

Big data problems we face today can be traced to the social ordering practices of the 19th century.

 blogs.lse.ac.uk/impactofsocialsciences/2015/10/13/ideological-inheritances-in-the-data-revolution/

10/13/2015

In the 19th century, changes in knowledge were facilitated not only by large quantities of new information pouring in from around the world but by shifts in the production, processing and analysis of that information. [Hamish Robertson](#) and [Joanne Travaglia](#) trace the connections between the 19th century data revolution and the present day one, outlining the implications this may have for the politics of big data in contemporary society. Two centuries after the first big data revolution, many of the problems and their solutions persist down to the present era.



This is not the first 'big data' era but the second. The first was the explosion in data collection that occurred from the early 19th century – [Hacking's](#) 'avalanche of numbers', precisely situated between 1820 and 1840. This was an analogue big data era, different to our current digital one but characterized by some very similar problems and concerns. Contemporary problems of data analysis and control include a variety of accepted [factors](#) that make them 'big' and these generally include size, complexity and technology issues. We also suggest that digitisation is a central process in this second big data era, one that seems obvious but which has also appears to have reached a new threshold. Until a decade or so ago 'big data' looked just like a digital version of conventional analogue records and systems. Ones whose management had become normalised through statistical and mathematical analysis. Now however we see a level of concern and anxiety, similar to the concerns that were faced in the first big data era.

This situation brings with it a socio-political dimension of interest to us, one in which our understanding of people and our actions on individuals, groups and populations are deeply implicated. The collection of social data had a purpose – understanding and controlling the population in a time of significant social change. To achieve this, new kinds of information and new methods for generating knowledge were required. Many ideas, concepts and categories developed during that first data revolution remain intact today, some uncritically accepted more now than when they were first developed. In this piece we draw out some connections between these two data 'revolutions' and the implications for the politics of information in contemporary society. It is clear that many of the problems in this first big data age and, more specifically, their solutions persist down to the present big data era.

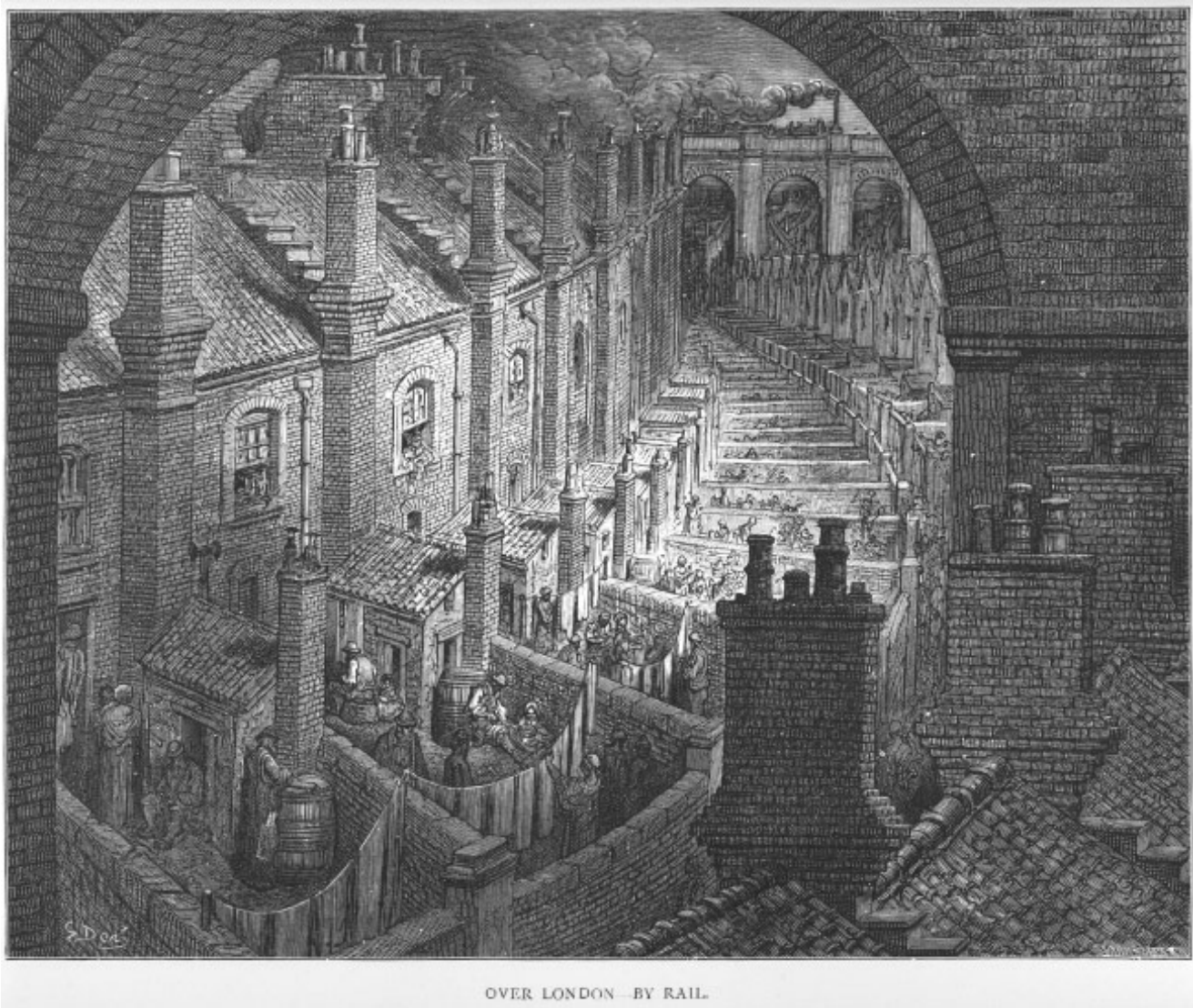


Image credit: [Gustave Doré's 19th-century engraving of London slums](#) (Public Domain)

Despite some discussion about the dates, there is general acknowledgement that the early 19th century was when the collection, analysis and production of various forms of information accelerated at a rate not previously seen in human history. More specifically, [Richards](#) called it the first information age. Linnaeus' botanical taxonomic approach proved so powerful a heuristic and practical device that it was swiftly applied to human social phenomena including the production of [racial taxonomies](#). The sciences as we know them were assuming their modern shape (Whewell coined the term 'scientist' in 1833), the social sciences were emerging from what were known as 'political arithmetic', 'social physics' and latterly the 'moral sciences', while science became an undertaking distinct from [natural philosophy](#).

Knowledge Technologies

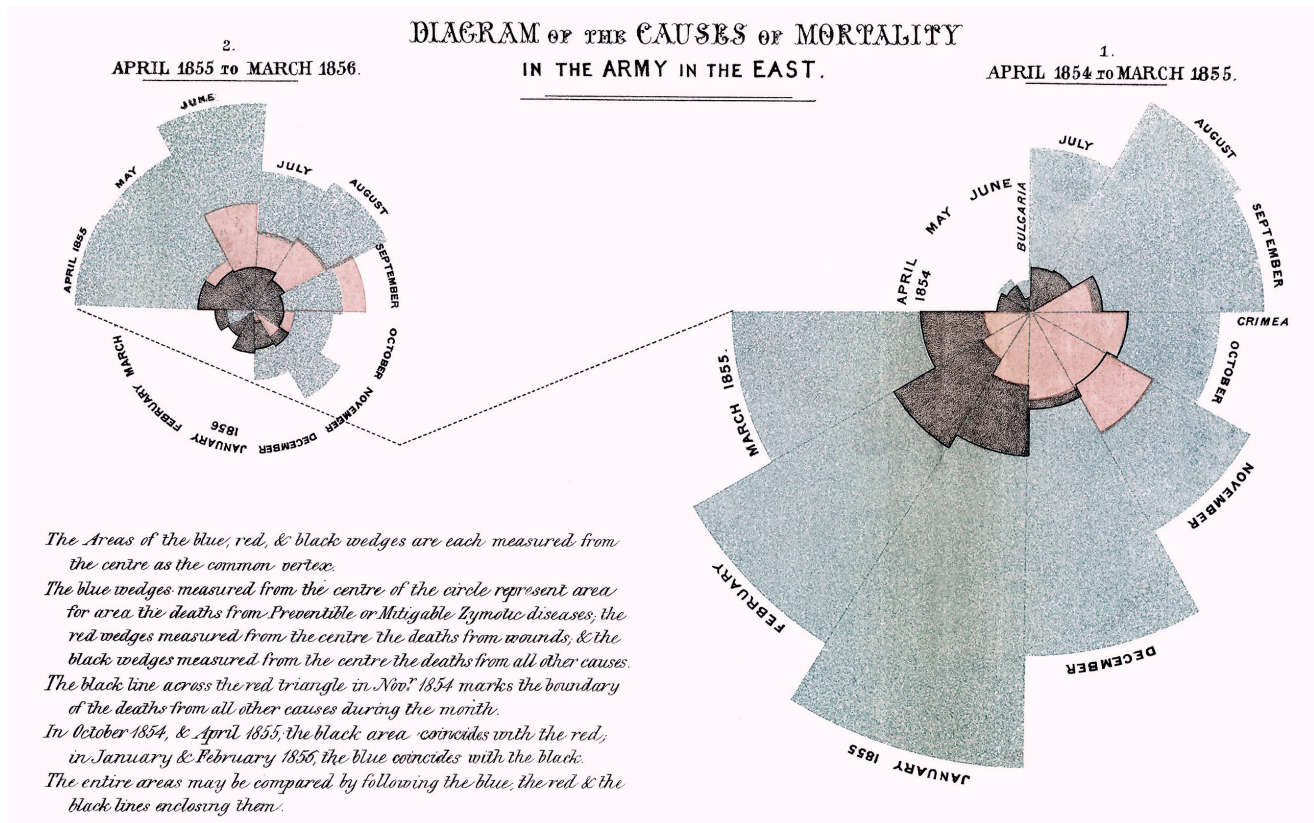
The 19th century was a pre-digital era in which the 'computer' was an individual at a desk doing the counting and calculations manually rather than an electro-mechanical or electronic device, but even this early infrastructure clearly set the scene for our current situation. The 18th century had already seen rapid developments in dictionaries of various kinds, including Diderot's 1751 *Encyclopédie* (based on Chamber's *Cyclopedia*) and Johnson's 1755 *Dictionary of the English Language* (not the first of its kind) illustrating a growing need to not just to collect but classify, categorise and order information to make it both meaningful and useful. The idea of and search for innate rules and regularity across a wide spectrum of phenomena emerged, with the search for [laws of nature](#) came in the

following century.

These information devices were supported by a growing number and variety of formalised knowledge production [processes and products](#) – the library, the museum, the census office, the printers and publishers with their books, newspapers, periodicals, magazines, journals, forms and envelopes. Cataloguing systems had existed for centuries but this period saw their emergence as formalized systems ranging from Brunet’s Paris Bookseller’s classification (1842) to the Dewey Decimal System (1876). Storage and retrieval also became an issue, leading to the development of library science, archival management strategies and mechanical handling systems.

In the context of colonial administration and scientific research fieldwork became a central concept, one which continues to be relevant to contemporary knowledge production in several disciplines and fields of practice (e.g. botany, geology, anthropology). The development of societies and associations also gained momentum as forums for identifying, exploring and formalizing new and expanding fields of knowledge. The convergence of this Foucauldian package of concepts, categories and practices generated a huge impetus towards data production in the Victorian period.

In the United Kingdom parliamentary Blue Books were being produced on an unprecedented scale as government increasingly concerned itself with the collection and analysis of data about this expanding information environment. They became such a phenomenon that many people despaired of their potential to overload bureaucratic knowledge systems that [lacked the capacity](#) to analyse the volumes of information being produced. Data visualisation and social mapping developed rapidly in response to this situation including the innovations of William Playfair (the line graph, bar and pie charts) and Florence Nightingale (polar diagrams) which provided new techniques for visualising these large and complex quantities of data.



“**Diagram of the causes of mortality in the army in the East**” by Florence Nightingale. (Public Domain)

These changes in knowledge were facilitated not only by large quantities of new information pouring in from around the world but by shifts in the production, processing and analysis of that information. Many of these methods are still with us including information taxonomies and knowledge trees to name but two. [Hacking](#) observed that while social categories are epistemic products their application can have marked ontological effects. Knowledge of the natural world was rapidly applied to the social world and the politicking of social identities began in earnest, supported by a rising tide of data and analytical methods. Conservatives and social critics alike relied on the production and dissemination of data, both large and small, to support repression and reform. The public inquiry emerged as another 19th century mechanism that persists in the present, with the same general focus – poverty, crime, health and systemic failures.

These new knowledge demands saw some contextual successes, such as in the demographic and statistical sciences, and some failures, such as Babbage's analytical engine design which was conceived but not completed during his lifetime. In some ways growing academic specialisations created a situation in which what was gained through a narrowing of focus and growth in sub-disciplinary activity was also lost in generalisability. This distinctly Victorian problem endures to the present day despite interdisciplinary projects of various kinds. [Floridi](#) writing on the philosophy of big data, has said quite specifically that the real big data problem we face today is less one of the quantity or quality of data or even technical skills but rather one of [epistemology](#).

Bureaucracy and Objectivity

Much of the data collected about human beings by bureaucratic systems has a history not simply of description or even understanding but one of control. Foucault's power/knowledge nexus is situated in a selection of bureaucratic and institutional forms for this reason. Every deviant or 'underperforming' social category is a warrant for action once documented. Consequently a great deal of social data is coercive in nature. Social data is rarely neutral and the persistence of 'wicked' social problems illustrates how regulation has been favoured in preference to their solution. That a census or a social survey is a snapshot of the way our societies are regulated is rarely remarked on and instead emphasis is given to the presumed objectivity of the categories and their data. This is the ideology of the small data era in action – the claim that it is science and not society that we are seeing through such instruments.

Classical sociology distinguishes between structure and agency. Many social policy and associated political debates essentially preference one or other of these dimensions as though they are both distinct and separate in theory and in social life. Structure is still equated with order against the potential horrors of anarchy, while agency remains couched in moral terms as personal responsibility. Bourdieu contradicted this separation using [habitus](#) to reconnect the two, but its epistemic influence is so powerful that many can no longer see the connection and debates lack the essential reflexivity he proposed.

The targets of social policy interventions for more than two centuries have essentially been the same categories of people – groups marked as moral outsiders (deviants) in their societies. The collection of data about these categories of people, in particular, was a marked feature of the first big data environment. These categories were operationalised through society's regulatory processes and institutions including education, the law and of course healthcare. These are the same locations where debates about structure, agency and morality continue to intersect and where the use of data and technology are represented as largely emancipatory. The risk is that 'big data' replicates the ideological underpinnings common to much of what has been produced under the small data paradigm.

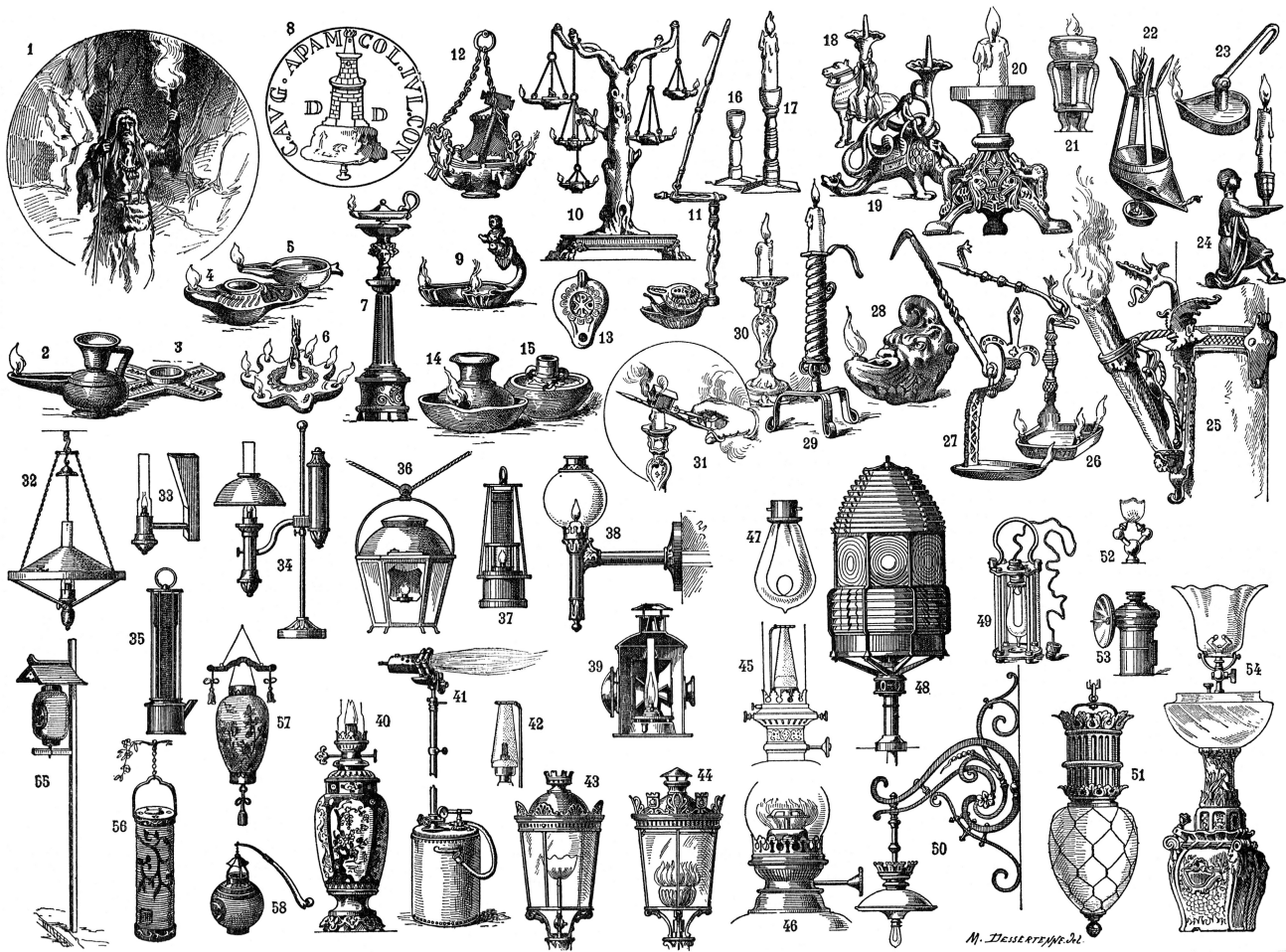


Image credit: [Maurice Dessertenne](#) – “Eclairage”, in *Nouveau Larousse Encyclopedia* (Public Domain)

Our question then is how do we go about re-writing the ideological inheritance of that first data revolution? Can we or will we unpack the ideological sequelae of that past revolution during this present one? The initial indicators are not good in that there is a pervasive assumption in this broad interdisciplinary field that reductive categories are both necessary and natural. Our social ordering practices have influenced our social epistemology. We run the risk in the social sciences of perpetuating the ideological victories of the first data revolution as we progress through the second. The need for critical analysis grows apace not just with the production of each new technique or technology but with the uncritical acceptance of the concepts, categories and assumptions that emerged from that first data revolution. That first data revolution proved to be a successful anti-revolutionary response to the numerous threats to social order posed by the incredible changes of the nineteenth century, rather than the Enlightenment emancipation that was promised.

Conclusion

Information is not new and nor is data – of whatever order of magnitude. We are in a period that can reasonably be seen as the second ‘big data’ revolution and it is revolutionary because it challenges our accepted understanding of the world and not simply because of the volumes and velocity of data generation in our new digital information technologies. Many social categories were designed to control, coerce and even oppress their targets. The poor, the unmarried mother, the illegitimate child, the black, the unemployed, the disabled, the dependent elderly – none of these social categories of person is a neutral framing of individual or collective circumstances. They are instead a judgement on their place in modernity and material grounds for research, analysis and policy interventions of various

kinds. Two centuries after the first big data revolution many of these categories remain with us almost unchanged and, given what we know of their consequences, we have to ask what will be their situation when this second data revolution draws to a close?

Like that first data revolution, this present one also has ambitions for people and their interactions with the new media emerging in its wake. These discussions are useful and necessary because discussion and negotiation are essential in the face of revolution. The responses to revolution in the late 18th and 19th centuries were often violent but we now have better methods available for the maintenance of social order as Foucault's technologies of the self and Bourdieu's habitus. Where we see this becoming highly problematic is in the continuity of ideologically informed notions of ourselves and others and the reproduction of such ideologies in and through our new digital environments. Following Floridi, this is a significant epistemic and ethical problem in our current big data era.

This is part of a wider series on the [Politics of Data](#). For more on this topic, also see Mark Carrigan's [Philosophy of Data Science](#) interview series and the Discover Society [special issue on the Politics of Data \(Science\)](#).

Note: This article gives the views of the authors, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Authors

[Hamish Robertson](#) is a geographer at the University of New South Wales with experience in healthcare including a decade in ageing research. He has worked in the private, public and not-for-profit sectors and he has presented and published on a variety of topics ranging from ageing, diversity, health informatics, Aboriginal health, patient safety and spatial science to cultural heritage research. Hamish is currently completing his PhD on the geography of Alzheimer's disease and recently finished editing a book on museums and older people.

[Joanne Travaglia](#) is a medical sociologist at the University of New South Wales with experience in the health field as a practitioner, manager, researcher and educator. Her research addresses various aspects of health services management and leadership, with a particular focus on the impact of patient and clinician vulnerability and diversity on the safety and quality of care.

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.