Nordic Journal of Library and Information Studies

Vol. 2, No. 1, 2021
ISSN (ONLINE) 2597-0593
DOI 10.7146/njlis.v2i1.127394

Herbjørn Andresen, Professor of Archival Science, Oslo Metropolitan University, Norway, herbj@oslomet.no

Book review:

Uncertain archives. Critical keywords for big data

Uncertain archives. Critical keywords for big data. (2021). Edited by Nanna Bonde Thylstrup, Daniela Agostihno, Annie Ring, Catherine D'Ignazio and Kristin Veel. The MIT Press, Cambridge, Massachusetts. ISBN: 9780262539883

Uncertain archives is an impressive book in many respects. A tempting place to start the review of this book is to refer to some numbers, as this book itself almost amounts some kind of 'big data'. The page count is 624. It contains 61 chapters, in addition to the introduction chapter, named 'Big Data as Uncertain Archives'. The book has 5 editors. With a few co-authored chapters, the number of authors is 73. There is a predominance of female authors, slightly more than two thirds. The book appears well-composed, well-crafted, and well-written, which is itself an achievement for a book of this size and complexity. However, the reasons why this book is important, and well worth reading, is not captured in these numbers. More importantly, the book provides rich, original, and valuable insights into fundamental issues in society, which at the same time run deep and are changing rapidly.

The book's subtitle includes the term 'keyword', which hints at the organization of its content. After its opening chapter, written by the editors, the succeeding chapter headings are keyword entries, presented in an alphabetical order. Some of these keyword terms will be familiar to most readers with a little knowledge of information science and Big Data discourse, such as *aggregation*, *metadata,* or *prediction*. Others may be more intriguing, such as *copynorms*, *hauntology,* or *remains*. Also, worth noting, the keywords that are discussed in this book do not include every household buzzword known from canonical big data literature. For instance, there are no keyword entries on the notorious 'three V's of big data', *volume*, *velocity,* or *variety*, which is nothing to miss. The 61 different keywords, a number too high to list individually in a book review, straddle different societal, cultural, and technological aspects of big data and algorithms. Most of the chapters, regardless of whether their

keyword headings at a first glance appear to be well-known and on-topic or not, are interesting, investigational, and often thought-provoking.

The encyclopaedia-mimicking order of the chapters is effective in several ways. First, it allows for many authors, and a variety of perspectives and argumentative styles. Second, it is inviting, as it softens the impression of a heavy brick and can be read in small chunks in any order the reader prefers. Third, and this may prove to be a clever move for a prolonged impact, it is a feature that helps the reader remember this book.

Each keyword/chapter is quite short, just under 10 pages on an average. The diversity of writing styles and academic directions is also a refreshing element. Most of the chapters avoid the tediousness that often characterizes academic prose and show an ability to convey knowledge and reasoning beyond a narrow clique. Nevertheless, they adhere to general academic virtues of stringency, transparency and referencing. Each chapter appears to be, and can be read as, a stand-alone contribution, thus resembling most academic anthologies. In terms of content, it is also more relevant to consider this book an anthology than an encyclopaedia or a loose collection of keywords. The different chapters carry perspectives and arguments that communicate with other chapters, to various degrees and on different levels.

The book offers different lines of arguments, that are more or less entangled. If (counterfactually) the editors had decided to organize this book as a linear anthology, grouping chapters into different thematic parts, there would be no single list of such 'parts' to emerge as the only logical way to organize it. The interconnections between the chapters are in a way kaleidoscopic. The introductory chapter includes a sub-heading named 'Thematic Flows and Constellations' that accounts for the complex relationships between the different chapters. The alphabetical order of the chapter headings (or keywords) has an effect of breaking away from the demands of a sequential line of argument. For instance, the keywords *proxies* and *stand-in* are related notions, but due to the alphabetical sorting they appear on different places in the book.

In order to provide a contracted account of what this book is about, some sort of grouping of subjects or thematic strands is necessary. This grouping is my point of view, it cannot be 'retrieved' in distinct parts of the book as the matter is organized.

One thematic strand is impacts of big data on individuals – on privacy, lifeworld and choices that are or are not available to us. Many keywords touch on aspects of this. On one hand, we are often willing to take part in datafication or quantification of activities to achieve something, yet the same sharing forms and perhaps even limits our options. We may contribute to data arsenals and algorithmic reach in a way that turns out to be disempowering. Another factor that could matter to individuals is a blurred or ambiguous distinction between human as a subject or as an object when interacting with 'leaf nodes' of data processing systems in the Internet of Things.

A second thematic strand is effects of big data on society. Some of the discussions concern inequality and bias as general aspects, others delve more specifically into feminist, gender, and intersectionality perspectives. The cultural field, including archival institutions, heritage, history and (digital) humanities, could also be considered as effects on society. Several chapters discuss poststructuralist ideas, with its variant of a constructionist view of archives, archives as structures that determine their possible content, and emphasizing the power exercised through appraisal, ordering and omissions. These poststructuralist perspectives may still appear somewhat foreign to the archival practitioners' field (at least from my vantage point), but this book makes a very probable case for their relevance.

Other strands – plural – concern technology, information analyses and data processing. These include both conventional computing and machine learning, knowledge organization and statistical concepts. There are very interesting discussions on predictions and predictability, uses of proxies, and the pervasive role of correlation (and, with it, 'the end of theory'). Particularly exciting in my opinion is the emphasis on a performative aspect of predictions. The distinction between predicting what will happen and making it happen is not necessarily always clear cut. Only a very few chapters are solely dedicated to technological or data processing issues. In most cases, these questions are intertwined with discussions of the social consequences of big data.

One last thematic strand, in my self-made list of themes for the purpose of this book review, is regulation and control. This theme is barely visible and occurs only in a few of the chapters. A notable exception is the keyword *ethics*, where the different strategies of either imposing ethical standards for algorithms to comply with, or of assuming ethical agency of algorithms, are compared. In some other chapters, small nods are sometimes given to needs for safeguarding basic human rights, privacy, and transparency, without actually discussing legislative options. The scarce discussions on legal regulation, control, and enforcement are somewhat surprising, considering that notions of justice and fairness are otherwise frequently invoked throughout the book.

As I have suggested, this brief account of what the book is *about* cannot do its rich content full justice. The strength of the arguments often lies in an unresolvable interplay between each chapter as a part and the book as a whole. There is, however, one recurring theme that deserves and requires special mentioning. It is in the book's main title, *Uncertain archives*. It is a notion that is discussed in many chapters, sometimes explicit and sometimes implied. The many approaches to discussing uncertainty indicate that it has probably been a common assignment from the editors to the authors. Anyhow, the discussions on uncertain archives are diverse, the editors seem to have held the writers on a long leash.

Uncertainty is, and has always been, integral to the archival endeavour. Archives are inherently incomplete, and they are both forming and formed by mandates and practices. The book's discussions of uncertain archives stand on firm ground of both classical and more emerging archival science viewpoints. For me, being an 'archival science guy', I must admit a sense of relief that the authors have not put the uncertainty of archives into an epochalist framing, which too easily leads to an ethos of defending one's turf and resisting change.

Still, uncertainty is an aspect that needs to be re-examined. Boundaries between what is fixed, and what is not, change. Data creators have less control, or even knowledge, of what collections they are contributing to, and how their data will be represented and interpreted. Archives, being structures that determine what they may hold, cannot expect to deal with self-contained collections from distinct record creators, modern collections hinge on multiple layers of platforms, services, known and unknown reinterpretations, and patterns of use. The introductory chapter offers an interesting perspective; controlling uncertainty has become an integral part of how modern society operates. Instead of being a unilateral signal of danger and risks to be avoided, uncertainties are a precondition for change, and is used for navigating dilemmas of security and freedom. Uncertainty raises the stakes in different directions at the same time. There is a rising risk of unreliable and even forged information, both on a detailed level and in the overarching narratives. On the other hand, uncertainty may also provide ways to 'restore to marginalized individuals and groups the ability to contest dominant organizations of information', to quote the more optimistic prospect that the introductory chapter concludes its discussion of uncertainty with. Dealing with uncertainties are in part practical matters, but it can also serve as a benign-ish redistribution of the cards different groups hold.

As stated earlier, my general view of this book is that it is well-composed and well-written. It is held up to a high academic standard. However, there are a couple of weak points, or maybe more aptly blind spots, that could be pointed out. The first one is that many of the chapters appear to discuss against an implied adversary or antagonist, be it tech firms, platforms, the market, surveillance capitalism, or governments. It is not a problem in the sense that it makes arguments and positions more difficult to comprehend, but it can to some extent undermine the general ambition of a more dualistic and open-textured approach to uncertainty. The second point is that some of the chapters mention recent events, for instance 'Christchurch' or 'Ferguson', without enough context to make sure it will still be understandable what they refer to, and why they are mentioned, a few years on. It must be emphasized that this only applies to a few chapters, in most cases events that are mentioned are sufficiently contextualized not to require fresh memory.

Uncertain archives is an impressive book, both in the sense that it is an extensive and well executed piece of work, and in the sense that it makes an impression on the reader. It is an impression that this is important to all of us, every discipline and every profession need some understanding of, and to raise critical questions towards, big data, datafication and analysis and their many related themes. The book is rich in ideas and perspectives, some tightly and some loosely linked. The authors provide profound knowledge and insights, and to various degrees argumentative positions, from different disciplines. The interdisciplinarity will ensure that virtually every reader learns something new and feels the joy (or sometimes agony) of new trains of thought being sparked. The book is also a pleasant read, in part due to its accessible and high-quality prose, but maybe even more due to its general tone. It is sincere, but not alarmist, nor is it evasive or fatalistic. This is science at work; identifying a problem field, setting an agenda, and examining it in a methodical way.

Who will, and who should, read this book? It is tempting to declare the book to be of interest for a general public. The problems that are discussed intervene in the lives of all of us, they are not only related to our academic interests. However, the complex interplay between parts and whole probably requires some academic training to understand and appreciate. For undergraduate archival students, I could easily find some gems among the chapters to be included in the literature of some of our subjects, but it could hardly serve as a complete textbook on a subject of archives and big data 'from a to z', it would be too complex for the purpose of acquiring skills. I will conclude the main target group is people like me, a never finished learning academic. So much of the reading of academic literature is devoted to the meticulous knowledge that is built stone upon stone within a narrow field. To move further on, it is sometimes really valuable to read a book that helps widen the perspectives. Perhaps it could be classified as academic self-help literature.