BETWEEN ONLINE AND OFFLINE: DOING ARCHIVAL RESEARCH IN THE DIGITAL AGE

by Gerben Zaagsma

INTRODUCTION

Changes brought about by archival digitisation affect archivists and historians alike. As Dutch historian and archivist Charles Jeurgens noted a decade ago, the history and archival science communities would do well to debate these together "because digitisation is fundamentally changing the relationship between the archive, the archivist and the researcher".[1] In this short contribution, I would therefore like to discuss how digitisation changes and (re)shapes the work of historians in the hope of providing a user perspective to current discussions about digitisation.[2]

First, it is important to historicise and qualify what is new regarding digitisation, which poses specific challenges but is only the latest in a long line of technologies of preservation and reproduction. Since the introduction of photography in archives and libraries in the late 19th century, the photostat, microfilm, and digitisation have all been deployed by heritage institutions in efforts to preserve and reproduce their materials.[3] The first "digital" archives were the social science data archives of the 1960s (kept on punched cards and magnetic tape), which became important sources for political scientists and historians alike.[4] When digitisation took off in the 1990s, it initially centred on capturing the information contained within archival materials through the construction of historical databases and the creation of text-based digital editions. Digital facsimile reproduction as we know it today, as well as mass digitisation, were mostly a post-2000s phenomenon.

The idea of digitisation as a preservation technique in its own right is also relatively new; early digitisation efforts were predominantly about access. As structured data now gave way to the exponential growth of unstructured data, new challenges of how to work with the information embedded in historical materials arose. In the past few years, more and more archives have engaged the question of how their digital collection practices can move beyond simply providing access to adopt a "collections as data" approach, which renders them open to computation.[5] It is here that the complementary research potential of digital archives truly comes to the fore. The field of computational archival science seeks to address some of the archival data curation challenges concerned at scale.[6]

ACCESSIBILITY AND THE POLITICS OF DIGITISATION

Digitisation has greatly increased accessibility, allowing for the reconstruction of hitherto dispersed collections, and enabled new comparative and transnational histories to be told, which, not so long ago, would have necessitated research trips to multiple archives. Scale notwithstanding, this change in modalities of access to archival materials has its historical precedents, too. Karl Krumbacher's Die Photographie im Dienste der Geisteswissenschaften, published in 1906, lauded photography's potential for historical research and was read across the Atlantic.[7] The work of American-Jewish historian Samuel Oppenheim on Jews in the Americas in the 1920s was partly enabled by his use of photostat copies of archival documents from the Netherlands.[8] Microfilm allowed for new ways of accessing increasing amounts of materials, and its use became especially ubiquitous after WWII. In fact, many historians have consulted original archival materials as well as microfilm for decades (and sometimes still do). Today, they often combine original with digitised archival documents, consulted in the reading room or online. In many cases, this enables the use of materials that would otherwise be out of reach for financial or other reasons, thereby expanding the research possibilities.

At the same time, digitisation has limits; even if many materials, especially books and newspapers, are digitised, many archives are not and will never be, as the results of the eNumerate Core Survey 4 from 2017 indicated.[9] That should not come as a surprise; long-term preservation through digitisation involves significant extra costs on top of the costs that analogue preservation already incurs.[10] Moreover, digitisation itself is not the only point of discussion. As the eNumerate survey also showed, only an estimated 58% of archives in Europe have their descriptive metadata published online. Caveats notwithstanding (the survey is voluntary), this suggests that more than 40% of (European) archival cultural heritage cannot be discovered online through institutional collection databases. This raises serious questions about whether to prioritise cataloguing or digitisation. Small surprise then,

that Jeurgens made a strong plea for digitizing archival access tools (inventories, catalogues, etc.), which enables the interlinking of relevant collections and, crucially, information about offline collections with those online. A recent, excellent example of this is the Dutch National Archive's collation of slavery-related digitized sources from institutions in the Netherlands, England, Guyana, and Suriname.[11] However, as Jeurgens also warned, the digitisation of access tools requires great care as "changes in the inventory have irrevocable consequences for how the contents of the archive are viewed".[12] The fact that archival digitisation is always only partial and that retro-digitisation is always a selection of an already existing selection raises questions about what is digitised and how selection processes ultimately shape historical research. Selection criteria for digitisation include the need to preserve fragile materials, providing easy access to collection highlights and frequently used materials, the research value of specific collections and academic research agendas. Memory politics, public discourses on the past, and the articulation of a country's imagined national identity are of similar importance, while legal, ethical, and copyright issues also frame and constrain digitisation strategies. All of these criteria are reflected in funding policies. In addition to the role these factors can play in the preservation efforts of official actors, bottom-up initiatives such as community archiving have their own agendas and can be animated by different concerns. To be sure, the basic questions of why, where, and how we can access what we can access, and which histories can (and cannot) be told with them, and by whom, have not changed after the digital turn. Indeed, archives have always enabled specific (re-)constructions and visions of the past and, as such, are powerful actors and potential gatekeepers in the production of historical knowledge. Yet, as historians increasingly make use of digital resources in their research, they have become ever more urgent. The politics of digitisation thus must be seen in the broader context of the politics of heritage and its preservation and involve questions about who digitizes what and why, what is metadated and OCR-/HTR-ed, how materials are classified and metadated, and how access is mediated.[13] An increasing number of institutions has started to provide (parts of) this information, a development that will hopefully see wider adoption across the GLAM sector, as understanding how digital resources are constituted is crucial for the historian's critical assessment.

WORKING IN AND WITH THE ANALOGUE AND DIGITAL ARCHIVE

Apart from questions about the state and politics of digitisation, of what is accessible and why, we can observe differences and changes in working with analogue and digital archives in practical, epistemological and methodological terms. The practice of archival research has changed fundamentally since Arlette Farge highlighted (and romanticised) what she famously dubbed le goût de l'archive.[14] 'Analogue' work in the reading room enables us to attend to the physicality of the archive, the smells, colours, look and feel of materials and a close reading in which the historian's imagination and analytical skills are brought to bear upon the original material/evidence. Today, however, archival work frequently involves photographing as many documents as possible to study them in more depth at home or in the office instead of the archival reading room. Meanwhile, 'digital' online work excludes the reading room altogether while introducing new materialities to be negotiated, thereby fundamentally changing, but not dispensing with, the 'taste' of the archive.[15]

Working with digital facsimiles of archival originals also influences the historian's interpretative praxis and thus has epistemological consequences.[16] Digitisation entails an "ontological transformation"[17] and the creation of a "new informational object"[18], altering notions of authenticity and

1 Charles Jeurgens: The Scent of the Digital Archive: Dilemmas with Archive Digitisation. In: BMGN – Low Countries Historical Review 128/4 (2013), S. 30–54. DOI: https://doi.org/10.18352/bmgn-lchr.9348

2 Various aspects discussed in this article are based upon and elaborated in more detail in: Gerben Zaagsma: Digital History and the Politics of Digitization. In: Digital Scholarship in the Humanities, 38/2 (2023), S. 830–851. DOI: https://doi.org/10.1093/llc/fqac050

3 For an important overview, see: Markus Friedrich: Vom Exzerpt zum Photoauftrag zur Datenbank. Technische Rahmenbedingungen Historiographischer Forschung in Archiven und Bibliotheken und ihr Wandel seit dem 19. Jahrhundert. In: Historische Anthropologie 22/4 (2014), S. 278–298. For a contemporary account see, for example: Otto Mente und Adolf Warschauer: Die Anwendung der Photographie für die archivalische Praxis. Leipzig 1909.

4 See, for instance: Ralph Bisco: Social Science Data Archives: A Review of Developments. In: The American Political Science Review 60/1 (1966), S. 93–109

5 Thomas Padilla et al. Always Already Computational: Collections as Data. Final Report. 2019. URL: https://collectionsasdata.github.io/ (aufgerufen am 05.12.2023).

6 Mark Hedges, Richard Marciano and Eirini Goudarouli: Introduction to the Special Issue on Computational Archival Science. In: Journal on Computing and Cultural Heritage 15/1 (2022). DOI: https://doi.org/10.1145/3495004

7 Karl Krumbacher: Die Photographie im Dienste der Geisteswissenschaften. Leipzig 1906.

8 Herbert Bloom: A Study of Brazilian Jewish History, 1623–1654, Based Chiefly Upon the Findings of the Late Samuel Oppenheim. In: Publications of the American Jewish Historical Society 33 (1934) S. 43–125. URL: https://www.jstor.org/stable/43058414

9 See: https://pro.europeana.eu/page/enumerate (zuletzt aufgerufen am 03.12.2023).

10 Jeurgens (Anm. 1), S. 46; Jakob Frohmann et al. Zum Verhältnis von Originalerhalt und Digitalisierung von Schriftlichem Kulturgut. In: ABI Technik 43/2 (2023), S. 103–109. DOI: https://doi.org/10.1515/abitech-2023-0018

11 See: https://www.nationaalarchief.nl/en/slavery (zuletzt aufgerufen am 03.12.2023).

12 Jeurgens (Anm. 1), S. 35–37.

13 These questions emanate from a model I proposed elsewhere to analyse the process of digitisation and its political dimensions. See: Zaagsma (Anm. 2), especially S. 837–844. This model, in turn, is based upon an elaboration of the political dimensions of archival and curatorial work as outlined in: Richard Harvey Brown and Beth Davis-Brown: The Making of Memory: The Politics of Archives, Libraries and Museums in the Construction of National Consciousness. In: History of the Human Sciences 11/4 (1998), S. 17–32. DOI: https://doi.org/10.1177/095269519801100402

14 Translated in English as the allure of the archives. See: Alette Farge: Le goût de l'archive. Paris 1989; Alette Farge. The Allure of the Archives. New Haven 2013.

15 See the project Le Goût de l'Archive à l'Ère Numérique by Frédéric Clavert and Caroline Muller: https://gout-numerique.net/ (zuletzt aufgerufen am 05.05.2023).

16 Gerben Zaagsma: On Digital History. In: BMGN – Low Countries Historical Review 128:4 (2013) S. 30–54, S. 25–6; Clavert and Muller (Anm. 15); Mareike König: Der Geschmack des digitalen Archivs zuhause auf dem Sofa #goutnum. URL: https://dhdhi.hypotheses.org/5552 (zuletzt aufgerufen am 03.12.2023).

17 Andreas Fickers: Authenticity: Historical Data Integrity and the Layered Materiality of Digital Objects. In: Digital Roots: Historicizing Media and Communication Concepts of the Digital Age. Hg. von Gabriele Balbi, Nelson Ribeiro, Valérie Schafer und Christian Schwarzenegger. Berlin 2021, S. 299–312, S. 305.

definitions of the original. Different materialities influence our reconstructions of the past as analogue and digital forms of analysis help shape the inferential process. Moreover, as Alan Munslow has suggested, "the historian normally and regularly crosses the line between inference and imagination", the latter being defined as "the application of the general capacity of the human mind for comparison, connection, analogy and difference to the study of the past and its sources".[19] The question then becomes how inference and imagination differ in analogue and digital contexts and how this, in turn, reshapes historical interpretation.

Finally, the digital archive allows for new methodological approaches. It enables historians to read and interpret their sources, converted into data, along two axes – close/distant reading and human/computational reading, whether their data is big or small (it is a common misconception that 'going digital' automatically entails working with big data and distant reading).[20] Data conversion does not only entail OCR or HTR, or techniques such as computer vision for visual elements but can include further extraction tasks such as named entity recognition or various natural language processing techniques. Many analytical options become available once historical information is made available for computation. These range from methods of text analysis, topic modelling, network analysis, GIS and spatial modelling to various visualization methods employed in research projects with a quantitative or qualitative bent or both.

All of these methods, it has to be noted, have long historical antecedents that have been investigated to only a limited extent.[21] Ever since mainframe computers entered universities, and especially since the advent of micro- and personal computing around the turn of the 1980s, historians have sought to harness computing power in their research for a variety of (analytical) tasks. One only needs to look at the various proceedings of the Association for History and Computing (AHC, 1987–2005) or the 40 volumes of the Halbgraue Reihe für historische Fachinformatik, published under the auspices of the former Max-Planck-Institut für Geschichte in Göttingen between 1989–1996. In that sense, we witness today an expansion, accelerating rapidly because of recent advances in AI and machine learning, of decades of computer-assisted and -enabled work building upon a steadily increasing amount of digitized materials and historical data.

CONCLUDING REMARKS

From gathering materials to processing the information contained within them to analysing them and disseminating the results, historical research practices have fundamentally changed as a result of the digital turn. As I argued a decade ago, hybridity has become the new normal for historians, most of whom combine traditional/analogue and new/digital materials and practices.[22] This hybridity of current archival research, in the reading room as elsewhere, shows that the analogue-digital dichotomy, in terms of historical research practice, is artificial. Much more productive than engaging in either/or debates, we would do well to think in terms of complementarity. That observation extends to digitisation. As Jeurgens has noted, "Digitised archives should complement, rather than replace, analogue collections".

For many historians, digitisation is about much more than easy access as it can open up new research possibilities. At the same time, however, the conundrum that archives face with regard to choices about digitisation is not always well understood by historians who too often look at archives and archivists as service providers instead of partners in the co-construction and -production of historical knowledge. As the era of mass digitisation is partly over[23], and small-scale as well as on-demand digitisation become more prominent, challenges facing the analogue archive should be higher on the agenda, no matter how paradoxical that might sound in the era of digital history. Instead of suggesting digitisation as a miracle cure for supposed problems of accessibility, online cataloguing to make visible what is sometimes misleadingly described as 'hidden' heritage seems an increasingly crucial task that concerns archivists and historians alike. It is to be hoped, then, that the debate Jeurgens already wished for will continue as we seek to preserve our cultural heritage for future generations.

ZWISCHEN ONLINE UND OFFLINE: ARCHIVFORSCHUNG IM DIGITALEN ZEITALTER

*Der Beitrag versucht, den Einsatz neuer Technologien in der historischen Forschung aus der Sicht der/s Historiker*in zu betrachten und die Unterschiede zwischen der Arbeit mit Online- und traditionellen Archiven von diesem Standpunkt aus zu reflektieren. Dabei wird erörtert, welche verschiedenen Faktoren zu berücksichtigen sind, welche politischen Dimensionen in der archivarischen und kuratorischen Arbeit bestehen und wie sie sich in einem digitalen Kontext verändern.*

Dr. Gerben Zaagsma
Université du Luxembourg
Luxembourg Centre for Contemporary and Digital History (C²DH)
11 Porte des Sciences, 4366 Esch-Belval Esch-sur-Alzette, Luxemburg
E-Mail: gerben.zaagsma@uni.lu

18 Jeurgens (Anm. 1), S. 34.
19 Alan Munslow. The Routledge Companion to Historical Studies. London 2006, S. 135.
20 Frédéric Clavert: Lecture des sources historiennes à l'ère numérique. 2012. URL: https://histnum.hypotheses.org/1061 (zuletzt aufgerufen am 03.05.2023).
21 A seminal special issue of the German historical journal zeitenblicke in 2011 was dedicated to „Historizität, Materialität und Narrativität. Zum Zusammenhang von Technikkultur und Historiographiegeschichte". See the entire issue, edited by Armin Heinen, here: https://www.zeitenblicke.de/2011/1/ (zuletzt aufgerufen am 05.05.2023). The author is currently working on a new book project about the history and genealogies of digital history, set within the broader context of how technology has shaped historical research practices and knowledge production since the late 19th century.
22 Zaagsma (Anm. 16), S. 17.
23 Adam Crymble: Technology and the Historian. Transformations in the Digital Age. Urbana 2021, S. 54–87. The exception to this are commercial mass digitisation projects by companies such as Ancestry.com whose activities raise serious questions about ownership of cultural heritage. See also: Jerome de Groot: Ancestry.com and the evolving nature of historical information companies. In: The Public Historian 42/1 (2020), S. 8–28.