

Big Data Histories: An Introduction

In recent years, we have seen a trend towards ‘datafication’: the translation of all aspects of daily life – its features, actions and interactions – into machine-readable information or ‘data’.¹ Much like other scholarly disciplines, media studies has benefited from this development, and specifically, from the availability of ever-increasing numbers of (user) data, generated by corporations, public administrations and heritage institutions, that can be subjected to analysis. This reality is shaping how research is conducted and which questions are asked – not only in, but also about the field, and about the humanities more broadly.

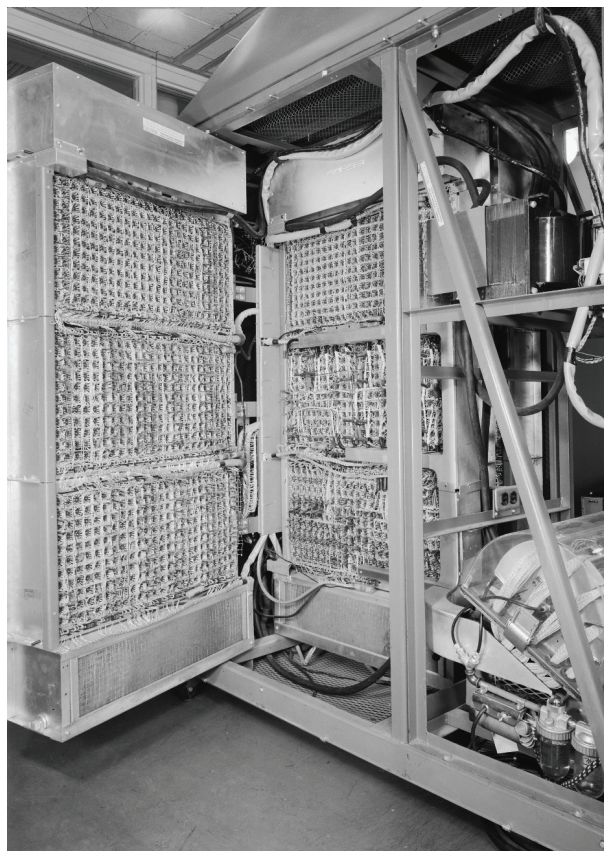
Others before us have pointed out that datafication, and the perceived ‘data deluge’ it has caused, is not a new phenomenon.² The same also applies to the collection, selection and analysis of so-called ‘big data’ (for the purposes of this issue: larger aggregations of information than a single, human researcher can oversee). The term ‘big data’ itself, of course, has been subject to critique – among others for the associations it elicits with some kind of a ‘revolution’.³ One reason for this is that using those terms carries the risk, precisely, that we conceive of data’s implications only in terms of newness and change, to the detriment of those continuities that are equally valuable in understanding (and successfully navigating) our datafied society. As one of us has argued elsewhere, developing a critical understanding of the current situation requires that we ‘debunk the exceptionalism inherent in the “big data” paradigm’.⁴

In an academic context, Danah Boyd and Kate Crawford argue, ‘big data’ is best defined not in terms of the quantities of data available, but rather in relation to our current capacity to search, aggregate and cross-reference large datasets – a capacity that rests on the interplay of technology (computer power, algorithmic accuracy), analysis (the ability to draw patterns from data that are economically, socially, or technically significant) *and* mythology (the belief that large datasets offer a ‘higher form of intelligence’).⁵ Even so, we need to be aware of historical precedents. One cannot deny of course that shifts have taken place in our field, much like in others; for instance, in terms of the scale of data-based scholarly endeavours, the extent of the automation involved, or the current need for interdisciplinary cooperation. But in order to make more convincing claims about both the nature and scope of such developments – and in the process, debunk undifferentiated accounts as to the novelty of either big data, or big data research – we need to carefully trace its historical lineages.

This volume is not the first to come to this conclusion. In early 2017, the history of science journal *Osiris* published a special issue on “Historicizing Big Data” that attempts to historicize modern data culture, exploring the relationship between technologies, practices and epistemologies. In their introduction, the editors point out that contrary to common belief, neither data,

nor digitization, nor even computing are specific to the present era, and that those processes were ‘performed in a variety of technological contexts, both pre- and post-electronic.’⁶ Comparing the notion of ‘big data’ to that of ‘big science’ (another alleged transformation in twentieth-century science and technology), they argue that the former is much more of a ‘moving target’, with more obvious roots in the past. Big data, for the authors, is merely a chapter in a longer history, or series of histories, ‘of observation, quantification, statistical methods, models and computing technologies.’ Those histories involved the development of strategies and technologies for dealing with information overload that in themselves ‘played a vital role in making knowledge.’⁷ In tracing them back in time, they argue, it is key to consider continuities as well as historically contingent differences or ruptures.⁸

This issue of *TMG – Journal for Media History* underwrites the position that there is a need for historicizing data and data practices, and continues the conversation on the subject. While there is some overlap with the *Osiris* issue also in terms of the specific topics tackled, the selection of “Big Data Histories” presented here focuses on questions that are particularly relevant to media scholars. On the one hand, it contains articles that take a media historical perspective on the phenomenon of datafication or on the analysis of data in different areas of cultural, social and economic life. In his programmatic article on how we should ‘do’ the history of big data,



IBM type 650 Magnetic Drum Data-Processing Machine (1957), the world's first mass-produced computer, marketed to the business, scientific and engineering communities. Collection: Fotocollectie Nederlandse Heidemaatschappij, Nationaal Archief.

David Beer has argued that attention is due not only to big data as a (material) phenomenon but also as a conceptual entity. The reason he quotes is that in many ways, the power dynamics of big data ‘are to be found just as much in the way that those data are labelled and described as (...) in the actual data themselves.’⁹ Inspired by Rob Kitchin, he argues that we therefore need to interrogate also its discourses and rationalities. The articles gathered here oftentimes do both: they engage with the discursive framings of big data as well as its materialities. On the other hand, the issue also contains some contributions that historicize, very specifically, practices of (big) data analysis by media scholars.

The peer-reviewed section the issue opens with consists of two parts. The first of those is entitled “Contemporary Big Data Practices in Historical Perspective”. In the opening contribution to this section, Niels Kerssens subverts persistent claims to the effect that big data these days are ‘revolutionizing’ managerial decision-making. To this end, he traces both the technological and the cultural origins of data-based management (specifically, in an American context) back to the 1970s and 1980s. His research is informed by close scrutiny of articles and adverts in the trade magazine *Datamation*, which are highly revealing of how the use of big data was ideologically framed at the time. In the next article, Markus Stauff shifts attention from the sphere of management to the realm of sport, considering its contribution to the emergence of a contemporary big data culture. He interrogates why and how media sports have become so entangled with big data, and how they have come to impact on its popularization, over time, as a cultural practice and cultural imaginary. A third contribution, by Frank Kessler and Mirko Tobias Schäfer, zooms in not so much on a specific *area* of data practice, but on a given *type* of data and data representations. In their article, Kessler and Schäfer place contemporary discourses on big data and trustworthiness in historical perspective, looking to the case of large-scale moving image archives of the nineteenth and twentieth centuries. These archives, it turns out, elicited very similar expectations among their audiences as today’s data, and specifically *images of data*: data visualizations or computer simulations. The final two articles in this section also deal with issues of data and bias, but specifically along the lines of gender, race, class and ability. Gerwin van Schie investigates the genesis of race-ethnic classification in the Dutch governmental data ontology. By looking closely at categories and their formulations in census reports of the last 120 years, he demonstrates how they are both socially, and socio-technically constructed. Rosa Wevers for her part interrogates the exclusionary aspects of biometrics. Taking the installation “Facial Weaponization Suite” (2011–2014) by the American artist Zach Blas as her case, she joins the maker in his critique of the now-dominant conception of biometrics as a ‘neutral’ set of practices. Unveiling its relations with a series of nineteenth-century pseudosciences, she points to its contested history, elaborating also on its implications for marginalized groups today.

Part 2 of the peer-review section, “Big Data in Media (History) Research: Developments and Historical Entanglements”, narrows the issue’s focus in an attempt to produce some exploratory genealogies of (big) data analysis in media scholarship, and specifically, media history research. First, Julia Noordegraaf, Kathleen Lotze and Jaap Boter review the usage, since its inception, of the online database Cinema Context – a key resource for many readers



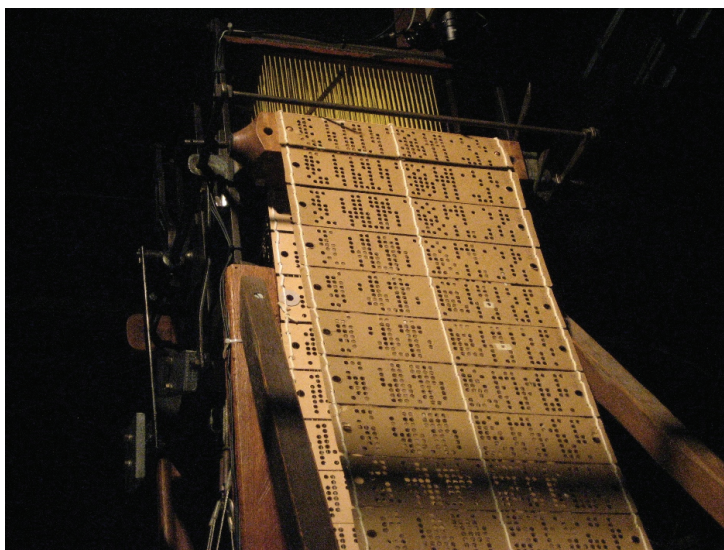
Data infrastructure at the Amsterdam location of the digital telephone exchange of PPT, the national mail and telephone service of the Netherlands (1985). Photo credit: Roland Gerrits; collection: fotocollectie Anefo, Nationaal Archief.

of this journal, and subject of discussion also in our previous issue on New Cinema History.¹⁰ In doing so, they reflect on the impact it has had on the study of historical film cultures, but they also identify its as yet unexploited affordances for computational analysis. Christian Gosvig Olesen and Ivan Kisjes close the section with a piece that likewise considers the analysis of film-related sources by computational means. Drawing on their experience with cinema distributor Jean Desmet's business archive, available in digital form in the CLARIAH Media Suite, they argue that current approaches to the study of such materials are largely text-centred – and as such, overlook their (material) complexities. In an effort to explain this trend, they trace the history of contemporary methodologies in New Cinema History to the *Annales* school's serial approach of the 1960s and 1970s. They end their contribution by suggesting that such methods be complemented with others; for instance, approaches involving tools for visual analysis.

The remainder of the issue is made up of a number of more informal pieces, in which contributors consider a broader and more varied set of relations between historical and contemporary data, data practice, and analysis of such practice. First, we present an interview section, divided in three parts. The first two contain records of shorter interviews, conducted by the issue's editors, in which four media and communications scholars discuss the data and

practices they research, or the methodologies they have developed for this purpose. In all cases, they take some time also to reflect on the historical antecedents or models for them, and how those continue to be relevant to their work today. The first section, entitled “Constructing Knowledge on Big Data: Methods in Historical Context”, features Alison Powell, developer of a methodology for data walking designed to generate knowledge about data in our everyday urban spaces, from the bottom up. In addition, Mirko Tobias Schäfer talks about his vision of ‘entrepreneurial research’, a form of participatory action research conducted at the Utrecht Data School. In the next section, “Issues in Big Data Practice: Histories and Historiographies,” Anne Helmond touches upon trends and issues in web historiography, and makes a plea for what she terms ‘website ecology’: an approach that involves websites’ various contexts and relations on the web in the writing of their histories. Subsequently, William Uricchio talks about what he perceives as a ‘colonization’ of the data imaginary by today’s public service institutions: an overly narrow conception of what data are, and of what they can do for their daily practice. He warns the reader of its possible consequences, and in the process, draws extensively on historical examples, identifying also the lessons they may teach us.

The issue concludes with two contributions – an interview and an exhibition review – that tackle more artistically-inspired investigations of big data and its histories. In the last interview section (called “Big Data Art, Present and Past”), we include the written record of a lengthier conversation, staged in 2017 in Utrecht, between new media artist Geert Mul and co-editor Eef Masson. In public communications, Mul refers to his work as ‘data-based art’ – with a term revealing both of his own process as a maker, and of his take on how contemporaries engage with, and make sense of, their daily realities. But it also provides a starting point for reflection on how his personal artistic practice ties in with much older traditions of play with data, and with various kinds of ‘rules’ for their recombination (linguistic, mathematical, or other). Finally, Maranke Wieringa contributes a review of the Algorithmic History Museum,



Punched card used in the Jacquard loom (early 1800s) – an inspiration for Charles Babbage, who used such cards for programme and data input into his Analytical Engine (first described in 1837). Collection: Ontario Science Centre; photo credit: John R. Southern.

an exhibition recently staged by media lab SETUP that challenges assumptions about algorithms' presumed neutrality by sketching fictional (and anachronistic) scenarios for the solution of historical problems with the help of such algorithms. Wieringa uses this as an opportunity to plead for a critical interrogation of the algorithms we live by, in terms of how they affect our daily lives today.

Notes

1. The term 'datafication', which has now become ubiquitous, was reportedly first used (in print) by Kenneth N. Cukier and Viktor Mayer-Schönberger, in "The Rise of Big Data: How It's Changing the Way We Think About the World," *Foreign Affairs* (May-June 2013), <https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data> (accessed 26 August 2017).
2. For instance, Ian Hacking, as referenced in David Beer, "How Should We Do the History of Big Data," *Big Data & Society* 3, vol. 1 (January-June 2016): 1–10, specifically 2.
3. See for instance Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences* (London: Sage, 2014).
4. Karin van Es and Mirko Tobias Schäfer, "Introduction: Brave New World," in *Studying Culture through Data*, ed. Mirko Tobias Schäfer and Karin van Es (Amsterdam: Amsterdam University Press, 2017), 13–22, specifically 13.
5. Danah Boyd and Kate Crawford, "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon," *Information, Communication & Society* 15, no. 5 (June 2012), 662–679, specifically 663.
6. Elena Aronova, Christine von Oertzen, and David Sepkoski, "Introduction: Historicizing Big Data," *Osiris*, no. 13 (2017), 1–17, specifically 2 (quote).
7. Aronova, Von Oertzen and Sepkoski, "Introduction," 3 and 6 (last two quotes).
8. *Ibidem*, e.g. 8.
9. Beer, "How Should We Do the History of Big Data," 2.
10. Clara Pafort-Overduin and Thunnis van Oort ed., "New Cinema History in the Low Countries and Beyond," *TMG – Journal for Media History* 21, no. 1 (2018).

Biographies

Karin van Es is an assistant professor of Media and Culture Studies at Utrecht University and coordinator of the Datafied Society research platform. Past publications in outlets such as *Television and New Media*, *Media, Culture and Society*, and *First Monday* focused among others on social TV, the concept liveness and the lure of objectivity in data visualization. Co-editor of the volume *The Datafied Society* (2017), she focuses in her current work on public service television and public values in the digital age.

Eef Masson is an assistant professor of Media Studies at the University of Amsterdam, where she teaches courses in film and media history and media archiving and preservation. She has published on non-fiction and non-theatrical films (among others, her book *Watch and Learn: Rhetorical Devices in Classroom Films after 1940*, 2012), media archives, museum media, and more recently, data visualization – specifically in artistic practice and media (history) research. Currently, she acts as senior researcher in UvA's The Sensory Moving Image Archive research project.