Big Data and Society* 4: 1–14.
- Taylor, L.** and **Broeders, D.** 2015: In the name of development: power, profit and the datafication of the Global South. *Geoforum* 64: 229–37.
- Thatcher, J., O'Sullivan, D.** and **Mahmoudi, D.** 2016: Data colonialism through accumulation by dispossession: new metaphors for daily data. *Environment and Planning D* 34: 990–1006.
- UNECA.** 2018: Africa data revolution report 2018: the status and emerging impact of open data in Africa. Available at: <https://repository.uneca.org/handle/10855/43341> (accessed 25 January 2022).
- United Nations.** 2013: Communique. Meeting of the high level panel of eminent persons on the post 2015 development agenda in Bali, Indonesia, 27 March. Available at: <https://www.un.org/sg/sites/www.un.org.sg/files/documents/management/Final%20Communique%20Bali.pdf> (accessed 25 January 2022).
- Wagoner, P.B.** 2003: Precolonial intellectuals and the production of colonial knowledge. *Comparative Studies in Society and History* 45(4): 783–814.
- Wiseman, L., Sanderson, J., Zhang, A.** and **Jakku, E.** 2019: Farmers and their data: an examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming. *NJAS—Wageningen Journal of Life Sciences* 90–91. Available at: <https://doi.org/10.1016/j.njas.2019.04.007> (accessed 25 January 2022).
- Ylijoki, O.** and **Porras, J.** 2016: Perspectives to definition of big data: a mapping study and discussion. *Journal of Innovation Management* 4(1): 69–91.
- Young, A.** 2014: Measuring the Sudanese Economy: a focus on national growth rates and regional inequality, 1959–1964. *Canadian Journal of Development Studies* 35: 44–60.
- Zuboff, S.** 2015: Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30: 75–89.