Introduction

New Brave World

Karin van Es & Mirko Tobias Schäfer

Optimistic reporting about 'big data' has made it easy to forget that data-driven practices have been part of the emerging information society since the nineteenth century (Beniger 1989; Porter 1996; Campbell-Kelly 2003). In lieu of an illustrative metaphor, the label 'big data' is used to describe a set of practices involving the collection, processing and analysis of large data sets. The term enables members of the general public to engage in debates, albeit often uninformed, on the ongoing transformation of our knowledge economy, but it disguises more than it reveals. Nevertheless, despite its vagueness, the term captures something of significance about contemporary Western societies, where economic value is generated through the processing of information and the monetization of knowledge. To develop a critical understanding of this current situation and its societal consequences, it is important to debunk the exceptionalism inherent in the 'big data' paradigm. For starters, we must stop feeding the hype about it and lay out what we know: the phenomenon we are dealing with is not 'big data', but 'the computational turn' (Berry 2012; Braidotti 2013). This turn began in the 1950s with the introduction of electronic computers and continues unabated today. It concerns the datafication of everything: all aspects of life are now transformed into quantifiable data (Mayer-Schönberger & Cukier 2013). As the social is extensively mined, its data are used to predict human behaviour and automate decision-making processes. As José van Dijck claims, 'datafication as a legitimate means to *access*, *understand* and *monitor* people's behaviour is becoming a leading principle, not just amongst techno-adepts, but also amongst scholars who see datafication as a revolutionary research opportunity to investigate human conduct' (2014: 198). Data analysis promises an 'objective' way to grasp the complex and dynamic reality we live in. Visualized via colourful dashboards, info-graphics and charts, it puts forth, persuasively and seductively, a seemingly accurate and unbiased assessment of reality. However, the translation of the social into data involves a process of abstraction that compels certain compromises to be made as the data are generated, collected, selected and analysed (Langlois et al. 2015).

The New Empirical

Because datafication is taking place at the core of our culture and social organization, it is crucial that humanities scholars tackle questions about how this process affects our understanding and documentation of history, forms of social interaction and organization, political developments, and our understanding of democracy. That datafication is a phenomenon that urgently demands investigation was acknowledged in sociology more than a decade ago, but this recognition did not necessarily lead to the adoption of novel data practices or the reassessment of existing research agendas. In a programmatic article, Andrew Abbott (2000) pointed out the challenges for researchers when confronted with new data resources available on an unprecedented scale:

> There is little question that a gradual revolution in the nature of knowledge is taking place: a slow eclipsing of print by visual representation, a move toward knowledge that is more experimental and even aleatory, an extensive commodification of important parts of previously esoteric knowledge. (298)

This forecast has been borne out by the developments of the past decade and a half, and these processes of change have indeed intensified. The Google search engine and commercial social media platforms such as Facebook, Twitter, YouTube and Instagram continually generate data from the interactions of millions of users. Access to data/tools is sold to marketeers and is employed to target, predict and manage these platforms' users. So-called application programming interfaces (APIs) make parts of vast databases accessible to third parties, including researchers. Concurrently, an industry has emerged whose companies collect, sell, combine and analyse data sets for all kinds of purposes, ranging from targeted advertising and market research to credit ratings, risk assessments and mass surveillance. The collection of data from massive data sets also yields a glimpse of a future when certain sorts of businesses will thrive on the exploitation of vast amounts of stored information.

The large corpus of empirical data and available tools for data collection and analysis is changing the ways knowledge is produced (Weinberger 2013; Meyer & Schroeder 2015). For the humanities, this transformation requires not only that we critically inquire into how technology affects our understanding of knowledge and how it alters our epistemic processes, but that we also employ the new data resources and technologies in new ways of scholarly investigation. Although data sets can provide new insights that offer opportunities for fine-grained detail previously not available, their possibilities are frequently overestimated (e.g. Anderson 2008; Schmidt &

Cohen 2014). Within academia, the blind trust in models, methods and data has been consistently criticized; recent big data enthusiasm has motivated a cohort of critical scholars to raise the alarm yet again (e.g. Couldry 2014; Gitelman 2013; boyd & Crawford 2011; Pasquale 2015). In this light, Rob Kitchin in *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (2014) identifies four fallacies sustaining big data empiricism:

1. Big Data can capture the whole of a domain and provide full resolution;
2. there is no need for a priori theory, models or hypotheses;
3. data can speak for themselves free of human bias or framing;
4. meaning transcends context or domain-specific knowledge. (133-137)

The unquestionable allure of new forms of empiricism makes it important for us to continue to acknowledge that humanities scholars' epistemological assumptions are different from those of their counterparts in the hard sciences. The purpose of humanities data research is not empirical validation and hypothesis testing, but the development of questions and the discovery of insights (Ramsay 2003: 173). Rather than import questions and methods from the hard sciences, we must develop our own approaches and sensitivities in working with data that will reflect the humanities' traditions.

The Humanities Scholar Revisited

There has been a tendency in academia to classify the emerging research practices of the digital humanities as a new specialism, a new field that can be neatly contained, whether within a department, an academic minor or research group. This position is troubling: computer-aided methods and data practices are not some new object like films or games once were. 'Digital humanities' is merely the *nom de guerre* of the computational turn in the humanities. Datafication and computerization will come to affect all research agendas and inform the skill sets of students and scholars alike. We predict that the term 'digital humanities' will sound increasingly pleonastic and will eventually disappear – it will lead not to the replacement of established methods in the humanities, but rather to an expansion in the curricula we study and the methods we use.

Widely associated with computational methods, the digital humanities have been criticized from within the humanities as a whole. In "The Dark Side of the Digital Humanities" (2014), Richard Grusin discusses the

tensions between scholars of the traditional humanities and those engaged in digital humanities, which have been based on the idea that the latter 'make things'. In this assertion he sees a devaluation of critique (or other modes of humanistic inquiry). The current focus on the digital humanities' ability to 'build' or to 'make' things rather than to critically comment on issues is misleading (not least because critique takes many forms, including making, building and application). At the core of this debate – though often it is not made explicit – is the question of how to approach the object of research. It has frequently been claimed that studying culture through data would necessitate thorough training in programming so as to allow researchers the wherewithal to 'look' into the 'black box' of the technology they are using, but such an emphasis misses the mark. Since the emerging algorithmic culture is characterized by the translation of rules and procedures into software, we need to develop an understanding of the mathematical concepts and models driving these programmes not in order to fully master them but rather to understand them sufficiently enough to approach new research objects from a critical perspective. We endorse Nick Montfort's understanding of programming as a means to develop intellectual capabilities that help us grasp how procedures of everyday life are translated into machine-readable language (Montfort 2016).

With regard to Grusin's call for humanities scholars to engage in critical inquiry, we are aware of the pressures brought on by academia's relentless corporatization and the overall neoliberal trend in society. Fields in the humanities are increasingly confronted with the demand that they justify their research activities. In such a climate it is tempting to employ computer-aided methods and quantitative analysis to feign a more 'scientific' appearance. Employing popular but meaningless terms such as 'big data' is symptomatic in this regard. However, the computational turn offers the humanities an incredibly important opportunity to study the contemporary transformation of society. We believe that access to large-scale empirical evidence and to analytic tools enables humanities scholars not only to describe this transformation empirically, but also to develop conceptual frameworks for understanding its impact. Understanding the 'digital humanities' as something instrumental, merely covering ways of doing research instead of conceiving of it as a more encompassing scholarly reaction to an increasingly datafied society, would simply mean following the folly of policymakers who believe that technological advancement will solve social problems.

As humanities scholars, we engage with 'the situated, partial, and constitutive character of knowledge production' (Drucker 2011: n.p.). The increasing translation of aspects of everyday life into machine-readable

information can be seen as yet another process of mediatisation that media scholars in particular are well suited to scrutinize. This does not mean that they can sit back and rely on their distinct methods and skills. On the contrary: the data revolution (Kitchin 2014) raises issues concerning research questions, methods and ethics. It calls for new literacies and the development of codes of conduct that make transparent the role of computational methods, that tackle ethical issues in data collection and sharing, and that address the role of humanities scholars in public debate and interdisciplinary cooperation.

To become 'experts' in data practices and simultaneously investigate how datafication affects institutions in our society, we need to work directly in the very fields where transformation through these practices manifests itself; whether as embedded researchers, activists or active participants applying scholarly expertise in the diverse contexts afforded by various social institutions. Here application and making are not only critical practices but also constitute learning acts. In response to the new demands of our times, we founded the Utrecht Data School at Utrecht University in the Netherlands. This research and education platform allows us to conduct data research with our students in areas that are rapidly changing through the advancing information processing technologies commissioned by companies, governments and non-profit organizations. But we are not a mere service provider: by initiating debates among stakeholders and policymakers, we can inform opinion-making processes and express critique as much in the application of data practices as in public debate. Unlike preaching from the pulpit of the academic lecture hall, our engagement in the field is risky. Nevertheless, we are convinced that the humanities have much to offer in societal debates about data through their profound understanding of cultural complexity and their critical inquiry into knowledge technologies.

Investigating the Datafied Society

This book is a collection of scholarly investigations into computer-aided methods and practices. While several contributors offer essays representing their skills, methods and exemplary research projects, others reflect on the sensibilities and competencies that scholars need to develop in order to study contemporary culture through data. This includes an expert understanding of the specific role of data analysis tools and data visualization in the process of knowledge production. In academic research, but also in many sectors of business and other areas of society at large, data

analysis unfolds via computer interfaces that display results that users often mistakenly regard as objective assessments. Such environments need knowledge workers who can grasp the processes of knowledge generation, from data collection through the various stages of analysis to visualization. These experts should be positioned to question the data sets as well as the mathematical models which determine the analysis.

In their historical investigation, Lorraine Daston and Peter Galison (2007) show that objectivity emerges as a symptom of epistemological fear: 'fear that the world is too labyrinthine to be threaded by reason; fear that the senses are too feeble and the intellect too frail; fear that memory fades, even between adjacent steps of a mathematical demonstration; fear that authority and convention is blind [...] Objectivity fears subjectivity, the core self' (373-74). The emergence of objectivity in scientific discourse also shaped a distinct self-understanding of the scientist. Delegating image creation to machines, the twentieth-century scientist became the expert reader of images whose 'trained judgement' afforded an accurate analysis and an alteration of the image for the depiction of patterns, the categorization of families of objects and so forth.[1] Daston and Galison note that contemporary scientific images are changing, both in quality and functionality. 'The image-as-tool seems to enter the scene inseparably from the creation of a new kind of scientific self – a hybrid figure, who very often works toward scientific goals, but with an attitude to the work that borrows a great deal from engineering, industrial application, and even artistic-aesthetic ambition' (2007: 413). One might add that over the past two decades of internet culture, open-source software development and online collaboration have also affected academic inquiry, which unfolds at the crossroads of universities and maker labs, hackathons or start-ups. Our current enthusiasm for computer-aided methods and data parallels the technology-induced crisis in representation and objectivity analysed by Daston and Galison. Their concerns must be taken into account in order to critically reflect upon the purported objectivity of computer-calculated results and visualizations.

The persuasive power of such claims to objectivity works on all levels of management and policymaking and requires that the scientific self be an eager advocate for critical inquiry into the working mechanisms of computer-aided and data-driven analysis. The media philosopher Vilém Flusser warned of the inscribed promise of scientific accuracy and objectivity in 'techno-images' (1997). The unbalanced enthusiasm for

---

1   See, for example, Daston and Galison (2007: 371).

data practices as processes for the development of accurate and 'neutral' (hence objective) results might prove just as problematic. As has been eloquently pointed out by David Gelernter, in discussing the uncritical reading of a map of potential ash distribution following the eruption of the Icelandic volcano Eyjafjallajökull in 2010, uncritical acceptance of computer-calculated results might have dangerous consequences: 'Firstly we'll be covered in an ash cloud of anti-knowledge and secondly a moral and intellectual passivity will emerge that won't doubt or argue against the images' (Gelernter 2010). This does not mean that we should reject those practices, but rather that we should employ them while being informed about their limitations, questioning their social impact and grasping their role in the epistemic process. The task ahead is to inform users, policymakers and the general public about the many factors that make up a data set, shape analysis and generate visualizations – and the many ways to read these digital analytics.

Although primarily directed at upper graduate students and researchers in the media studies, *The Datafied Society* is a useful collection of essays for anyone interested in studying culture during the era of the computational turn. The edited volume has been structured into four parts: (1) Studying Culture through Data; (2) Practices; (3) Concerns; and (4) Key Ideas in Big Data Research.

Part 1, 'Studying Culture through Data', covers different research methods. In her contribution, Eef Masson explores how two sets of epistemic traditions that used to be relevant to more or less distinct groups of scholars – hermeneutic and empirical – encounter each other in humanistic data research. She discusses how recent literature that reflects on the current state of the digital humanities tends to focus either on returning to the core, interpretative tasks of humanists or on bridging the two epistemic traditions. The tension between these traditions is also addressed in the contribution by Christian Gosvig Olesen, who argues that Cinemetrics, Cultural Analytics and ACTION, which quantify and visualize stylistic patterns in films and other cultural products, promote an inductive, exploratory form of analysis, and thus challenge the perception that cinemetric methodology is primarily a scientistic mode. These tools possess the potential to make cinemetrics more compelling to film scholars, who have been sceptical of its approach due to its association with a positivist epistemology.

The focus then shifts to Cultural Analytics, with Lev Manovich's proposed alternative to the distinct traditions carved out by Social Computing and Digital Humanities. Manovich successfully avoids taking sides between

the goals and methods of the humanities and of the sciences and instead explores how to toggle between the two disciplinary paradigms in order to pursue opportunities missed by both. He discusses the project he heads up, *On Broadway*, as an exemplary use of Cultural Analytics. The project uses data and images from various sources such as Twitter and Instagram posts, New York City taxi trips and data of economic indicators to create a novel view of city life.

The next two contributions focus on digital methods. First, Richard Rogers investigates the role of query design in digital methods, discussing how digital methods repurpose mediums and outputs for social (and medium) research. He explores how Google can be employed as an epistemological machine in research and discusses query design as a distinct analytical approach. Then, Natalia Sánchez-Querubín examines issue networks as part of the digital methods initiative to 'follow the medium'. She explores how streams of hashtags rather than hyperlinks can provide a way for redoing issue network analysis for social media.

Part 2 is dedicated to data practices in digital data analysis. It considers how researchers can engage with the datafied society. In his contribution, William Uricchio explores how algorithms paired with big data redefine long-held subject-object relations, raising important epistemological questions. He makes suggestions as to how the humanities agenda can be revised so that the new order's implications can be properly understood. In the following chapter, Bernhard Rieder and Theo Röhle discuss what they regard as the five main challenges of digital methods and explore the concept of the 'digital Bildung' (Berry 2011) as a means of facing these challenges. Taking up three examples, they demonstrate that the tools we have come to use mobilize a wide array of knowledge. The singular focus on code as a form of knowledge that is required when working with data distracts from considering what is actually coded. The 'content' of software is not code per se but rather a procedure *expressed* in code; and knowledge about these procedures is what needs to be developed.

The next contribution zooms in on the tools used in digital data analysis, as Johannes Paßmann and Asher Boersma consider an approach to algorithmic black boxes. They develop a concept of transparency that outlines the skills necessary for researchers to deal with the parts of the box that remain 'black' or opaque. Next, Cornelius Puschmann and Julian Ausserhofer show us different aspects of APIs from the perspective of social scientists using them for data collection. They describe the origin of APIs in software development, conduct a survey of popular Web APIs by type, and discuss issues with regard to the reliability, validity and representativeness of data

retrieved from APIs before offering speculation about future developments in this area. Tommaso Venturini et al. then reflect on a particular way of analysing data through the visualization of networks. They illustrate the narrative and storytelling potential of networks by examining the *Iliad's* network of characters, thus moving away from a discussion of the mathematical properties of networks to a reflection on how networks mediate and structure the phenomena they represent. Lastly, Karin van Es, Nicolás López Coombs and Thomas Boeschoten advocate for reflexive data analysis. They provide a series of questions about the various stages involved in doing digital data research that underscores how data and data visualizations are constructed by researchers and the tools they use.

Part 3 is dedicated to ethics, encompassing concerns ranging from moral issues that need to be tackled when embarking on research to reflections on how big data discriminates. Gerwin van Schie, Irene Westra and Mirko Tobias Schäfer discuss research ethics in light of their own experience in scraping patient data from an online platform. They reflect on the strained relationship between existing ethical guidelines and big data research, particularly in relation to the idea of informed consent. They propose a research structure that allows big data research to be conducted in an ethical manner. Taking a broader view, Annette Markham and Elizabeth Buchanan consider their previous work to provide a cohesive framework for assisting internet researchers, review boards, students and ethicists in ethically navigating the murky waters of internet research. Concluding this section, Koen Leurs and Tamara Shepherd explore the social biases of data sets and discuss the extent to which inequality, racism and prejudice are reflected in data sets.

Part 4, 'Key Ideas in Big Data Research', comprises a series of four short interviews exploring two topics; first, with Nick Couldry and Carolin Gerlitz on the challenges in researching the datafied society. Couldry specifically tackles the 'myth of big data' and Gerlitz the problems of making data points countable and comparable. With Evgeny Morozov and Mercedes Bunz, we then consider how algorithms affect everyday life. Morozov opposes the exceptionalism of algorithms; Bunz stresses the need to engage in dialogue with technology and to learn how to understand 'algorithmic thought'.

The chapters in this book can be read separately, but, taken together, they make a contribution that will stimulate and engage humanities scholars via their perspectives on debates and reflections on the theory and practices of digital data research. In addition to enhancing understanding of the field itself, they provide some hands-on guidelines to help direct research in an

ethical and transparent manner, promoting awareness of how researchers and their tools affect knowledge production.

References

Abbott, Andrew. 2000. "Reflections on the Future of Sociology." *Contemporary Sociology – Journal of Reviews.* 29 (2) (March): 296-300.

Berry, David M. 2012. "Introduction: Understanding the Digital Humanities." In *Understanding Digital Humanities*, ed. David M. Berry, 1-20. New York: Palgrave Macmillan.

boyd, danah, & Crawford, Kate. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, And Scholarly Phenomenon." *Information, Communication & Society* 15 (5): 662-679.

Braidotti, Rosi. 2013. *The Posthuman.* Cambridge: John Wiley & Sons.

Couldry, Nick. 2014. "Inaugural: A necessary disenchantment: myth, agency and injustice in a digital world." http://onlinelibrary.wiley.com/doi/10.1111/1467-954X.12158/abstract.

Daston, Lorraine & Peter Galison. 2007. *Objectivity.* Cambridge, MA: The MIT Press.

Dijck, José van. 2014. "Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology." *Surveillance & Society* 12 (2): 197-208.

Drucker, Johanna. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5 (1).

Flusser, Vilém. 1997. *Medienkultur.* Frankfurt a.M.: Fischer.

Gelernter, David. 2010. "Gefahren der Softwaregläubigkeit: Die Aschewolke aus Antiwissen." *Frankfurter Allgemeine Zeitung*, April 26. www.faz.net/aktuell/feuilleton/debatten/digitales-denken/gefahren-der- softwareglaeubigkeit-die-aschewolke-aus-antiwissen-1606375.html.

Gitelman, Lisa & Virginia Jackson. 2013. "Introduction." In *"Raw Data" Is an Oxymoron* ed. Lisa Gitelman, 1-14. Cambridge, MA: The MIT Press.

Grusin, Richard. "The Dark Side of Digital Humanities: Dispatches from Two Recent MLA Conventions." *Differences: A Journal of Feminist Cultural Studies* 25 (1): 79-92.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Infrastructures and Their Consequences.* London: SAGE.

Mayer-Schönberger, Viktor & Kenneth Cukier. 2013. *Big Data: A Revolution that Will Transform How We Live, Work, and Think.* London: John Murray Publishers.

Meyer, Eric T. & Ralph Schroeder. 2015. *Knowledge Machines: Digital Transformations of the Science and Humanities.* Cambridge, MA: The MIT Press.

Montfort, Nick. 2016. *Exploratory Programming for the Arts and Humanities.* Cambridge, MA: The MIT Press.

Pasquale, Frank. 2015. *The Black Box Society.* Cambridge, MA: Harvard University Press.

Ramsay, Stephen. 2003. "Toward an Algorithmic Criticism." *Literary and Linguistic Computing* 18 (2): 167-174.

Weinberger, David. 2014. *Too big to know: Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room.* Basic Books.