Erik Koenen, Christian Schwarzenegger and Juraj Kittler

Data(fication)

"Understanding the World Through Data" as an Everlasting Revolution

Abstract: In this chapter we set out to historicize and trace the pre-digital roots of the concept of datafication of communication and society. Collecting and processing data as well as governing data storage and access to it are not to be seen as a particularity of the digital era. Data and datafication produced, already long before the digital revolution, exclusive arrangements of infrastructures and knowledge orders and they can hence be seen as building blocks of culture and society. We illustrate this argument in four steps using different historic examples. We first provide a glimpse into the beginnings of datafication in ancient times. We then present data as early social science instruments in the modern welfare states since the mid-nineteenth century used for social control and to grasp facets and consequences of social modernization. Thirdly, data were also crucial in the service of oppression during the National Socialist era, in which cutting-edge data technologies contributed to the planning and implementation of the Holocaust. Finally, the shift of data from the numerical to the digital information age in the second half of the twentieth century and its consequences for a "datafication of everything" is discussed.

Keywords: big data, data processing, social engineering, technological solutionism, dark data, social history

The digital age is an age of data. Digital communication research rhetoric suggests an unprecedented relevance and depth of data for virtually all domains of social and everyday life and for shaping the very construction and perception of reality. Data is seen as a fabric, output, fuel, lubricant, and currency of the digital age. Digital communication is based on algorithms, metrics and complex processes of datafication. Whereas in principle everything can be turned into data (Mejias and Couldry 2019), the notion of datafication refers to processes of rendering information into machine-readable quantifiable data for the purpose of aggregation, analysis, and anticipation of human behavior and social interaction (Mayer-Schönberger and Cukier 2013; Mau 2019; Southerton 2020). In this trajectory, Couldry and Hepp (2017, 34–56) have argued that datafication marks an entirely new wave of mediatization, causing profound transformations of our everyday life as well as the organization and evaluation of social

https://doi.org/10.1515/9783110740202-008

issues. Following Southerton's (2020, 1) contribution to the "Encyclopedia of Big Data", datafication is also used to describe "a logic that sees things in the world as sources of data to be 'mined' for correlations or sold . . . ." Data is hence occasionally referred to as the new oil or the most valuable resource, which to control promises wealth, power, and influence in shaping the social world. Hence, datafication "combines two processes: the transformation of human life into data through processes of quantification, and the generation of different kinds of value from data" (Mejias and Couldry 2019).

But, putting aside the rhetoric of technological newness and the fascination for the peculiarity of the current historical moment, we can follow Rödder and ask how new all of this truly is (Rödder 2015). And if so, to what extent and regarding what respects are the processes of digital datafication novel. When the notion of datafication was originally introduced by Mayer-Schönberger and Cukier (2013), it was coined with regard to the economy of digital platforms and since used in relation to processes of making virtually all aspects of human behavior processable for large scale Big Data analysis. Data and the interpretation of data in this respect is seen as a means of looking at and understanding the world (Borgman 2016; boyd and Crawford 2012; Mejias and Couldry 2019; van Dijck 2014). Recent publications have argued that data(fication) should not be viewed as a unique characteristic of the digital era, but rather as a historical category and as an ongoing historical process with manifold precursors (Beer 2016; Borck 2017; Aronova, von Oetzen, and Sepkowski 2017; van Es and Eef Masson 2018). In a similar vein, Beer (2016) has called for understanding "Big Data" "both as a material phenomenon and as a concept", which is changing historically and only reached a new level in the digital era. "Big Data," Aronova and colleagues (2017, 7) furthermore argue, "is often associated with the era of digital electronic databases, but this association potentially overlooks important continuities with data practices stretching back to much earlier material cultures. While technologies have changed – from paper-based to mechanical to electronic devices – database practices have been more continuous than the technologies and tools" (Aronova, von Oetzen, and Sepkowski 2017, 7).

In this chapter, we set out to historicize the concept of datafication of communication and society and trace the pre-digital roots of what is nowadays considered a defining process of the digital era. Rather than imagining data and datafication as a particularity of the digital era, we provide a perspective on data as a category and datafication as a historical process. Our perspective is inspired by Raphael's (1996) take on the scientification of the social. In a perspective on a social history of data and datafication, it becomes evident that gathering data was never a naïve project but always happened on purpose, with particular goals and intents regarding its impact on shaping social reality.

Collecting and processing data as well as governing data storage and access to using it was produced already way before the digital revolution's exclusive arrangements of infrastructures and knowledge orders. Furthermore, datafication has long since amounted to distinct communicative and social practices. A look back at earlier epochs can thus help us to compare today's processes of datafication with previous historical contexts and put their alleged uniqueness in perspective.

We illustrate this argument in four steps using different historic periods as examples: before we leap into the mid-nineteenth-century and the role of datafication of the social through the emerging social sciences, we first provide 1) a glimpse on the beginnings of datafication in ancient times. We then present 2) data as early social science instruments in the modern welfare states used since the mid-nineteenth century for social control and to grasp facets and consequences of social modernization; 3) data in the service of oppression during the National Socialist era, in which cutting-edge data technologies (e.g., Hollerith-systems) contributed to the planning and implementation of the Holocaust; 4) the shift of data from the numerical to the digital information age in the second half of the twentieth century in which the new computer technologies rapidly increased the processes of datafication and made data gathering increasingly a private and corporate enterprise. When Mayer-Schönberger and Cukier (2013) coined the notion of datafication, they predicted Big Data would bring a revolution that will transform how we live, work, and think. Historicizing datafication, we show that the aspiration of "understanding the world through numbers" is indicative of old hopes, renewed promises and an everlasting revolution rather than a singular turning point.

1 Data(fication) as a Building Block of Culture and Society – From Early Cultures to the Age of Enlightenment

It can be argued that the history of datafication is closely intertwined with the history of writing and hence an early wave of mediatization (Herrenschmidt 2007; Krotz 2012; Ong 1982). According to anthropologists and historians, the main driving forces behind the development of symbolic forms of data gathering and recordkeeping were the necessity of the earliest complex social formations to keep records of their legislative measures and taxes. Thus, the ability of the state to coordinate substantial public works and the need of private entrepreneurs to

keep records of their business operations were also matters of record keeping from the beginning (van de Mieroop 1999, 13; Law et al. 2015, 212). Such tangible records were external to an individual, which means that they could be inspected, verified, and cross-examined (Law et al. 2015, 208). Data collection hence from the very beginning of written records became a paramount objective for governing bodies as well as for private individuals and families engaged in complex webs of economic activities and logistics. Data collections contained the documentation of loans, sales, and rentals as well as public records of military and religious personnel and were kept in private and public administrative archives, which can be found practically everywhere in Mesopotamia where cuneiform writing was diffused.

Additionally, the sophisticated urban civilization of the Greeks soon articulated the need for public and private record keeping (Sickinger 1999). From the speeches of Attic orators, valuable insights can be gleaned not only about the quantity and the places in which they were stored but also the cultural, economic, and political importance ascribed to these records by their contemporaries. "An excellent thing, fellow citizens, an excellent thing is the preservation of the public acts," claimed Athenian politician Aeschynes (3.75) in 330 B.C.E. "For the record remains undisturbed, and does not shift sides with political turncoats, but whenever the people desire, it gives them opportunity to discern who have been rascals of old but have now changed face and claim to be honorable men" (Aeschines 1958, 367–369). Thus, anticipating later developments, already in ancient Greece the collection of data holds out the prospect of objective and undisputed access to social reality. Data should make it possible to open up reality independently of bias, whim, and interpretation.

Less direct knowledge exists about the record-keeping and data gathering activities in Classical Rome. Contrary to Greek social life which took place in the public square, Roman society, even during the republican era, was organized on the principle of patronage where wealthy and influential patrons received their clients in their own homes. Therefore, each patron's household had a separate space storing private as well as important public records belonging to the office administered by the head of the household (Culham 1989, 104). We hence can identify widely privatized and decentralized practices for the collection and storage of data. Historical evidence indicates that even such essential public data as the census, the masterpiece of Roman bureaucracy and backbone of its taxation system, were not gathered, processed, and stored in one central location.

In the late medieval period, the earliest parish registries appeared but it was only during the Reformation when they became mandated by the Catholic Church, and subsequently also by various Protestant Churches in order to keep track of their own populations (Emigh, Riley, and Ahmed 2015, 174–175). After

the Council of Trent (1545–1563), the Catholic Church required all parish priests to keep registers containing lists of persons who received the basic sacraments, such as baptism or marriage, or who had a Christian burial. By doing this, it *de facto* created a permanent, standardized census tracking the fundamental demographic data by registering those who were born, got married and died (Culham 1989, 105).

The relatively high degrees of literacy and numeracy required in the commercial world were simultaneously translated into the civic life of the late medieval urban communities and reflected in the fact that medieval cities and city-states started organizing their own information-gathering and record-keeping systems (Barber 1992, 258).

Historians argue that it was the reign of Louis XIV in France which for the first time demonstrated that an early modern nation-state could prosper despite its relatively vast and differentiated geography. This was possible mainly due to the improving systems of transportation and communication but also to the collection and management of large amounts of statistical data. Jean-Baptiste Colbert (1619–1683), the king's chief minister, is often presented as a posterchild of this trend. Over his two decades of service to the king, Colbert amassed vast amounts of information through the various sectors of public administration that he oversaw – from finance and the navy to foreign diplomacy – using data as a foundation for the rationalization of state operations (Soll 2009). In this effort, he harnessed many of the techniques developed previously by scholars, merchants, and churchmen and systematically applied them to government.

Following Mejias and Couldry (2019), and as illustrated by the examples above, datafication began in the domain of business and administration and not social life. The collection of data was closely linked to those in politically or economically privileged positions (positions of power) for doing so and the command over data helped to reinforce and expand their power. The data collected typically addressed those bound to the powerful through contracts or other obligations and the purpose of the data was to govern, administer, coordinate and control and further accumulate privilege. But data could also be used to rationalize processes based on datafied knowledge. While the records that were kept are partially of surprisingly rich detail, for instance contracts in late medieval Genova could not only hold the dates but also the hours and minutes of when the contracts were made if this information was deemed necessary, it is also a common pattern that the data gathered was more or less directly linked to its initial purpose and focused on essential elements. It was only later that the collection of data expanded from the functional collection of data to the "datafication of everything" (Mayer-Schönfelder and Cukier 2013, 93–94), with the functions of the data only to be discovered later in the vast collections.

Datafication practices from their very historical precursors and origins were located at the intersections of (recordable) knowledge and power. The study of datafication, in historical as well as contemporary contexts, is thus also a critical perspective on processes of gathering and capturing human experience in data and processing it for economic or political purposes.

2 Open the Box: The Scientification of Data and the Datafication of Social Life

As shown, the practices of experiencing, collecting and organizing data can already be observed from early cultures to the Age of Enlightenment. With the advent of modernity, a new period begins in which societies develop and explore instruments to systematically observe and describe themselves with data, not least in order to manage, organize, and regulate social processes based on this data.

The very ideas that datafication opens vistas for beholding the world and that society can be discovered, explored and understood through data resonate in the aspirations regarding the potential of the then emerging social sciences. "[W]e can in principle *control everything by means of calculation*," as Max Weber (2004, 13) stated. The "disenchantment of the world" through the modern social sciences, which Max Weber (1864–1920) told to his Munich students a hundred years ago, was above all to be achieved by the accumulation of data, calculations, and the technical means which allowed for the collection and processing of data. "The nineteenth century," Jürgen Osterhammel argues, "can be seen as the century of counting and measuring. The idea of an all-embracing taxonomy now grew into a belief that the power of number – of statistical processing or even 'social mathematics', as the Marquis de Concordet, a bright star of the late Enlightenment, put it – could open up truth itself to human reason. It was in the nineteenth century that societies measured themselves for the first time and archived the results" (Osterhammel 2014, 29).

Over the course of the nineteenth century, social data collection became a common practice in Europe and the US (Burke 2012), providing spaces for "ongoing institutional self-observation" (Osterhammel 2014, 25) and helping statistics to become "what it is today: the most important tool for the constant self-monitoring of society" (Osterhammel 2014, 26). As early as in the 1830s and 1840s, statistical societies were founded in many industrial cities in England in order to collect social data for deep insights into the entire social life of the lower classes. One of the most important studies produced in this context was Charles Booth's (1840–1916)

long-term study "Life and Labour of the People of London" (1886–1903), in which he explored the citizens of London over a period of 16 years (Schubert 1994). The scope of these data collections went way beyond legal obligations, contracts or financial dependencies but tried to provide insight into broad facets of everyday life.

This modern empirical view on societies was primarily motivated by the dismay about the great social upheavals occurring during the industrial revolution. At the same time as the industrial revolution, a structural social change began which was characterized by enormous population growth, migration from the countryside to the cities and urbanization – a transformation with massive social consequences. Governments, legal and administrative bodies required reliable data and figures to gain an idea of the magnitude and dimensions of these social shifts. The aim was to use the collection of social data to learn something about the social reality transformed by industrialization and the many new social questions that emerged as a consequence. In this sense, ideas and concepts about the usefulness of data and their collection were framed at first in a philanthropic and idealistic manner. Social data should serve for sociological elucidation and socio-political ambitions, ideas and interests. Furthermore, and as already briefly stressed above, the rise of the modern nation state as a model of political organization is deeply entangled with and enabled by the advent of modern statistics and social sciences. In conjunction with the rise of nations, a new chapter in the social history of datafication begins because data and the consequent needs for calculation are paramount preconditions to making a state administration work. The coordination and efficacy of governing measures depend on the complex integration and interpretation of diverse data, measures, and calculations. Henceforth, the methods, techniques and technologies for logistics, data collection, and processing also needed to progress and new, innovative ways were found. The history of the computer can be told in relation to processes of datafication and the nation state's need to count and calculate (Balbi and Magaudda 2018, 31–33).

In Germany, the Verein für Socialpolitik (founded in 1873) and its members were pioneers in the development of instruments and practices for social investigation by data (Gorges 1986). In the mechanical era of datafication, the Verein für Socialpolitik created important paper tools to gather social knowledge about society. One of the most frequently used tools was the "Enquete" (Embden, Cohn, and Stieda 1877; Horst 1980). In the spirit of the English social reformers, the Verein für Socialpolitik defined the "Enquete" as scientific preparatory work for the enactment of laws: each "Enquete" was expressly intended to have a political impact. To become effective as a "mean of control, agitation and power of the *modern state*," the "Enquete" was a "complex body" of social data collection (Kern 1980, 89). Ideally, an

"Enquete" integrated different practices and techniques of empirical social research such as field studies, questionnaires, special reports by experts, statistics, and surveys. What becomes evident from this example is that the collection of data in the "Enquete" clearly exceeds the documentation of what is a current state of affairs. Instead, they mark a shift from aiming to record society and social relations as they are at the moment of data collection towards the ambition of effectively shaping how society moves forward based on datafied knowledge.

In this manner, the "Enquete" and the method-mix associated with it turned into the standard instrument "of social factfinding in Germany before the first World War": "All the leading historical economists, Schmoller, Bücher, Brentano, Adolph Wagner, as well as those of the next generation who are known to us as the founders of sociology, Tönnies and Max Weber, became involved in planning and directing surveys, writing out questionnaires and analyzing the returns" (Oberschall 1965, 3). The classical philologist and economist Karl Bücher (1847–1930), to take just one example from this list, was not only the father of the "law of mass production" and founder of "newspaper science" as the precursor of communication research in Germany but also a busy empirical social researcher. Prepared by a series of historical-statistical studies, in 1888 and 1889, he conducted a census and a housing-"Enquete" in Basel and with the "Investigations on the Situation of the Craft in Germany" (1892–1897) organized a large-scale study for the Verein für Socialpolitik (Bücher 1890a, 1890b, 1895–1897, 1919). Bücher's studies and the many studies of his colleagues once again reveal the specific view that early empirical social researchers wanted to gain of society with the help of social data. Although the individual researchers and their studies differ in their points of view and objectives, they are united in their political purpose to identify social patterns and regularities in the data in order to use this knowledge to solve serious social problems and imbalances. Thus, it was never just about developing and training a data-based "factual view" (Bonß 1982) on the world, but always also about policy advice, public information, and social reforms. In doing so, early empirical social researchers contributed to the advent of datafication by providing a complex system of reliable instruments, practices and techniques with which modern societies began to discover, observe, and manage themselves with the help of data.

2.1 The Birth of Data Processing

Despite all the ambitions, dedication and efforts of empirical social researchers, collecting social data was only a first step. The huge amounts of data on paper they produced were nothing but worthless data garbage without the right data

processing techniques. In contrast to the tradition of the cameralistic statistics in the Age of Enlightenment, which were primarily interested only in the description of "state peculiarities" ("Staatsmerkwürdigkeiten") (Kern 1980, 19–27), the insights for modern empirical social researchers were hidden in the complex combinations and relations making up the collected data.

Therefore, the statistical methods and procedures had to be refined and efficient calculation aids were required. For his Basel studies, Bücher used a manual method of data processing with a paper tool called "counting slips" ("Zählblättchen"), which was commonly used in European official statistics prior to the punch card (Rauchberg 1890; von Oertzen 2017). "The counting slip," writes von Oertzen in her important study on the innovative data practices of Prussia's statistical office in the second half of the nineteenth century to handle census data, "was an intermediate, movable data carrier designed to facilitate and enhance the counting and sorting of data compiled in lengthy enumeration lists, praised for its ability to greatly enhance statistical complexity. [. . .] The main difference was that the information on counting slips had to be transferred from enumeration lists with a pen, not, as with punch cards, via punched holes. And whereas punch cards could be sorted and counted by machines, counting slips had to be sorted into stacks and counted by hand" (Aronova, von Oertzen, and Sepkowski 2017, 137). The next step was from manual to automated, machine-based data processing. In 1889, Herman Hollerith (1860–1929) was awarded the gold medal at the Paris World Exhibition for a prototype of his Electric Tabulating System. A year later this system was used for the first time in a large-scale experiment for processing data from the American census. Punch cards as data carriers and resources for electro-mechanical data processing, however, only became widely accepted in the following decades (Austrian 1982; Heide 2009). Thereafter, the punch card remained dominant for machine-based data processing until the availability and application of means of magnetic storage possibilities, thus through to at least the 1970s.

Even while the technical processes of data processing by counting slips and punch cards were very different – manual in the former case and machine-based in the latter – both data practices were characterized by common basic techniques for handling large amounts of data. Clearly, the Hollerith machines were faster but "manual and mechanical data processing rested on the same principle: a movable paper tool carrying all relevant data of one person, which enabled statisticians to sort and compile census data in new ways" (Aronova, von Oertzen, and Sepkowski 2017, 149). Data as such do not exist. Data are always complex constructions and so even then a lot of human work was hidden in data to make it usable at all. In this sense, they both revolutionized the organization of data handling by establishing basic data procedures and routines

(e.g., formatting, merging, sorting, synchronizing), which were not yet performed by machines alone, but by many people, who acted as "human computers" (Grier 2007). In these days, computers were not only partly human, as Grier suggests, but oftentimes female: the skillful tasks necessary for serving in the capacity of human computers were often performed by women (Abbate 2012; Edwards and Harris 2017; Hicks 2017). This detail in the history of datafication was later somewhat obscured in computer history, which has strongly reproduced myths of computer development as driven by the great deeds of exceptional men. Whether with or without the support of the Hollerith machine, human-based data processing thus needed "carefully planned choreographics of sorting, sustained labor management, skillful counting techniques, relentless tracing of errors, and strict control" (Aronova, von Oertzen and Sepkowski 2017, 132). This is also indicative of a development characteristic of datafication in the digital era, namely the integration of data collection and data processing in entangled processes.

The modern empirical social research that emerged in the second half of the nineteenth century is only one example of the data enthusiasm prevailing at that time. Especially for the organization of politics and the rationalization of bureaucracy, data became increasingly important. "Seeing like a state," as James Scott has pointedly summarized it, in around 1900 meant seeing the world more and more with the aid of numbers. (Scott 1999). "For governments and state authorities," claims Mau, "the numerical medium is essential in a chaotic reality in order to define problems adequately and devise suitable intervention programmes" (Mau 2019, 32). The numerical approach through data enabled by empirical social research techniques and tools furthermore began to spread. Gradually, almost all areas of social life based their work on data and adopted the "language of numbers" to describe, decipher, and organize social affairs (Mau 2019).

2.2 Data and Social Engineering

Simultaneously, expectations and ideas about the potential and usefulness of data began to shift significantly. The increasing possibilities for measuring social behavior and for surveying societies by data are noticeably intertwined with discourses and concepts of social planning as well as the rationalization and regulation of social life. From this perspective, the many social questions raised by modernity were no longer questions of political action, but rather questions of a more technical nature. The increasing measurement of the social with ever new instruments, machines and methods was a basic prerequisite for this. Social engineering is a generic term for a variety of ideas to adapt social life and societies according to the pace and rhythm of industrial modernity by

means of "social interventions" (Peukert 1987, 132–149; Raphael 2011, 149–157). Thomas Etzemüller points out that the idea of social engineering at that time was nourished by a popular longing for social order, the scientificisation of the social and a general belief in the blessings of technology (Etzemüller 2017). Etzemüller has thus defined social engineering as a "combination of (social) technological solutions, a specific idea of social order and a decided design imperative," which could appear in "various contexts" – from the left to the right of the political spectrum (Etzemüller 2017). Such techno-utopian visions are reflected by what in contemporary datafication discourses is addressed as technological solutionism (Morozov 2013); the oftentimes naïve, and partially ignorant of collateral effects, ideological belief that big data "will allow us to make large-scale and sophisticated interventions in politics, culture, and everyday life. Technology will allow us to solve problems in highly original ways and create new incentives to get more people to do the right thing" (Morozov 2013). However, visions of improving society through social interventions based on data were always connected with utopias of a "new human being": "Coercion through the transformation of things or through authoritarian political intervention and the learning adaptation of those concerned were the two complementary poles of such programs" (Raphael 2011, 151, translation by the authors).

With the increasing measurement of the social and growing technical possibilities, social planning and rationalization became a key political concept far beyond the economic field in the first third of the twentieth century. The 1920s are known for the rationalization of housing, urban planning and social policy from data-driven, functional aspects, but rationalization also intervened more and more in people's individual lifestyles and privacy and covered issues such as controlling "proper" family planning and sexuality in order to optimize the reproduction of the population (Raphael 2011).

3 Times of Dark Data

However, whereas the datafication of human life and social behavior allegedly aimed for the betterment of society and an optimization of social life, there was also a dark side to the datafication of the social world and humanity. Notions of a "new human being" and of the socio-technical control and optimization of the social through data reflect a world view that perceives "the self-dynamics of social change as a jeopardy for the nation" (Raphael 2011, 155, translation by the authors). A means to counter this danger was seen in the biological and social selection of individuals and populations. Since the middle of the nineteenth

century, numerous anthropological doctrines have been established in the grey area between natural and social sciences, which promoted the measurement of man and the social in order to identify the supposedly "deviant and marginalized" (Bernard 2017, 119). Stephen Jay Gould has critically reviewed this "mismeasurement" of man as driven by the idea of biological determinism (Gould 1981). Well-known examples are Alphons Bertillon's (1853–1914) anthropometry, Paul Broca's (1824–1880) craniometry or Cesare Lambombroso's (1835–1909) concept of the "born criminal." These are all concepts that insist on a correlation between the measured external bodily characteristics and mental and moral anomalies and which found widespread support, especially in police practice (Bernard 2017). Classification, categorization and standardization of individual features, social characteristics and normed categories in general play an important role in shaping the modern world, as was stressed by Bowker and Star (2000). Classifications order human interactions and depending on what is classified for what purposes, the classification of standards and deviation can impact on society for the better or the worse.

Even before the First World War, the German historical economist Rudolf Goldscheid (1870–1931) drew up the plan for a comprehensive "economy of human beings" ("Menschenökonomie"), "which would ensure that 'human material' was not wasted, but on the contrary to optimize the population of a country by improving its genetic make-up and rationalizing its reproductive conditions" (Exner 2004; Raphael 2011, 154). In the interwar period, questions of "improving" the biological and social quality of the population increasingly came to the attention of a diverse field of experts, including demographers, eugenicists, medical scientists and social politicians. Their demands, ideas, and models for a planned "healthy regeneration" of human populations based on complex data also reached social policy, which intervened in various countries with political interventions aimed at "negative" (for example marriage prohibition, forced sterilization) and "positive eugenics" (child benefit, laws to protect the health of mother and child etc.) (Raphael 2011).

The times of dark data that loomed on the horizon found their breakthrough in the era of National Socialism, in which the various instruments of biologistic measurement and eugenic selection became the foundations of Nazi racial politics. As Götz Aly and Karl Heinz Roth have extensively researched, experts, practices and techniques of systematic identification, isolation, and selection of people based on the datafication of health, ideology, race, social status, and religion became in fact the administrative backbone of the race-hygienic extermination policy (Aly and Roth 2004). For this purpose, state-of-the-art Hollerith punch card technology was used, which was developed and provided by the German IBM subsidiary DEHOMAG (Black 2002).

4 Data as the Functional Logic of Societies in Digital Transformation

"What may the state, what may the statisticians know about the individual?" Questions such as these may sound anachronistic today, with more data available to states and private enterprises than ever before and with comparatively little critical opposition against the widespread collection and availability of data. But in 1983, Götz Aly and Karl Heinz Roth published their book on the "Nazi Census" in the light of these questions and employed the Nazi State as a deterrent and explicit statement to protest against the German census planned for 1983 (Aly and Roth 2004). The 1980s were a time in which it became obvious that more and more areas of life are organized by data and also the private sphere is increasingly permeated by data. In the German discourse, the fear of the "gläserner Mensch" – a man made of glass, transparent and open for inspection of all the secretive bits and details of their personal lives – was a widespread catchword. From the perspective of historical surveillance studies, Sven Reichhardt (2016) shows how data-based infrastructures, practices and techniques of surveillance, and self-surveillance as essential processes were embedded ever deeper into everyday life, privacy and societies. The fear regarding surveillance through data transparency also echoed in pop culture. The 1981 song "Computerwelt" by pioneering German band Kraftwerk is but one example of the fear of misuse of computer data and digital surveillance by German and international institutions: "Interpol und Deutsche Bank / FBI und Scotland Yard / Flensburg und das BKA / Haben unsere Daten da" ("Interpol and Deutschr Bank, FBI and Scotland Yard, Flensburg (location of the Federal Motor Transport Authority) and BKA (Federal Criminal Police), they all have our data"). The critical reactions to datafication processes reflect a quantitative and qualitative shift and increase in keeping with the prevalence and relevance of data for social processes.

In the second half of the twentieth century, data has become "a fundamental organizational principle of modern societies" (Bächle 2016, 157, translation by the authors). According to Nassehi (2019), in the course of the computerization and digitization of society, data has finally established itself as a functional logic of modern societies. Through these processes of social transformation, data become both a trade secret as well as the leading currency and the raw material of modern societies. In this sense, data for societies has an enormous potential to generate social order, which, according to Nassehi, primarily lies in their inexhaustible "recombination possibilities": "The simplicity of the data is the key to its effectiveness" (Nassehi 2019, 145). In digital societies, data increasingly accumulates in abundance and everywhere.

Whereas in the previous historical examples data were seen as a means of recording social relations and experiences, in this current period data itself moves to the center of interest. In this process, the relationship between the recorded data and social life is partially inverted and data changes from being an ex-post codification of, resource for, and the result of economic operations and legislative measures. We have shown that throughout history, social and economic relationships as well as observable patterns of human behavior were made into data, e.g., translated into contracts or codified into laws; additionally, with social science data gathering there was an interest to be defined and then investigated, data was accumulated to serve a particular goal or to provide insights to specific areas. With the rise of digital platforms and big data, human behavior itself, particularly the use of and interaction with digital devices, generates data. The prospect of datafication in the digital era hence was widened to capture all aspects of life from digital traces and to later find patterns in these enormous volumes of data, which can be used for various interests.

Over time, the concept of datafication has since become more widespread and is used in a variety of fields and (sub)disciplines, also beyond online platforms and the traces left behind by navigating through or with digital devices. The "quantification of the social" renewed historical promises of providing access to a "rationalized world of data" (Mau 2019, 36), which allows for anticipating and controlling human behavior as well as for minimizing risks while doing so. While the character of datafied societies is sometimes rather seen in a widespread "data voluntarism" (Mau 2019) of people, which is cleverly driven by promises of social participation as well as a "cult of numbers that masquerades as rationalization" (Mau 2019), datafication does not only affect those who provide data voluntarily. Instead, big data can also create new and reinforce old divides regarding access to data, interpretation of data, representation in data, and the ethics of data and their processes of capturing (boyd and Crawford 2012). Datafication, in addition to its use as a notion describing a phenomenon and process of social transformation, is hence also increasingly used to name critical research perspectives, discussing potential biases, omissions, or discriminations caused by big data or its analytical application (Leurs and Shepherd 2017; Milan and Treré 2020; Dencik and Kaun 2020). Interestingly, historical references and interpretative schemes prominently feature in such discussions. Employing perspectives from (post-)colonial studies, datafication is seen to create new centers of power and exploited or dependent peripheries, oftentimes mirroring colonial dependencies from the past (Thatcher, O'Sullivan, and Mahmoudi 2016). While colonialism might seem like a thing of the past, as Couldry and Mejias (2019) write, the historic appropriation of land, bodies, and natural resources is mirrored today in this new era of pervasive datafication.

This is just another way of looking at how research into datafication today can learn from historicizing the concept and idea.

Conclusion

Institutions, practices, and processes of datafication, according to the guiding thesis of this chapter, did not simply surface out of nowhere nor come upon us in the wake of digitalization, but have a long prehistory with diverse historical roots. The ambitions and initiatives for understanding the world through data can be traced through time as a persistent and ongoing historical enterprise. A look at earlier epochs can help us to compare and put in perspective processes of datafication today with the data practices of other historical contexts. Historicizing data(fication) does not focus on what is distinct about datafication in the digital age but examines data(fication) with regard to its interrelations with changing social conditions and evolving media environments. Dirk Baecker (2013, 164, 184), hence, proposes seeing "society or culture as metadata," which is characterized by and has enabled "historically varying" processes of datafication. It can then be observed how societies were repeatedly challenged by new forms and quantities of data records in the face of media (r)evolutions (Burke 2000, 2012). Changing media environments affected the generation, storage, and exchange of data. Beer (2016, 1) thus proposes investigating and contextualizing processes of datafication in "historical, political and sociological terms."

The first traces of data-processing instruments and techniques can already be discovered from ancient to early modern times. The main driving forces behind the development of symbolic forms of data gathering and recordkeeping were the necessity of the earliest complex social formations to keep records of their legislative measures and taxes. Apart from that, Nassehi (2019) has pointed out in his studies of the "patterns" of modern societies how closely the nineteenth century as an age of analog data-driven discovery of society by the state, statistics and sociology is linked to the current digital data-driven permeation of our present. Gradually, over the course of the twentieth century, more and more fields of society changed their working basis to data. As a result, data have become a kind of "key currency" not only to quantify but also as a possibility to code and recode almost all facets of the social (Mau 2019). This can already be observed in around 1900 in discourses and ideas of social planning and social engineering to optimize social life and societies and experienced its darkest excesses in the use of data for the race-hygienic extermination policy of the Nazi

regime. Today, data permeates and influences our lives as a matter of course, but as a historical topic, datafication has yet to greatly resonate.

Following a perspective on the social history of data and datafication in the longue durée, we illustrate that the production, collection, and processing of data not only predate digitalization but also, in the immediate decades before the digital revolution, produced exclusive infrastructures, knowledge orders, and practices. Historicizing the concept of data(fication) provides an analytical matrix for identifying persistent questions and changing answers through the ages. In order to capture historically long-term processes of datafication, this research framework can be further systematized by asking not only for the social contexts but also for the respective ideas, discourses, infrastructures, media, practices, and techniques around data in a particular epoch. Using these lenses, we can trace and discuss the shifts and persistence of data collection institutions (e.g., state or private), the governance of access (e.g., open or restricted) and the processing of data, the means and the ends of data collection, and the discursive transparency about (and potential resistance against) socially collected and relevant data and what happens with it.

References

Abbate, Janet. *Recoding Gender: Women's Changing Participation in Computing*. Cambridge, MA: The MIT Press, 2012.

Aeschines. *The Speeches of Aeschines*, translated by Charles Darwin Adams. Cambridge, MA: Harvard University Press, 1958.

Aly, Götz, and Karl Heinz Roth. *The Nazi Census. Identification and Control in the Third Reich*. Philadelphia: Temple University Press, 2004.

Aronova, Elena, von Oetzen, Christine, and David Sepkowski. "Introduction: Historicizing Big Data." *Osiris* 32 (2017): 1–17.

Austrian, Geoffrey D. *Herman Hollerith. Forgotten Giant of Information Processing*. New York: Columbia University Press, 1982.

Bächle, Thomas Christian. *Digitales Wissen, Daten und Überwachung zur Einführung*. Hamburg: Junius, 2016.

Baecker, Dirk. "Metadaten. Eine Annäherung an Big Data." In *Big Data. Das neue Versprechen der Allwissenheit*, edited by Heinrich Geiselberger and Tobias Moorstedt, 156–186. Berlin: Suhrkamp, 2013.

Balbi, Gabriele, and Paolo Magaudda. *A History of Digital Media: An Intermedia and Global Perspective*. New York: Routledge, 2018.

Barber, Malcolm. *The Two Cities: Medieval Europe, 1050–1320*. New York: Routledge, 1992.

Beer, David. "How should we do the history of Big Data?" *Big Data & Society* 3 (2016): 1–10.

Bernard, Andreas. *Komplizen des Erkennungsdienstes. Das Selbst in der digitalen Kultur*. Frankfurt am Main: S. Fischer, 2017.

Black, Edwin. *IBM and the Holocaust. The Strategic Alliance between Nazi Germany and America's most powerful Corporation*. New York: Three Rivers Press, 2002.
Bonß, Wolfgang. *Die Einübung des Tatsachenblicks. Zur Struktur und Veränderung empirischer Sozialforschung*. Frankfurt am Main: Suhrkamp, 1982.
Borck, Cornelius. "Big Data. Praktiken und Theorien der Datenverarbeitung im historischen Querschnitt." *N.T.M.* 25 (2017): 399–405.
Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: The MIT Press, 2016.
Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things out: Classification and Its Consequences*. Cambridge, MA: The MIT Press, 2000.
boyd, danah, and Kate Crawford. "Critical Questions For Big Data." *Information, Communication & Society* 15, no. 5 (2012): 662–679.
Brückweh, Kerstin, Schumann, Dirk, Wetzell, Richard F., and Benjamin Ziemann, eds. *Engineering Society. The Role of the Human and Social Sciences in Modern Societies 1880–1980*. Basingstoke: Palgrave Macmillan, 2012.
Bücher, Karl. *Die Bevölkerung des Kantons Basel-Stadt am 1. Oktober 1888*. Basel: H. Georg, 1890a.
Bücher, Karl. *Die Wohnungs-Enquete in der Stadt Basel vom 1.-19. Februar 1889*. Basel: H. Georg, 1890b.
Bücher, Karl, ed. *Untersuchungen über die Lage des Handwerks in Deutschland mit besonderer Rücksicht auf seine Konkurrenzfähigkeit gegenüber der Großindustrie*. Leipzig: Duncker & Humblot, 1895–1897.
Bücher, Karl. *Lebenserinnerungen 1847–1890*. Tübingen: H. Laupp, 1919.
Burke, Peter. *A Social History of Knowledge: From Gutenberg to Diderot*. Cambridge: Polity Press, 2000.
Burke, Peter. *A Social History of Knowledge II: From the Encyclopedia to Wikipedia*. Cambridge: Polity Press, 2012.
Couldry, Nick, and Andreas Hepp. *The Mediated Construction of Reality*. Cambridge: Polity Press, 2017.
Couldry, Nick, and Ulises Ali Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Culture and Economic Life. Stanford: Stanford University Press, 2019.
Culham, Phyllis. "Archives and Alternatives in Republican Rome." *Classical Philology* 84, no. 2 (1989): 100–115.
Dencik, Lina, and Anne Kaun. "Datafication and the Welfare State." *Global Perspectives* 1, no. 1 (2020): 12912.
Edwards, Sue Bradford, and Duchess Harris. *Hidden Human Computers: The Black Women of NASA*. Minneapolis: Essential Library, an imprint of Abdo Publishing, 2017.
Embden, Gustav, Cohn, Gustav, and Wilhelm Stieda. *Das Verfahren bei Enqueten über sociale Verhältnisse. Drei Gutachten nebst einem Anhang nach dem Englischen von J.M. Ludlow*. Leipzig: Duncker & Humblot, 1877.
Emigh, Rebecca Jean, Riley, Dylan, and Patricia Ahmed. *Antecedents of Censuses from Medieval to Nation States: How Societies and States Count*. London: Palgrave Macmillan, 2015.
Es, Karin van, and Eef Masson. "Introduction: Big Data Histories." *Journal for Media History/ Tijdschrift for Mediageschiedenis* 21 (2018): 1–6.

Etzemüller, Thomas, ed. 2009. *Die Ordnung der Moderne: Social Engineering im 20. Jahrhundert*. Bielefeld: Transcript, 2009.
Etzemüller, Thomas. "Social engineering." *Docupedia-Zeitgeschichte*. 2017. http://docupedia.de/zg/Etzemueller_social_engineering_v2_de_2017.
Exner, Gudrun. "Rudolf Goldscheid (1870–1931) and the Economy of Human Beings. A New Point of View on the Decline of Fertility in the Time of the First Demographic Transition." *Vienna Yearbook of Population Research* 2 (2004): 283–301.
Giedion, Sigfried. *Mechanization Takes Command. A Contribution to Anonymous History*. New York: Oxford University Press, 1948.
Gorges, Irmela. *Sozialforschung in Deutschland 1872–1914. Gesellschaftliche Einflüsse auf Themen- und Methodenwahl des Vereins für Socialpolitik*. Frankfurt/Main: Anton Hain, 1986.
Gould, Stephen Jay. *The Mismeasure of Man*. New York: W.W. Norton, 1981.
Grier, David Alan. *When Computers were Human*. New Jersey: Princeton University Press, 2007.
Heide, Lars. *Punched Card Systems and the Early Information Explosion 1880–1945*. Baltimore: The Johns Hopkins University Press, 2009.
Herrenschmidt, Clarisse. *Les trois écritures. Langue, nombre, code*. Paris: Gallimard, 2007.
Hicks, Mar. *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*. History of Computing. Cambridge, MA - London: The MIT Press, 2017.
Kern, Horst. *Empirische Sozialforschung. Ursprünge, Ansätze Entwicklungslinien*. München: C.H. Beck, 1980.
Krotz, Friedrich. "*Zeit* der Mediatisierung – Mediatisierung der *Zeit*." In *Medien & Zeit*, no. 2 (2012): 25–34.
Law, Danny, Haicheng, Wang, Nissen, Hans J., and Gary Urton. "Writing and Record-Keeping in Early Cities." In *The Cambridge World History*, vol. 3, edited by Norman Yoffee, 207–25. Cambridge: Cambridge University Press, 2015.
Leurs, Koen, and Tamara Shepherd. "15. Datafication & Discrimination." In *The Datafied Society*, edited by Mirko Tobias Schäfer and Karin van Es, 211–232. Amsterdam: Amsterdam University Press, 2017.
Mau, Steffen. *The Metric Society. On the Quantification of the Social*. Translated by Sharon Howe. Cambridge: Polity Press, 2019.
Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data. A Revolution That Will Transform How We Live, Work and Think*. London: John Murray, 2013.
Mejias, Ulises A., and Nick Couldry. "Datafication." *Internet Policy Review* 8, no. 4 (2019).
Milan, Stefania, and Emiliano Treré. "The Rise of the Data Poor: The COVID-19 Pandemic Seen From the Margins." *Social Media + Society* 6, no. 3 (2020): 1–5.
Morozov, Evgeny. *To Save Everything, Click Here: The Folly of Technological Solutionism*. First edition. New York: PublicAffairs, 2013.
Nachum, Iris. "Heinrich Rauchberg (1860–1938). A Reappraisal of a Central European Demographer's Life and Work." *Austrian History Yearbook* 50 (2019): 78–98.
Nassehi, Armin. *Muster. Theorie der digitalen Gesellschaft*. München: C.H. Beck, 2019.
Oberschall, Anthony. *Empirical Social Research in Germany 1848–1914*. The Hague - Paris: Mouton & Co., 1965.
Oertzen, Christine von. "Machineries of Data Power. Manual versus Mechanical Census Compilation in Nineteenth-Century Europe," *Osiris* 32 (2017): 129–150.
Ong, Walter J. *Orality and literacy: The technologizing of the word*. New York: Methuen, 1982.

Osterhammel, Jürgen. *The Transformation of the World. A Global History of the Nineteenth Century*. Translated by Patrick Camiller. Princeton - Oxford: Princeton University Press, 2014.

Peukert, Detlev J.K. *Die Weimarer Republik. Krisenjahre der Klassischen Moderne*. Frankfurt/Main: Suhrkamp, 1987.

Raphael, Lutz. "Die Verwissenschaftlichung des Sozialen als methodische und konzeptionelle Herausforderung für eine Sozialgeschichte des 20. Jahrhunderts." *Geschichte und Gesellschaft* 22 (1996): 165–193.

Raphael, Lutz. *Imperiale Gewalt und mobilisierte Nation. Europa 1914–1945*. München: C.H. Beck, 2011.

Rauchberg, Heinrich. "Übersicht über den Stand und die neuesten Fortschritte der Technik auf dem Gebiete der Bevölkerungsstatistik." *Allgemeines Statistisches Archiv* 1 (1890): 99–116.

Rauchberg, Heinrich. "Die elektrische Zählmaschine und ihre Anwendung, insbesondere bei der österreichischen Volkszählung." *Allgemeines Statistisches Archiv* 2 (1891–1892): 78–126.

Reichhardt, Sven. "Überwachungsgeschichte(n). Facetten eines Forschungsfeldes." *Geschichte und Gesellschaft* 42 (2016): 5–33.

Rödder, Andreas. *21.0. Eine kurze Geschichte der Gegenwart*. München: C.H. Beck, 2015.

Schubert, Dirk. "Charles Booth – Entdecker der 'Zwei-Drittel-Gesellschaft' und die 'Arithmetik des Jammers'." *Jahrbuch für Soziologiegeschichte* (1994): 117–140.

Scott, James. *Seeing Like a State. How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press, 1999.

Sickinger, James P. *Public Records and Archives in Classical Athens*. Chapel Hill: The University of North Carolina Press, 1999.

Soll, Jacob. *The Information Master: Jean-Baptiste Colbert's Secret State Intelligence System*. Ann Arbor: University of Michigan Press, 2009.

Southerton, Clare. "Datafication." In *Encyclopedia of Big Data*, edited by Laurie A. Schintler and Connie L. McNeely, 1–4. Cham: Springer International Publishing, 2020.

Thatcher, Jim, O'Sullivan, David, and Dillon Mahmoudi. "Data Colonialism through Accumulation by Dispossession: New Metaphors for Daily Data." *Environment and Planning D: Society and Space* 34, no. 6 (2016): 990–1006.

Thorvaldsen, Gunnar. *Censuses and Census Takers: A Global History*. New York: Routledge, 2018.

Van de Mieroop, Marc. *Cuneiform Texts and the Writing of History*. London: Routledge, 1999.

Van Dijck, José. "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology." *Surveillance & Society* 12, no. 2 (2014): 197–208.

Weber, Max. "Science as Vocation." In *Max Weber: The Vocation Lectures*, edited by David Owen and Tracy B. Strong, translated by Rodney Livingstone, 1–31. Indianapolis: Hackett Publishing Company, 2004.