

Big data, small freedom?

Informational surveillance and the political

Burkhardt Wolf

In 2010, 'big data' was described as 'datasets that could not be captured, managed and processed by general computers within an acceptable scope'.¹ Today's definitions boil down to three Vs: Variety, Volume and Velocity. Big data deals with mostly unstructured, heterogeneous and non-validated data, whose size is so big that it requires parallel processing in supercomputers. Only this setup can cope with huge amounts of data that are constantly changing 'through the absorption of complementary data collections, through the introduction of previously archived data or legacy collections, and from streamed data arriving from multiple sources'.² Big data may encompass structured databases of every origin, but also transaction and interaction data from communication networks, data from cloud computing, and the rapidly growing 'internet of things' – from smart devices to sensors, and cameras.

Big data does not pivot merely on disposing of huge stacks of information, but rather on the possibility to aggregate or analyse them, and therefore to find unexpected cross-connections within the material. When taken out of – as some enthusiasts once claimed – the fluid, open and free environment of electromagnetic streams, the data is, so to speak, petrified in order to 'mine' it. Data mining, in turn, is supposed 'to uncover previously unknown, useful and valuable knowledge, patterns, relations' through the use of 'sophisticated evolutionary algorithms of classical techniques such as statistics, pattern recognition, artificial intelligence, machine learning'.³ And since, according to Moore's law, sensory and computational capacities are growing exponentially, this procedure is run with less and less regard to time and quantity.

Big 'dataveillance'?

Big data will get bigger and bigger. More and more quickly, it will disclose complex structures within, at first glance, unstructured data masses. Within standard procedures, those findings can be converted into descriptions and explanations of our past and present, or into predictions and probability-based statements about our possible futures. This opens a delightful prospect for technocrats: quite a few medical advisers, for example, are jolly about the promise of totally individualized medical treatment deemed feasible by processing the genomic big data of any patient. A good many advocates of e-learning are excited about the possibilities of digital education, whereby any learning success or failure is fed back into the personal data record. And on the level of social technologies, big data lends new capacities to early-warning systems that rely on streaming data, in order to analyse it in real time by adaptive, sequential and learning algorithms. It thus supports police consultants who investigate the probability of crime; it assists anti-terrorism units tirelessly watching for threats to the public; and it backs up counter-insurgency analysts who are alerted by the chances of 'wrong' political events.

Critics of big data are not only puzzled by the easily legalized informational exploitation for – at first sight purely – commercial reasons. They are troubled about the

threat to 'our freedom', since the almost unlimited data processing obviously harms our right to 'informational self-determination' (as it is called in the German constitution): the right to know and control what others know about us. And it is perhaps not so much the existence and scope, but rather the fabrication of this 'knowledge' that appears to be the corporate secret of today's 'control society'. Even though big data does not, in the first place, concern individuals, but specifically encompasses 'metadata' (information about the structures of certain data, but not about their content), people may swiftly be identified and then classified as soon as metadata are combined. This may lead to a triple secrecy of social sorting: using allegedly confidential data, keeping the persons in question unaware of it, and applying 'proprietary algorithms' for producing statements about those persons.

Highly individualized medical treatment, for instance, whereby the body is 'defined in terms of its genetic profile, nicotine or medication intake, disease history, etc.'⁴ may easily determine personal risk assessment with severe consequences for finding appropriate health insurance or for future opportunities in the job market. Learning analytics may lead to a complete survey of the learner (especially when short-circuited with data about his 'general interests' and general behaviour), eventually denying him his freedom of future development and replacing traditional 'education' with personal surveillance and deterministic profiling. Or some highly speculative makeshift security alert based on big data correlations could lead to the long-term observation of any person, or even to immediate preventative measures against unsuspecting suspects. So are we at a turning point? Does 'dataveillance' create new state secrets, new *arcana* that, for the sake of governance, have to be concealed from public view? Or is this sort of 'statistical rule' nothing new at all?

History of statistical rule

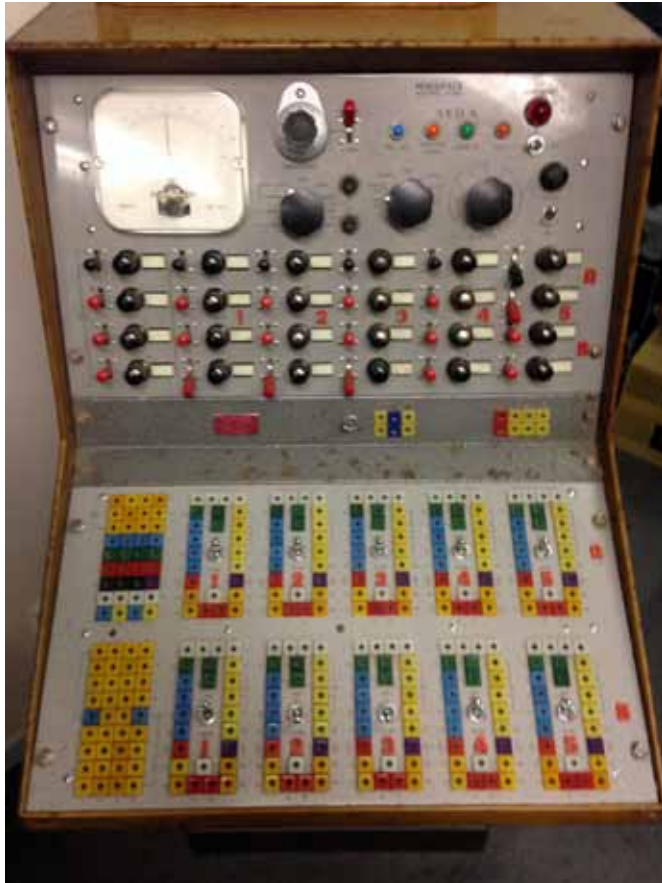
We have to look to early modern times in order to find the first institutionalized dataveillance: in sixteenth-century Europe, *police* was understood as a statist measure for 'good order' in public affairs, for the productivity and well-being of the people. However, this police (designed for operating beyond the supervisory authority of traditional law and administration) was not yet focused on defensive action or protection against dangers and threats. Its task was to orient and to inspect the people, to which end it made use of early queries and registers. But it operated on a local, mostly urban, scale.

The first attempt at 'statistical' rule was made in Germany in the seventeenth century, when the state, its territory, its inhabitants, its economy and its institutions were described in an idiographic manner. 'Statistics' here was defined as the 'science of the actual state of the state' (*Staatszustandswissenschaft*), a science that should collect and arrange diverse data as an expedient for the ruler and his decisions. But these massive amounts of data were kept secret. They were *arcana* of state, made use of exclusively by the sovereign. The main concern of this whole enterprise was, in a quite characteristically German manner, to grasp the territorial diversity of the state as a totality within the heterogeneous. Therefore, these statistics were not based on proper analytic categories, let alone on computations.

Around the same time, first attempts at numerical statistics were made in England, where, due to local liberal traditions, no statistics in the sense of 'state data' were available. But the first surveys did have birth records and 'bills of mortality' kept by the parishes to fall back on. The data found there was not merely for the record and for a quantitative description of England's state of affairs, but was investigated for regularities in its development. In this way, certain laws of fertility and mortality were concluded, and the population as a whole was assessed. John Graunt, William Petty and Edmond Halley, the pioneers of 'political arithmetic', as this approach was named,

set an example for future numerical surveys of state affairs and introduced random sampling as a statistical method.

A real calculus for statistics was advanced neither in Germany nor in England, but in France. There, the administration was centralized and professionalized for fiscal reasons, a general record of the kingdom's resources was part of the royal *arcana*, and the new mathematical sciences were promoted for governmental reasons. While German statisticians were confined to words and simple numbers, and English researchers used merely a kind of 'shopkeeper's arithmetic', scientists in seventeenth-century France (e.g. Blaise Pascal) formulated the first doctrines of chance while devoting themselves to theories of gambling. François Quesnay's *Tableau économique* (1758) finally brought together the prospect of totality (as claimed in Germany) with the promise of precise measurements (as advocated in England).



During the Napoleonic era, extensive and methodical *enquêtes* of the whole country were commissioned, statistical bureaus established, and exhaustive statistics published. The calculus of probability was refined especially by Pierre Simon Laplace, a distinguished scholar as well as one of Napoleon's ministers. In Laplace's view, probability compensated for our confined subjective knowledge, but as such it was no proof against the total computability of the world: it could only be totally understood and predicted if there was an intelligence, a spirit or a ghost overseeing all data and laws. For Laplace, this absolute intelligence was only a hypothesis and did not necessarily have to be called God. But this hypothesis was certainly no less justified than today's conceptions of all-calculating supercomputers.

The later nineteenth century was crucial for statistical rule, since it allowed for an exchange between technologies of administration and the knowledge of natural sciences, as well as of the humanities (or 'moral sciences'). Two final steps had to be taken to achieve the goal of ruling society via probabilities. First, Jakob Bernoulli's paradigm of mass statistics, the 'main theorem', developed most prominently in his *Ars Conjectandi* (1713), was reinterpreted. Bernoulli's analogy to society concerned a ballot box containing black and white balls in a specific, but unknown, proportion. The more often you draw a ball, the closer you get to the actual proportion, he concluded. In the nineteenth century, Siméon Poisson, a disciple of Laplace, confirmed this convergence even when the investigated elements were completely heterogeneous. From this, the statistical 'law of large numbers' was derived as a kind of universal law for nature and society. Second, nineteenth-century astronomy had to deal with disturbing errors in measurement, as it tried to determine the constant data of celestial movements. Thus, the concept of statistical variance around the true value due to measurement errors was introduced and projected onto a mathematical graph. There, a bell-shaped curve, the so-called 'Gaussian distribution', was visualized, presenting the 'normal value' and a section of errors or – as one could interpret it – possibilities around it.

The Belgian astronomer Adolphe Quételet merged these two innovations: he applied the distribution of normality and its deviances to mass phenomena like the members of a society administered under the 'law of large numbers'. In such a way, he 'discovered' the *homme moyen* (*the mean or average man*) who was marked by the statistical median and drew an analogy to the true value of astronomical observation. In relation to this *homme moyen*, which he regarded as a *real* norm and, at the same time, as an *ideal* of homeostasis within society, empirical people were, so to speak, errors not in measurement but in embodiment. Thus, two crucial epistemological steps were taken on the track to statistical rule: first, making several (confined) observations of one and the same object; and second, building up a new reality by observing different objects and establishing a relationship between them. In a way, this referred back to the old scholastic problem of the existence of a 'general entity'.

In any case, in order to invent the concept of 'society', statistical objects and correlations had to be reified as 'collective things'. This realistic notion of virtual macrosocial objects led to two important inception: sociology was founded as a new science that focuses solely on this – half real, half imaginary – object named 'society'. And, as a showcase of statistical rule, public insurance was founded on a large scale, especially in Germany between 1881 and 1889, when the state introduced obligatory health, accident and old-age insurance on the basis of extensive statistical data. This 'political technology', as François Ewald has called it, established a totally new, non-legal and immanent form of social contract: now risk and, therefore, statistics-based probability was considered a collective issue, was made calculable, and was regarded as capital to invest.⁵ Against this background, one could ask whether statistics, as a technical way to rule people and to achieve consensus, is a replacement of politics – especially after Bismarck had famously established public insurance to get rid of the socialists.

Implementation of dataveillance

Yet the outlined 'ideological' and 'scientific' advancement of statistical rule is only half the story. To be implemented, it had to fall back on certain media: to tables, charts, schedules and, of course, early computing devices, which had already been developed in the eighteenth century. To be sure, real computer databases did not emerge until the late nineteenth century, when Herman Hollerith invented his programmable and electromagnetic tabulating machine. This 'statistical computer', as he called it, automatized the counting and sorting of punched cards.⁶ After having been deployed for the American census and imported to Europe, it turned out to be extremely useful for insurance companies and for official statistics, so that it was still being used by the Nazis, among other things for finding and excluding the so-called 'abnormal'. In contrast to big data, its data pool was highly structured. But it set a paradigm for the statistical use of calculating machines, which was pushed forward especially during the Cold War with its 'operational research' and its political programme of 'cybernetic societies'.

As nineteenth-century rule had demonstrated, controlling and prevention traditionally operate on two different levels: personal files and statistical databases, which are then interpreted in order to find sectors of normality and deviance. But does data-veillance continue to operate on those two levels in the present? In 'positive dragnet investigation', which was used in the form of *Rasterfahndung* in Germany's war on 'left-wing-terrorism' in the 1970s, personal files of suspected or searched individuals were compared to statistical databases. In 'negative dragnet investigation', on the other hand, police started with a mere pattern of possible offenders (who, notably, could be characterized by feigned normality), so that computers would track down suspicious persons by scanning diverse statistical databases. Searchers thus were empowered to stigmatize certain patterns as 'criminogenic', regardless of actual evidence. Today's

dataveillance is, of course, much more extensive and effective in its searching functions and crime pattern analyses.

In a genuine big-data approach – along with structured data sets like documents, mail or telephone calls – motion patterns, voice analyses and camera shots are also used. Here, neither the content of the documents nor the intention of the person is crucial: only a little metadata and its correlations are enough to identify a person;⁷ and as soon as its pool is enlarged, metadata is sufficient to produce a profile prediction. From the perspective of big data management, people are the sum of their interactions, contacts and relationships. Because of that, a personal file is transformed from traditional parameters like name, age, gender or residence into a prognosis for the personal future. And on these grounds, dataveillance draws from information potentially important in days to come. Consequently, those ‘data predictions about individuals may be used to, in effect, punish people for their propensities, not their actions. This denies free will and erodes human dignity.’⁸ In aiming particularly at predictions, surveillance increasingly abstains from substantial explanation and from proper political action – as if ‘society’ were a natural, immutable phenomenon, and not man-made. In the end, by this kind of dataveillance, power knows us, our relations and correlations, our present, past and future, much better than we ourselves do.

Fictions of freedom and the political

Power seems all-knowing and all-seeing, even if its knowledge and its observations are grounded in the grey zone between reality and fiction. Varying Michel Foucault’s famous term, ‘synopticism’ has come to be understood as an enhanced form of panopticism, placing the individual on a statistical level, allowing for limited freedom and administering the ‘market of risks’, instead of searching for and dealing with the real causes. If this exercise of power governs individuals and collectives in their mutuality, if it refrains from any self-contained reason of state, and if it merely tries to regulate society according to its own ‘inherent’ or ‘natural’ laws – then you could name this kind of rule, once again in Foucault’s terms, ‘governmentality’. It is not any more a counterbalance to individual or social freedom, but rather a stimulation and protection of liberal freedom – that is, of market freedom. The state has become a centre for limited intervention and a network of security *dispositifs* ensuring man’s freedom, so long as he makes use of it in a productive way. Freedom and security are two sides of the same coin. And even if we should say that ‘normalization’ is a limit to our freedom, one could answer that ‘normalism’ (Jürgen Link) itself has been loosened and made ever more flexible: nineteenth-century normalism, as described by Foucault, was based on collecting and interpreting data, by creating a zone of normality, by defining zones of harmless deviations and by excluding specific zones. In ‘control society’, as described by Gilles Deleuze, the hypothesis and ideal of an average man and the corresponding concept of deviation seem as obsolete as the method of proper statistics; nothing is excluded, but everything is constantly put in motion, modulated and permanently related to permanently changing data pools.

The old nightmare scenario of Orwellian totalitarianism seems to have been superseded by liberal and ‘liquid’ surveillance, as Zygmunt Bauman has termed it. However, because the new flexible normalism is not interested in real causalities or inner reasons, individuals become black boxes: they are not to be interpreted as subjects any more, but have to be assessed according to their outputs and, by this, correlated to data stocks. This is why Deleuze speaks of contemporary ‘dividuals’. Put under the rule of probability, power is constantly confronted with pure chance, but uses big data and algorithms to make chance real. Whereas probability traditionally concerns itself with the future, future possibilities are now projected onto things and become of prime importance for their qualification. Therefore reality is duplicated: power and

business are geared to the fiction of a probable reality – to a ‘fiction’ that is much more than pure fantasy.⁹ But as soon as probability and its fictitious reality become more crucial than certainty and reality, data processing becomes the main task. And as soon as proactive or preventative measures are more important than coping with present conditions or actual causalities, fictions, as David Lyons has put it, ‘make a real difference. They have ethics, politics.’¹⁰ Against this background, it may be not freedom but rather the political that is at stake.

Reinserting political knowledge

To be sure, depoliticization through statistical rule has been a topic ever since the introduction of data in government. Perhaps most succinctly, Carl Schmitt countered ‘the political’ with the ‘fanciful’ or ‘fictitious’ stance of modern government, with its ‘statistical apparatus’ and its negation of real causes.¹¹ For Schmitt, liberal ‘politics’ is characterized by its businesslike ‘good governance’ and its production of general consensus that is mostly based on data and the deliberation of experts. But how could one place ‘the political’ in opposition to contemporary dataveillance without falling prey to Schmitt’s totalitarian drift or without sidestepping into discussion of essentially abstract philosophical concepts, as elaborated for instance by Chantal Mouffe, Claude Lefort or Alain Badiou? How is the ‘political’ beyond the ‘control society’ to be practically set off? Is it an issue beyond discourse, the mere disturbance and disruption of digital power?

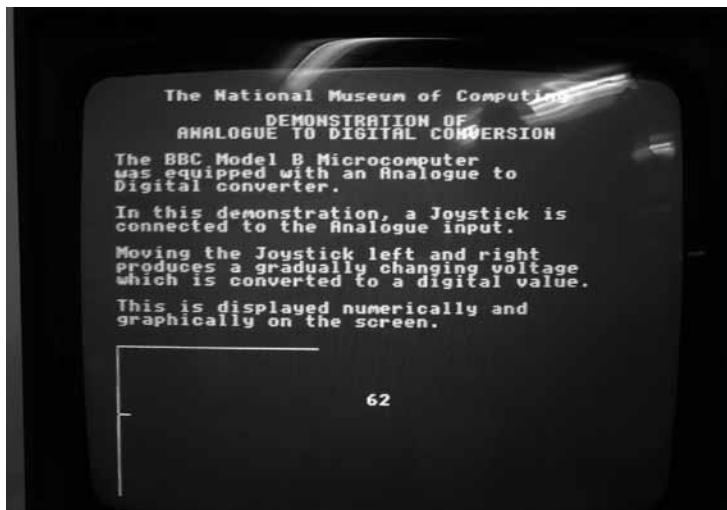
Deleuze advocated something like this, and some activists try to realize the ‘political’ through active ‘blackboxing’, by denying the supply of data, and by becoming absent, ‘dividual’ and practical non-being. But, given that statistical rule will proceed as long as its disturbance remains a mere disturbance, could there be a kind of third way in aiming at the interface between discourse and data processing? The vigilance of cultural critics is a general and traditional way; the systematic indiscretion of whistleblowers a more recent and spectacular way of making data-veillance again accessible for political discourse. If part and parcel of ‘the political’ is decision-making or, at least, contesting decisions, then one should certainly fight not only over the measures to be taken, after data has already been construed, but over data interpretation and the mining of data itself. The algorithms and codes are, aside from the hardware and infrastructure themselves, the deepest level of the political. ‘Code is law’, as Lawrence Lessig put it in 2000, because it shapes our society as a marketplace.¹²

On the level of ‘mining’, interpreters of big data far too often pretend to deal in a secure way with insecure things or persons. Regardless of the fact that they are dealing with insecure knowledge, under the market pressure of efficiency and velocity, some data experts are interpreting data as swiftly as it is being processed: they make a quick search for correlations and speedily draw their conclusions. But the value, significance and meaning of correlations has been disputed since the very beginning of statistical rule: in the seventeenth century, John Graunt dwelled on the superstitious correlation between coronations and plagues;¹³ the medical statisticians of the eighteenth century were compelled to distinguish between mere correlations and proper, medically reliable causalities; and the dispute about Quételet’s *average man* led to discussions about fake or real, constant or merely probable causes and their link to underlying concepts like ‘nature’ and ‘society’.¹⁴

Through correlations without fair scientific expertise, one can render nearly anything plausible. Francis Galton, for example, was not only the creator of modern ‘eugenics’, but also the founder of ‘correlation analysis’. In contrast to big data’s analysts, he started from a fuzzy hypothesis: namely, the strict correlation between intelligence or ‘social value’ and physical characteristics. But, already in the

'small-data-age', he demonstrated how to build 'objectively' upon correlations between heterogeneous and, at first sight *and* final inspection, not substantially linked data. In the age of big data, as 'datasets are far too big and the area under consideration is probably far too complex', a 'valid substantive hypothesis' seems to be negligible. Proponents of this approach claim: 'Our results may be less biased and more accurate, and we will almost certainly get them much faster'¹⁵ – as if social and political issues were generally impartial, calculable and dependent on the quickest possible solution.

Even though big data analyses may disclose more than linear relationships and detect more complex, thus far ignored correlations, it's still doubtful whether big-data analysts, urged towards high-speed delivery, will find the time and have the subject-specific ability to thoroughly consider their conditions and consequences. Providing 'answers to questions that previously could not even be posed' makes the big-data approach comparable to modern scientific experiments¹⁶ – but no one would contend that experimentation could take place without meticulous subject-specific knowledge. In and of themselves, patterns or correlations mean nothing and do not refer to any reality, unless they are integrated into a language or grammar of evaluations and knowledge. But the 'end of theory' that was proclaimed (most prominently by Chris Anderson in 2008) for the new age of big data has dawned by choice, if data mines are merely exploited, but not really investigated. Data without theory is blind, as theories without data are empty. Thus, on the level of contemporary data power, the political consists in reconnecting code and discourse.



This, to be sure, is easily said – but not easily accomplished without the data experts that actually have the material at their disposal. Big-data mining, especially when touching on political issues like surveillance and prevention, tends to consist of mere observations plus fuzzy inferences if there is no political and scientific expertise involved. Usually (as is the case with Kalev H. Leetaru and his 'Global Database of Events, Language, and Tone'), pioneering big-data

projects concerning political issues are funded commercially and put into their sponsors' service (in this particular instance: Yahoo and Google), but not disclosed to the public or to science. And the problem is not only that many of 'the statistical packages or algorithms that are being developed in a highly competitive marketplace have been specialized for the generation of predictive models and simulations'.¹⁷ Within the big-data industry, the analyst often stands alone when dealing with his or her data – whereas in the age of small data work on statistical results was 'performed by professionals other than the mathematicians, computer scientists, and statistic experts who analysed the data'.¹⁸ Control replaces comprehension, if data-driven approaches completely supersede hypothesis-driven ones, whose proxies, of course, are always insecure and tentative – but, at the same time, debatable on a scientific, theoretical and political level. Enthusiasm about technical doability plus some humanist qualms (regarding 'free will' or 'human dignity') are not good enough. What seems to be indispensable is to set knowledge against the exercise of data-driven power by reclaiming data interpretation and creating new political concepts. If there is no subject-specific debate about merely correlated data, 'society' becomes a black box and is surrendered to purely technocratic reasoning.

The most obvious problem of ‘big dataveillance’ seems to be then, first, one-sided expertise: mostly, data analysts unite the skills of statistics, programming, design, communication and sale, but they are not scientists or specialists on politically sensitive data. Whereas in the old days of statistical rule most data experts were at least concerned about the consequences of their doing for the shape and range of politics and the political. Second, the easy availability of big data leads to numerous unsound data interpretations; this wouldn’t necessarily be a problem, unless, third, many politically relevant data investigations were outsourced to commercial agencies, left to the free market and its interests. Against this background, the old alliance between state and university seems somewhat nostalgic, especially with regard to the new type of scholar who is no longer firmly related, but maybe just ‘correlated’, to a neoliberalized university dependent on commercial sponsors or other external funds.

In times of dataveillance, even the old *arcana* of government that concerned an exclusive knowledge of sovereigns and scientists seem something to be yearned for: state archives and statistics served as decision guidances, and, although they were brought under reason of state and were time and again treated as secrets, they were firmly tied to political discourse and left to political experts. Even if today’s *arcana*, big-data codes and algorithms, are not employed off the record, they have become business secrets. And a statistical rule that is brought under economic dictates dissolves not only ‘the political’, but politics as well. Apart from providing an indispensable counteraction against the extensive scope of dataveillance (executed by states and their police or secret services) that threatens ‘our freedom’, genuinely political action consists in protecting scientific, theoretical and historical knowledge about the political. The task is to reinsert theory and science into the exercise of statistical politics – and not to sell them as some small things on the market of big data.

Notes

1. Min Chen, Shiwen Mao, Yin Zhang and Victor C.M. Leung, ‘Introduction’, *Big Data: Related Technologies, Challenges and Future Prospects*, Springer, Cham/Heidelberg/New York/Dordrecht/London, 2014, p. 2.
2. Jules J. Berman, ‘Introduction’, *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*, Morgan Kaufman, Amsterdam and Boston MA, 2013, p. xx.
3. Ali Serhan Koyuncugil and Nermin Ozgulbas, ‘Preface’, *Surveillance Technologies and Early Warning Systems: Data Mining Applications for Risk Detection*, Hershey, New York, 2011, p. xv.
4. Irma van der Ploeg, ‘The Body as Data in the Age of Information’, in Kirstie Ball, Kevin D. Haggerty and David Lyon, eds, *Routledge Handbook of Surveillance Studies*, Routledge, London and New York, 2012, S. 177.
5. François Ewald, ‘Insurance and Risk’, in Graham Burchell, Colin Gordon and Peter Miller, eds, *The Foucault Effect: Studies in Governmentality*, Harvester Wheatsheaf, London, 1991, p. 207.
6. Cf. Geoffrey D. Austrian, *Herman Hollerith: A Forgotten Giant of Information Processing*, Columbia University Press, New York, 1982, pp. 40–41, 62–5.
7. Cf. Klaus Mainzer, *Die Berechnung der Welt. Von der Weltformel zu Big Data*, Beck, Munich, 2014, pp. 246–7.
8. Kenneth Cukier and Viktor Mayer-Schönberger, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, Boston MA and New York, 2013, p. 170.
9. Cf. Elena Esposito, *Die Fiktion der wahrscheinlichen Realität*, Suhrkamp, Frankfurt am Main 2007, p. 120.
10. David Lyon, ‘Surveillance as Social Sorting. Computer Codes and Mobile Bodies’, in David Lyons, ed., *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination*, Routledge, London and New York, 2003, p. 27.
11. Carl Schmitt, *Political Romanticism* (1919), trans. Guy Oakes, MIT Press, Cambridge MA and London, 1986, p. 30; and *The Crisis of Parliamentary Democracy* (1923), trans. Ellen Kennedy, MIT Press, Cambridge MA and London, 2000, p. 16.
12. Lawrence Lessig, ‘Code is Law: On Liberty in Cyberspace’, in *Harvard Magazine*, January/February 2000.
13. Cf. John Graunt, *Natural and Political Observations...*, in *The Economic Writings of Sir William Petty*, vol. II., Cambridge University Press, Cambridge, 1899, p. 369.
14. Cf. Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning*, trans. Camille Naish, Harvard University Press, Cambridge MA, 1998, pp. 103–4.
15. Cukier and Mayer-Schönberger, *Big Data*, pp. 54–5, and hereafter, p. 61.
16. Hans-Jörg Rheinberger, *The Epistemology of the Concrete: Twentieth-Century Histories of Life*, Duke University Press, Durham NC, 2010, p. 171.
17. Oscar H. Gandy Jr, ‘Statistical Surveillance: Remote Sensing in the Digital Age’, in Ball, Haggerty and Lyon, eds, *Routledge Handbook*, p. 128.
18. Berman, *Principles of Big Data*, p. 130.