1 Toward Critical Data Studies

Charting and Unpacking Data Assemblages and Their Work

Rob Kitchin and Tracey P. Lauriault

A Critical Approach to Data

Societies have collected, stored, and analyzed data for a couple of millennia as a means to record and manage their activities. For example, the ancient Egyptians collected administrative records of land deeds, field sizes, and livestock for taxation purposes, the Domesday Book in 1086 captured demographic data, double-entry bookkeeping was used by bankers and insurers in the fourteenth century, and the first national registry was undertaken in Sweden in the seventeenth century.[1] It was not until the seventeenth century, however, that the term "data" was used for the first time in the English language, thanks to the growth of science, the development of statistics, and the shift from knowledge built from theology, exhortation, and sentiment to facts, evidence, and the testing of theory through experiment.[2] Over time the importance of data has grown, becoming central to how knowledge is produced, business conducted, and governance enacted. Data provide the key inputs to systems that individuals, institutions, businesses, and the sciences employ in order to understand, explain, manage, regulate, and predict the world we live in and are used to create new innovations, products, and policies.

The volume, variety, and use of data have grown enormously since the seventeenth century, and there has long been the creation and maintenance of very large data sets, such as censuses or government administrative and natural resource databases. Such databases, however, have typically been generated every few years or are sampled. In contrast, over the

past fifty years we have begun to enter the era of big data, with such characteristics as being

- huge in *volume*, consisting of terabytes or petabytes of data;
- high in *velocity*, being created in or near real time;
- diverse in *variety*, being structured and unstructured in nature;
- *exhaustive* in scope, striving to capture entire populations or systems (n = all);
- fine-grained in *resolution* and uniquely *indexical* in identification;
- *relational* in nature, containing common fields that enable the conjoining of different data sets; and
- *flexible*, holding the traits of *extensionality* (new fields can easily be added) and *scalability* (data sets can expand in size rapidly).[3]

While there are varying estimates, depending on the methodology used, as to the growth of data production caused in the main by the production of big data, in addition to a steep growth in small data such as personal video, photo, and audio files (all of which consume large amounts of data storage), it is clear that there has been a recent step change in the volume of data generated, especially since the start of the new millennium.[4] Gantz and Reinsel have estimated that data volumes had grown by a factor of nine in the preceding five years, and Manyika et al. have projected a 40 percent rise in data generated globally per year.[5] In 2013 EU Commissioner for the Digital Agenda Neelie Kroes reported that 1.7 million billion bytes of data per minute were being generated globally.[6] Such rises and projections for further increases are due to the continuous and exhaustive, rather than sampled, production of born digital data, in combination with the nature of some of those data (e.g., image and video files) and the increased ability to store and share such data at marginal cost. For example, in 2012 Facebook reported that it was processing 2.5 billion pieces of content (links, comments, etc.), 2.7 billion "Like" actions, and 300 million photo uploads *per day*, and Walmart was generating more than 2.5 petabytes (2^{50} bytes) of data relating to more than 1 million customer transactions *every hour*.[7]

These massive volumes of data are being produced by a diverse set of information and communication technologies that increasingly medi-

ate and augment our everyday lives, for example, digital CCTV, retail checkouts, smartphones, online transactions and interactions, sensors and scanners, and social and locative media. As well as being produced by government agencies, vast quantities of detailed data are now being generated by mobile phone operators, app developers, Internet companies, financial institutions, retail chains, and surveillance and security firms, and data are being routinely traded to and between data brokers as an increasingly important commodity. More and more analog data held in archives and repositories are being digitized and linked together and made available through new data infrastructures, and vast swaths of government-produced and held data are being made openly accessible as the open data movement gains traction.[8]

This step change in data production has prompted critical reflection on the nature of data and how they are employed. As the concept of data developed, data largely came to be understood as being pre-analytical and prefactual—that which exists prior to interpretation and argument or the raw material from which information and knowledge are built. From this perspective data are understood as being representative, capturing the world as numbers, characters, symbols, images, sounds, electromagnetic waves, bits, and so on, and holding the precepts of being abstract, discrete, aggregative (they can be added together), nonvariant, and meaningful independent of format, medium, language, producer, and context (i.e., data hold their meaning whether stored as analog or digital, viewed on paper or screen, or expressed in different languages).[9] Data are viewed as being benign, neutral, objective, and nonideological in essence, reflecting the world as it is subject to technical constraints; they do not hold any inherent meaning and can be taken at face value.[10] Indeed the terms commonly used to detail how data are handled suggest benign technical processes: "collected," "entered," "compiled," "stored," "processed," and "mined."[11] In other words it is only the uses of data that are political, not the data themselves.

This understanding of data has been challenged in recent years. Contrary to the notion that data is pre-analytic and prefactual is the argument that data are constitutive of the ideas, techniques, technologies, people,

systems, and contexts that conceive, produce, process, manage, and analyze them.[12] In other words, how data are conceived, measured, and employed actively frames their nature. Data do not pre-exist their generation; they do not arise from nowhere, and their generation is not inevitable: protocols, organizational processes, measurement scales, categories, and standards are designed, negotiated, and debated, and there is a certain messiness to data generation. As Gitelman and Jackson put it, "raw data is an oxymoron"; "data are always already 'cooked.'"[13] Data then are situated, contingent, relational, and framed and are used contextually to try and achieve certain aims and goals.

Databases and repositories are also not simply a neutral, technical means of assembling and sharing data but are bundles of contingent and relational processes that do work in the world.[14] They are complex socio-technical systems that are embedded within a larger institutional landscape of researchers, institutions, and corporations and are subject to socio-technical regimes "grounded in . . . engineering and industrial practices, technological artifacts, political programs, and institutional ideologies which act together to govern technological development."[15] Databases and repositories are expressions of knowledge/power, shaping what questions can be asked, how they are asked, how they are answered, how the answers are deployed, and who can ask them.[16]

Beyond this philosophical rethinking of data, scholars have begun to make sense of data ethically, politically and economically, spatially and temporally, and technically.[17] Data can concern all aspects of everyday life, including sensitive issues, and be used in all kinds of ways, including to exploit, discriminate against, and persecute people. There are then a series of live moral and ethical questions concerning how data are produced, shared, traded, and protected; how data should be governed by rules, principles, policies, licenses, and laws; and under what circumstances and to what ends data can be employed. There are no simple answers to such questions, but the rise of more widespread and invasive data generation and more sophisticated means of data analysis creates an imperative for public debate and action. In addition data are framed by political concerns as to how they are normatively conceived and contested as public

and private goods. The open data and open government movements, for example, cast data as a public commons that should be freely accessible. Business, in contrast, views data as a valuable commodity that, on the one hand, needs to be protected through intellectual property regimes (copyright, patents, ownership rights) and, on the other, should be exploitable for capital gain. Indeed data often constitute an economic resource: for government they are sold under cost-recovery regimes and for business they are tradable commodities to which additional value can be added and extracted (e.g., derived data, analysis, knowledge). In the present era data are a key component of the emerging knowledge economy enhancing productivity, competitiveness, efficiencies, sustainability, and capital accumulation. The ethics, politics, and economics of data develop and mutate across space and time with changing regimes, technologies, and priorities. From a technical perspective, there has been a focus on how to handle, store, and analyze huge torrents of data, with the development of data mining and data analytics techniques dependent on machine learning, and there have been concerns with respect to data quality, validity, reliability, authenticity, usability, and lineage.

In sum we are starting to witness the development of what Dalton and Thatcher call critical data studies—research and thinking that apply critical social theory to data to explore the ways in which they are never simply neutral, objective, independent, raw representations of the world but are situated, contingent, relational, contextual, and do active work in the world.[18] In their analysis Dalton and Thatcher set out seven provocations needed to provide a comprehensive critique of the new regimes of data:

- situating data regimes in time and space;
- exposing data as inherently political and identifying whose interests they serve;
- unpacking the complex, nondeterministic relationship between data and society;
- illustrating the ways in which data are never raw;
- exposing the fallacies that data can speak for themselves and that big data will replace small data;

- exploring how new data regimes can be used in socially progressive ways; and
- examining how academia engages with new data regimes and the opportunities of such engagement.

We agree with the need for all of these provocations. In a short presentation at a meeting of the Association of American Geographers one of us set out a vision for what critical data studies might look like: unpacking the complex assemblages that produce, circulate, share/sell, and utilize data in diverse ways; charting the diverse work they do and their consequences for how the world is known, governed, and lived in; and surveying the wider landscape of data assemblages and how they interact to form intersecting data products, services, and markets and shape policy and regulation. It is to this endeavor that we now turn.

Charting and Unpacking Data Assemblages

Kitchin defines a data assemblage as a complex socio-technical system that is composed of many apparatuses and elements that are thoroughly entwined and whose central concern is the production, management, analysis, and translation of data and derived information products for commercial, governmental, administrative, bureaucratic, or other purposes (see table 1-1).[19] A data assemblage consists of more than the data system or infrastructure itself, such as a big data system, an open data repository, or a data archive, to include all of the technological, political, social, and economic apparatuses that frame their nature, operation, and work. The apparatuses and elements detailed in table 1-1 interact with and shape each other through a contingent and complex web of multifaceted relations. And just as data are a product of the assemblage, the assemblage is structured and managed to produce those data.[20] Data and their assemblage are thus mutually constituted, bound together in a set of contingent, relational, and contextual discursive and material practices and relations. For example, the data assemblage of a census consists of a large amalgam of apparatuses and elements that shape how the census is formulated, administered, processed, and communicated and how its

findings are employed. A census is underpinned by a realist system of thought; it has a diverse set of accompanying forms of supporting documentation; its questions are negotiated by many stakeholders; its costs are a source of contention; its administering and reporting are shaped by legal frameworks and regulations; it is delivered through a diverse set of practices, undertaken by many workers, using a range of materials and infrastructures; and its data feed into all kinds of uses and secondary markets. Data assemblages evolve and mutate as new ideas and knowledges emerge, technologies are invented, organizations change, business models are created, the political economy changes, regulations and laws are introduced and repealed, skill sets develop, debates take place, and markets grow or shrink. And while data sets once generated within an assemblage may appear fixed and immutable (e.g., a compiled census), they are open to correction and revision, reworking through disaggregation and reaggregation into new classes or statistical geographies, parsing into other data systems, data derived and produced from them, and alternative interpretations and insights drawn from them. Data assemblages and their data are thus always in a state of becoming.

This notion of a data assemblage is similar to Foucault's concept of the *dispositif*, which refers to a "thoroughly heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral[,] and philanthropic propositions" that enhance and maintain the exercise of power within society.[21] The dispositif of a data infrastructure produces what Foucault terms "power/knowledge," that is, knowledge that fulfills a strategic function: "the apparatus is thus always inscribed in a play of power, but it is also always linked to certain coordinates of knowledge which issue from it but, to an equal degree, condition it. This is what the apparatus consists in: strategies of relations of forces supporting, and supported by, types of knowledge."[22] In other words, data infrastructures are never neutral, essential, objective; their data are never raw but always cooked to some recipe by chefs embedded within institutions that have certain aspirations and goals and operate within wider frameworks.

TABLE 1-1. Apparatus and elements of a data assemblage

APPARATUS	ELEMENTS
Systems of thought	Modes of thinking, philosophies, theories, models, ideologies, rationalities, etc.
Forms of knowledge	Research texts, manuals, magazines, websites, experience, word of mouth, chat forums, etc.
Finance	Business models, investment, venture capital, grants, philanthropy, profit, etc.
Political economy	Policy, tax regimes, incentive instruments, public and political opinion, etc.
Governmentalities and legalities	Data standards, file formats, system requirements, protocols, regulations, laws, licensing, intellectual property regimes, ethical considerations, etc.
Materialities and infrastructures	Paper/pens, computers, digital devices, sensors, scanners, databases, networks, servers, buildings, etc.
Practices	Techniques, ways of doing, learned behaviors, scientific conventions, etc.
Organizations and institutions	Archives, corporations, consultants, manufacturers, retailers, government agencies, universities, conferences, clubs and societies, committees and boards, communities of practice, etc.
Subjectivities and communities	Data producers, experts, curators, managers, analysts, scientists, politicians, users, citizens, etc.
Places	Labs, offices, field sites, data centers, server farms, business parks, etc., and their agglomerations
Marketplace	For data, its derivatives (e.g., text, tables, graphs, maps), analysts, analytic software, interpretations, etc.

This cooking of data is revealed through the work of Ian Hacking, who drew inspiration from Foucault's thinking on the production of knowledge.[23] Hacking posits that within a data assemblage there are two interrelated processes at work that produce and legitimate its data and associated apparatuses and elements, shaping how its data do work in the world, that in turn influence future iterations of data and the constitution of the assemblage. In both cases he posits that a dynamic nominalism is at work, wherein there is an interaction between data and what they represent, leading to mutual changes.

The first of these processes is what Hacking terms the "looping effect."[24] The looping effect concerns how data are classified and organized, how a data ontology comes into existence, and how it can reshape that which has been classified. The loop (fig. 1-1) has five stages:

1. classification, wherein things that are regarded as having shared characteristics are grouped together or, in cases of deviance, forced into groupings;
2. objects of focus (e.g., people, spaces, fashions, diseases, etc.) wherein, in the case of people, individuals eventually start to identify with the class into which they are assigned or, in the case of nonhuman objects, people come to understand and act toward the objects according to their classification;
3. institutions, which institutionalize classifications and manage data infrastructures;
4. knowledge, which is used to formulate, reproduce, and tweak classifications; and
5. experts, being those within institutions who produce and exercise knowledge, implementing the classification.

Through this looping effect Hacking argues that a process of "making people up" occurs in data systems such as the census or the assessing of mental health, wherein the systems of classification work to reshape society in the image of a data ontology. Examples could include people defining themselves or being defined by mental health symptoms, as well

as a system of mental health facilities being built and staffed by specialist professionals.

The second of the processes consists of what Hacking terms "engines of discoverability" that extend beyond simply methods. He discusses these methods using a medical lens, which Lauriault has modified to incorporate the making up of spaces as well as people.[25] Hacking posits that there are a number of such engines, the last three of which are derived engines that are

a. counting the volumes of different phenomena;
b. quantifying: turning counts into measures, rates, and classifications;
c. creating norms: establishing what might or should be expected;
d. correlation: determining relationships between measures;
e. taking action: employing knowledge to tackle and treat issues;
f. scientification: establishing and adopting scientific knowledge;
g. normalization: seeking to fashion the world to fit norms (e.g., encouraging diets to meet expected body mass indices);
h. bureaucratization: putting in place institutions and procedures to administer the production of expectations and to undertake action; and
i. resistance to forms of knowledge, norms, and bureaucracy by those who are affected in negative ways (e.g., homosexual and disabled people's resistance to medicalized models that class, position, and treat them in particular ways) or those forwarding alternative systems, interpretations, and visions.[26]

Together these engines undertake the work of a data assemblage at the same time as it legitimates and reproduces such work and the assemblage itself. For example, a census counts a population and aspects of people's lives, turns that information into measures, establishes baseline rates, assesses relationships between factors, and is transformed into knowledge, which leads to practices of normalization and is enacted by dedicated and related bureaucracy. Each stage reinforces the previous, and collectively they justify the work it does. The knowledge produced and indeed the whole assemblage can be resisted, as with the census boycotts

in Germany in the 1980s or with campaigns to ensure that Irish ethnicity is not undercounted in the UK, that "New Zealander" is accepted as an ethnicity in New Zealand (instead of "New Zealand European"), and that women's unpaid work is accounted for, or the knowledge produced can be transgressed, as in the case of those who report their religion as Jedi.[27] It can indeed even be canceled, as in the 2011 long-form census of Canada.

Data assemblages form part of a wider data landscape composed of many interrelated and interacting data assemblages and systems. Within the public sector, for example, there are thousands of data systems (each one surrounded by a wider assemblage) that interact and work in concert to produce state services and forms of state control at the local, regional, and national levels. Often this data landscape extends to the pan-national and the global scale, through interregional and worldwide data sets, data-sharing arrangements and infrastructures, and the formulation of protocols, standards, and legal frameworks (e.g., Global Spatial Data Infrastructures, INSPIRE). Firms within industry likewise create and occupy a complex data landscape, selling, buying, and sharing data from millions of data systems, all part of wider socio-technical assemblages. For example, the data landscape of big data consists of hundreds of companies, ranging from small and local to large and global, that provide a range of complementary and competing services, such as cooked data, specialty compilers and aggregators, data analytics, segmentation tools, list management, interpretation and consulting, marketing, publishing, and research and development. We have barely begun to map out various data landscapes, their spatialities and temporalities, their complex political economy, and the work that they do in capturing, analyzing, and reshaping the world. It is to the latter we now turn.

Uncovering the Work of Data Assemblages

As noted in the previous section, data assemblages do work in the world. Data are being leveraged to aid the tasks of governing people and territories, managing organizations, producing capital, creating better places, improving health care, advancing science, and so on. This leveraging takes many forms, but the central tenet is that data, if analyzed and exploited

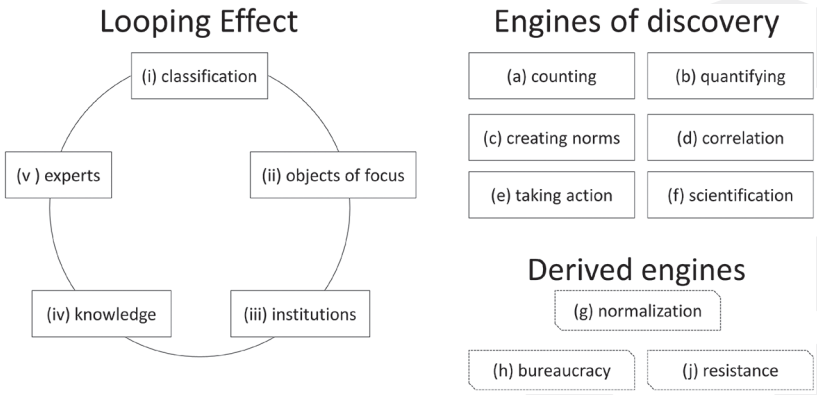

FIG. 1-1. The working of a data assemblage, following Hacking, "Philosophie et histoire des concepts scientifiques," and Laney, 3D *Data Management*. Created by R. Kitchin and T. Lauriault.

appropriately, produce information and knowledge that can be used to reshape operating procedures and organizational structure, identify new products, segment markets, reduce uncertainty and risk, and increase efficiency, productivity, competitiveness, and sustainability.[28] While much of the work to which data are put is beneficial to wider society, with data being used to improve quality of life and to tackle humanitarian and environmental issues, there is also a darker side to much data work. Here we want to consider the latter, highlighting four ways in which data are being employed to produce pernicious social and economic relations: dataveillance and the erosion of privacy, profiling and social sorting, anticipatory governance, and secondary uses and control creep. These practices are currently the subject of much debate, and there is an urgent need for critical studies that can inform the arguments being made.

As the revelations of WikiLeaks, Edward Snowden and other whistle blowers, the Maher Arar case, and other legal challenges with respect to erroneous record keeping and the mistreatment of individuals have demonstrated, from 9/11 onward there has been a step change in the extent and nature of state-led surveillance and securitization in many nations. Vast quantities of everyday communications (telephone calls, text messages,

emails, social media), as well as general Internet use, are being routinely and systematically gathered by organizations such as the U.S. National Security Agency and analyzed for strategic intelligence.[29] All nation-states similarly gather large databases of information about citizens with respect to all aspects of their lives—income, tax, welfare, health, education, and so on. Likewise companies now routinely generate data with respect to all aspects of their business, including their customers and their patterns of consumption. Indeed given the mediating role of software in tasks such as working, traveling, consuming, communicating, and playing, it is increasingly difficult to take part in daily life without leaving a digital trace.[30] For example, the Dutch Data Protection Authority estimates that the average Dutch citizen is recorded in 250 to 500 databases, with some in up to 1,000 databases—a figure that is growing.[31] These databases not only include individuals' digital footprints (data they themselves leave behind) but also individuals' data shadows (information about them generated by others). Those to whom the data refer often have little control over the data generated, their form, extent, or how they are used.[32] Individually these databases provide limited views of people, but they gain power when combined, revealing detailed patterns and enabling what has been termed dataveillance—the sorting and sifting of data sets in order to identify, monitor, track, regulate, predict, and prescribe.[33] The widespread generation of data and the practices of dataveillance raise many questions concerning privacy and rights to anonymity and confidentiality that are only just starting to be thought through and responded to.[34]

Data have long been used to profile, segment, and manage populations, but these processes have become much more sophisticated, fine-tuned, widespread, and routine with the application of data analytics employing machine learning techniques.[35] While the state might profile its citizens for the purposes of security and policing, commercial enterprises are seeking to reduce risk and maximize yield through more effective targeting of products. Whereas earlier generations of profiling sought to create aggregated population or area profiles, which then shaped decision making with regard to marketing and product placement (e.g., geodemographic profiling), new generation analytics can work at the level of the individual,

combining data from various sources such as credit and store card transactions, clickstreams, social media posts, and other kinds of personal data to produce a detailed customer profile.[36] These profiles are used to socially sort customers, identifying some for preferential treatment and excluding others, and to predict the likelihood that customers might be able to meet payments or to judge their projected lifetime value if they remain loyal, and how likely they are to move their custom.[37] They are also being used to underpin new forms of dynamic and personalized pricing, tailored to a consumer's profile and purchase history, that are designed to leverage optimal spending.[38] Consumers are thus being routinely measured and ranked, and they receive differential services, based on their associated data and where they live.

One particularly pernicious form of predictive profiling is anticipatory governance. It involves predictive analytics that are used to assess likely future behaviors or events and to direct appropriate action. Such anticipatory governance has been a feature of air travel for a number of years, with passengers profiled for risk and levels of security checks prior to starting their journey.[39] More recently it has been extended to general policing, with a number of U.S. police forces using it to identify potential future criminals and to direct the patrolling of areas based on an analysis of historical crime data, records of arrests, and the known social networks of criminals.[40] In such cases individuals' data shadows do more than follow them; the data shadow precedes them, seeking to police behaviors that may never occur.[41] As a consequence, people are treated differently in anticipation of something they may or may not do. Given their effects vis-à-vis individual lives and their black-boxed nature, the practices of predictive profiling, social sorting, and anticipatory governance require much more attention, as do the companies that develop and undertake such tasks.

The work that data systems do in all of these cases is based on generating an excess of data. Indeed big data is premised on generating, hoarding, and linking as much data as possible in the hope that value and insight can be leveraged from them. Rather than being generated and used to fulfill a specific task, data can be repackaged, sold, and repurposed for all kinds of secondary uses. Such a strategy runs counter to the policy of data

minimization, one of the foundations of privacy and data protection in the European Union and North America. This policy stipulates that data should only be generated and used to perform a particular task and that they should be retained only for as long as they are required to perform that task.[42] A clear example of where the premise of data minimization is being breached is with respect to control creep, in which data generated for one form of governance is appropriated for another.[43] Clearly control creep has mostly occurred with respect to security, with airline industry and government administrative data being repurposed for profiling and assessing passenger risk; with congestion charge cameras installed for that sole purpose also being used for general policing; and with social media data being repurposed to conduct criminal investigations and undertake predictive profiling.[44] But control creep is also in evidence across a range of other domains, for example, using personal location, consumption, and social media data to assess credit risk or suitability for employment.[45] Given the implications for civil liberties from secondary data use, there is a need to examine its consequences and to design new approaches to data protection, such as privacy by design.[46]

Conclusion

Dalton and Thatcher conclude their call for critical data studies by setting out five questions that they believe require further study, all relating to big data:

- What historical conditions lead to the realization of big data such as they are?
- Who controls big data, its production, and its analysis? What motives and imperatives drive their work?
- Who are the subjects of big data and what knowledges are they producing?
- How is big data actually applied in the production of spaces, places, and landscapes?
- What is to be done with big data and what other kinds of knowledges could it help produce?[47]

There are many more questions that can be added to this list, not least by widening the lens to open data, as well as data archives and repositories, but also by considering the wider data landscape, data assemblages, and data markets. Rather than produce an extensive list of questions, we want to conclude by calling for greater conceptual work and empirical research to underpin and flesh out critical data studies.

The ways in which data are being generated, the analytics used to process and extract insight from them, the industries growing up around them, their wider political economic framing, and how they are employed all demand critical engagement. While there is a rich and diverse tradition of critical social theory that can be directed toward data assemblages and the wider data landscape, such theory needs to be refined and fine-tuned to make sense of data and their work in the world, with new theory developed where needed. Yet we have barely begun to critically conceptualize data and their apparatus and elements. Such thinking needs to be complemented with more normatively oriented reflection on the ethics and politics of big data, open data, and data systems of different varieties.

Such conceptual and normative assessments need to be accompanied by a diverse set of empirical case studies that examine all facets of data-driven governance, business, and science, that unpack data assemblages, and that map the wider data landscape. Our suggested approach is to employ methods such as ethnographies, interviews, focus groups, and participant observation to delve into the workings of assemblages, to trace out genealogies of how the data landscape has changed over time and space, to map the materialities and infrastructures that constitute data infrastructures, and to deconstruct the discursive regime accompanying data-driven initiatives.[48]

Undertaking this conceptual and empirical work is what our own research will focus on over the next few years as part of the Programmable City project, building on our initial large-scale studies.[49] This extensive project is examining the intersections of big and open data, ubiquitous computing, software and algorithms, and smart city developments in Dublin and Boston, unpacking a set of data assemblages and charting the data landscape of each city. We have no doubt that many others will

be engaging in similar studies, given the growth in data-driven forms of science, business, and government. We hope that what this research will produce is a diverse set of vibrant critical data studies.

Notes

1. Dupaquier and Dupaquier, *Histoire de la démographie*; Bard and Shubert, *Encyclopedia of the Archaeology*; Poovey, *History of the Modern Fact*; Porter, *Rise of Statistical Thinking*.
2. Poovey, *History of the Modern Fact*; Garvey, "*facts* and FACTS"; Rosenberg, "Data before the Fact."
3. Kitchin, "Big Data and Human Geography," 262; boyd and Crawford, "Critical Questions for Big Data"; Dodge and Kitchin, "Codes of Life"; Laney, *3D Data Management*; Marz and Warre, *Big Data: Principles*; Mayer-Schönberger and Cukier, *Big Data: Revolution*; Zikopoulos et al., *Understanding Big Data*.
4. See, for example, Hilbert and López, "World's Technological Capacity"; Gantz and Reinsel, *Extracting Value from Chaos*; and Short et al., *How Much Information?*
5. Gantz and Reinsel, *Extracting Value from Chaos*; Manyika et al., *Big Data: Next Frontier*.
6. Rial, "Power of Big Data."
7. Constine, "How Big Is Facebook's Data?"; Open Data Center Alliance, *Big Data Consumer Guide*.
8. Lauriault et al., "Today's Data"; Kitchin, *Data Revolution*.
9. Floridi, "Data"; Rosenberg, "Data before the Fact."
10. Pérez-Montoro and Díaz Nafría, "Data."
11. Gitelman and Jackson, "Introduction."
12. Bowker and Star, *Sorting Things Out*; Lauriault, "Data, Infrastructures"; Ribes and Jackson, "Data Bite Man"; Kitchin, *Data Revolution*.
13. Gitelman and Jackson, "Introduction," 2, citing Bowker, *Memory Practices*.
14. Star and Ruhleder, "Steps toward an Ecology"; Kitchin and Dodge, *Code/Space*.
15. Ruppert, "Governmental Topologies"; Hecht, "Technology, Politics, and National Identity," 257.
16. Lauriault, "Data, Infrastructures"; Ruppert, "Governmental Topologies."
17. Kitchin, *Data Revolution*.
18. Dalton and Thatcher, "Critical Data Studies."
19. Kitchin, *Data Revolution*, 24.
20. Ribes and Jackson, "Data Bite Man."

21. Foucault, "Confession of the Flesh," 194.
22. Foucault, "Confession of the Flesh," 196.
23. Hacking, "Biopower"; Hacking, "Making Up People"; Hacking, "Tradition of Natural Kinds"; Hacking, "Philosophie et histoire des concepts scientifiques"; Hacking, "Kinds of People."
24. Hacking, "Tradition of Natural Kinds"; Hacking, "Philosophie et histoire des concepts scientifiques"; Hacking, "Kinds of People."
25. Lauriault, "Data, Infrastructures."
26. Hacking, "Philosophie et histoire des concepts scientifiques."
27. Hannah, *Dark Territory*; UK Census, *Irish in Britain*; Middleton, "Email Urges 'New Zealander'"; Waring, *If Women Counted*; Singler, "SEE MOM IT IS REAL."
28. Kitchin, *Data Revolution*.
29. Amoore, "Biometric Borders"; Bamford, *Shadow Factory*.
30. Kitchin and Dodge, *Code/Space*.
31. Koops, "Forgetting Footprints."
32. CIPPIC, *On the Data Trail*.
33. Clarke, "Information Technology"; Raley, "Dataveillance and Countervailance."
34. Solove, "Taxonomy of Privacy"; Elwood and Leszczynski, "Privacy, Reconsidered."
35. Weiss, *Clustering of America*; Goss, "'We Know Who You Are'"; Parker, Uprichard, and Burrow, "Class Places and Place Classes"; Singleton and Spielman, "Past, Present and Future."
36. Siegel, *Predictive Analytics*.
37. Graham, "Software-Sorted Geographies"; Minelli et al., *Big Data, Big Analytics*.
38. Tene and Polonetsky, "Big Data for All."
39. Dodge and Kitchin, "Flying through Code/Space"; Amoore, "Biometric Borders."
40. Siegel, *Predictive Analytics*; Stroud, "Minority Report."
41. Stalder, "Privacy Is Not the Antidote"; Harcourt, *Against Prediction*.
42. Tene and Polonetsky, "Big Data for All"; CIPPIC, *Submissions to the House of Commons*.
43. Innes, "Control Creep."
44. Lyon, *Surveillance Studies*; Pither, *Dark Days*; Gallagher, "Staking Out Twitter."
45. O'Reilly, "Creep Factor."
46. Information and Privacy Commissioner/Ontario, *Seven Principles of Privacy*.
47. Dalton and Thatcher, "Critical Data Studies."
48. Kitchin, *Data Revolution*.
49. See Programmable City, http://www.nuim.ie/progcity/; Lauriault, "Data, Infrastructures"; and Kitchin, *Data Revolution*.