

## BOOK REVIEW

Anne Helmond, Mediastudies Department, University of Amsterdam.

**Lisa Gitelman (ed), *“Raw Data” Is an Oxymoron* (Cambridge: MIT Press, 2013), 208 pp., ISBN: 9780262518284 (pbk), £20.95.**

This edited collection is a timely and important intervention into the ‘Big Data’ hype by addressing the core of the Big Data discourse through its history. It does so by not only asking the important question of *what data are*, but also ‘[w]hat are we to data and data to us?’ (2013, p. 1). The authors contributing to this collection, edited by media historian Lisa Gitelman, place these fundamental questions into the larger history of data by addressing the distinct histories of data, data collection and data analysis from various disciplines. In doing so, they provide a historical lens on the data economy with the important – but often side-stepped – question of *what compromises the data* in ‘Big Data’ driving this new economy and *how is it constructed?*

In the introduction, Lisa Gitelman and Virginia Jackson remind us that data are not simply ‘there’, waiting for us to be collected, stored and analysed, but instead data are always imagined and created in particular historical, social, cultural and economical circumstances. This draws us to the main theme and goal of the book, which is to reflect on the suggestion of Bowker that data are never ‘raw’ but always ‘cooked’ (2013, p. 2). Precisely because data are constructed, Gitelman and Jackson argue that data ‘need to be understood as framed and framing, understood, that is, according to the uses to which they are and can be put’ (2013, p. 5). The process of this framing requires further investigations into the changing socio-economic contexts in which data are created and the heterogeneous social and cultural phenomena that are reduced to data. In other words, we cannot take the data in Big Data for granted.

In introducing the chapters in the collection, Gitelman and Jackson formulate three closely aligned general precepts of data: data are abstract, aggregative and mobilized graphically (2013, pp. 6–12). The first pair of chapters addresses the abstractive nature of data in distinct ways. In the opening chapter, Daniel Rosenberg provides a historical understanding of the concept of data to show how its connotation shifted over time as a way of understanding data as a rhetorical concept (2013, p. 36). In the second chapter, Travis D. Williams provides a reading of early modern mathematics as a cultural practice to illustrate how our interpretation of data should be situated and understood within ‘their original historical and cultural milieus’ (2013, p. 42). While data may be abstract, it is neither timeless nor contextless and for Gitelman and Jackson, these two attributes provide important points of entry for a critical analysis into the imagining of data. The third and fourth chapters

provide such a data contextualization by explaining the mutual shaping of disciplines and their objects over time (2013, p. 7). Kevin R. Brine and Mary Poovey study the treatment of data within the discipline of economy through a case study on Fischer's late nineteenth-century economic models. Matthew Stanley addresses the difficulty of working with historical data sets and the problems of reevaluating ancient data in the field of astronomy. These chapters also work through the second precept of the aggregative quality of data which give data 'their potential power, their rhetorical weight' (2013, p. 8). The third pair of papers that follows provide us with distinctive 'prehistories of databases' (2013, p. 9) by focusing on the materiality of early, paper-based data aggregation infrastructures. Ellen Gruber Garvey details the work of the Grimké sisters who collected runaway slavery advertisements by 'data mining' newspapers in the nineteenth century. Here, the aggregation of this data into a paper system enabled its analysis: 'It was the work of trimming, sifting, and aggregating the material that recreated it as a database and not just a collection of anecdotes. This work allowed for its recontextualization and analysis'. The sisters used the ads as rhetorical basis in their war against slavery where the power laid in the aggregation and recombination of the public data they compiled about slavery through ads. They published their work as *American Slavery As It Is* (1839) which became 'one in a multitude of projects that helped to create the modern concept of information, by isolating and recontextualizing data found in print'. The other chapter in this pair by Markus Krajewski is at times very dense, and details the paper slip box as an early type of database. The apparatus is described as a self-aware device which reconfigures the relation between the information system and its user by pointing to the agency of the device in the production of knowledge: 'Through its elements, prefabricated for connectivity, it always offers a configuration of potential states of knowledge, which are only realized, that is to say retrieved, by the user through certain combinations at a given time' (2013, p. 112). The final two chapters further elaborate on the role and agency of the invisible infrastructures used to capture, store and process data in our current era. Rita Raley looks at the online practices of dataveillance and counterveillance to argue that they are fundamentally entangled (2013, p. 139). She uses the example of the installation of plugins to opt-out of the data-mining practices of advertisers which at the same time 'make the browser more distinct and thus facilitate browser fingerprinting' (2013, p. 132). This leaves us with a disturbing scenario in which disabling the connections with trackers actually makes us *better* suitable subjects for tracking. The final chapter by David Ribes and Steven J. Jackson describes the various problems faced by scientists in making their data 'future proof' and how in this context 'data – long-term, comparable, and interoperable – become a sort of actor, shaping and reshaping the social worlds around them' (2013, p. 148). While the final chapter provides valuable insights into the active role of data in the process of its preservation, the chapter feels a bit disconnected from the previous chapter and the afterword that follows. In the afterword, Geoffrey C. Bowker returns to the central concept of 'raw data' that has been the central theme throughout the book.

This leaves us with the final precept as outlined in the introduction, namely that 'data are mobilized graphically' (2013, p. 12). Gitelman and Jackson argue that data visualization

‘amplifies the rhetorical function of data’ and that ‘recognizing the power of data visualization is an important part of living with data’ (2013, p. 12). However, none of the chapters directly engages with the rhetoric of data visualization and its framing effects. Fitting their preferred type of representation, we find that data visualization is represented in the colour plates that follow the introduction. However, the collection could have benefited from extending the critical reflections on data, as demonstrated throughout the chapters, into the visualizations of data by critically addressing their persuasive visual strategies.

With the increasing number of books published on the topic of Big Data, often heralding the potential value of big data as simply waiting to be uncovered, this collection has illustrated how this celebratory attitude is problematic. Rita Raley situates us ‘in the midst of what is exuberantly called a “Data Renaissance”, in which new marketing worlds await exploration and raw material – raw data – awaits extrapolation, circulation, and speculation’ where this raw data are the currency of the new data economy (2013, p. 123). However, as Raley and the other authors have argued throughout the collection, there is no such thing as raw data, since data are always cooked. Which insights does cooked data bring to Big Data analysis and what are the implications? Manovich (2012) and boyd and Crawford (2012) have previously opened up the agenda for such critical questions towards Big Data and what this collection offers is a critical reflection on how data are made by providing us with a much-needed historical lens.

### **Notes on contributor**

Anne Helmond is a PhD candidate with the Digital Methods Initiative, the New Media PhD program at the Department of Media Studies, University of Amsterdam. In her research, she focuses on software-engine relations in the blogosphere, cross-syndication politics in social media and data flows between web platforms. Her research interests include digital methods, software studies, platform studies, social media, algorithms, syndication protocols and web archives. She also teaches new media courses in the Media Studies department.

### **References**

- boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication and Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.
- Manovich, L. (2012). Trending: The promises and the challenges of big social data. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 460–475). Minneapolis: The University of Minnesota Press.