

Introduction

Like Hamlet, historians have now reached the crossroads of “to be or not to be;” either they accept the challenge and attain to new heights of achievement or else reject it and be swamped by the tidal wave of accumulated and expanding knowledge as was the art savant in “Penguin Island.”¹

[I]f we do not wake up soon to the new realities of big data, computer scientists will leave us behind, biting the dust in this road to knowledge.²

The Urgency of Digital History

As long as there have been computers, there have been scholars pulling at historians, challenging them to use these computers for historical research. Historians need to adapt to the new technological possibilities, otherwise they risk becoming irrelevant. The reason for such calls is not just to become a modern profession, but because historians are supposedly faced with ever increasing amounts of sources, big data, or even knowledge. These challenges are then argued to require computational approaches. Such calls have largely gone unanswered, as can be seen from the two quotes above, which are 70 years apart, yet pose a similar challenge to historians. Already in 1948, the historian Murray G. Lawson stated that “historians have not been sufficiently conscious of the benefits to be derived from the technological revolution which has transformed contemporary society.”³ In 1968, the historian Emanuel Le Roy Ladurie made the (in)famous statement that “the historian will be a programmer or he will be nothing”.⁴ In 1990, the historians Onno Boonstra, Leen Breure and Peter Doorn wrote that “[t]he historian who refuses to use a computer as being unnecessary, ignores vast areas of historical research and will not be taken serious anymore”.⁵

1 Murray G. Lawson, “The Machine Age in Historical Research,” *American Archivist* 11, no. 2 (1948): 149.

2 Roberto Franzosi, “A Third Road to the Past? Historical Scholarship in the Age of Big Data,” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 50, no. 4 (2017): 14, <https://doi.org/10.1080/01615440.2017.1361879>.

3 Lawson, “The Machine Age in Historical Research,” 142.

4 Quoted in Lawrence Stone, “The Revival of Narrative: Reflections on a New Old History,” *Past & Present* 85, no. 85 (1979): 13.

5 Onno Boonstra, Leen Breure and Peter Doorn, “Past, Present and Future of Historical Information Science,” *Historical Social Research / Historische Sozialforschung* 29, no. 2 (2004): 4.

Fourteen years later, they were disappointed, and although a group of enthusiasts in history had formed, computational methods had far from diffused in the historical profession.

This lack of diffusion was partially due to critical responses from historians. Especially the first wave of computational methods in history, consisting of quantitative analyses, was criticised since “almost all important questions are important precisely because they are not susceptible to quantitative answers”.⁶ Yet turning to qualitative data did not lead to much more enthusiasm. The historian Peter Denley instead noted that “we have sacrificed at the altar of the microchip the thirteenth century fief rolls of Champagne and the fifteenth century baptismal records of Pisa, the naturalisation lists of fourteenth century Freiburg in Switzerland and tenth century Cluniac charters”.⁷ Yet more important than such critical exchanges was the lack of engagement; historians interested in computational approaches simply failed to convince their peers.⁸

This is not to say that historians have entirely missed the so-called “digital turn”. Every historian nowadays has a computer on their desk, writes their monograph in word processing software and searches for information on Google or some specific online database.⁹ A renewed interest in digital methods has arisen now that libraries and archives are increasingly publishing sources in online databases. Vast quantities of sources have been digitised in the large-scale digitisation projects of the past decades. Yet although historians and archives are highly interdependent, historians have largely remained silent about questions regarding the consequences of digitisation.¹⁰ The digitisation of sources

6 Arthur Schlesinger, “The Humanist Looks at Empirical Social Research,” *American Sociological Review* 27, no. 6 (1962): 770, <https://doi.org/10.2307/2090404>, cited in Stephan Thernstrom, “The Historian and the Computer,” in *Computers in Humanistic Research: Readings and Perspectives*, ed. Edmund A. Bowles (1967), 73–81.

7 Quoted in William A. Speck, “History and Computing: Some Reflections on the Achievements of the Past Decade,” *History and Computing* 6, no. 1 (1994): 30.

8 Boonstra, Breure and Doorn, “Past, Present and Future of Historical Information Science,” 85–59; Speck, “History and Computing.”

9 Max Kemman, Martijn Kleppe and Stef Scagliola, “Just Google It,” in *Proceedings of the Digital Humanities Congress 2012*, ed. Clare Mills, Michael Pidd and Esther Ward (Sheffield, UK: HRI Online Publications, 2014).

10 Ian G. Anderson, “Are You Being Served? Historians and the Search for Primary Sources,” *Archivaria* 58 (2004): 81–129; Andreas Fickers, “Veins Filled with the Diluted Sap of Rationality: A Critical Reply to Rens Bod,” *BMGN – Low Countries Historical Review* 128, no. 4 (2013): 155–63; Andreas Fickers, “Update Für Die Hermeneutik. Geschichtswissenschaft Auf Dem Weg Zur Digitalen Forensik?,” *Zeithistorische Forschungen – Studies in Contemporary History* 17, no. 1 (2020): 157–68; Tim Hitchcock, “Confronting the Digital: Or How Academic History Writing Lost the Plot,” *Cultural and Social History* 10, no. 1 (2013): 9–23,

and workflows and the introduction of search engines are often thought of as practical revolutions, while the effect on research is treated as a secondary by-product.¹¹ The speedup of archival exploration is perceived as an advantage, mainly because it leaves more time for close reading.¹²

Such interpretations treat the digital form as an equivalent surrogate to the original source, merely more accessible. Computers are, however, envisioned to allow much more comprehensive interaction with the historical material. Historians that subscribe to this vision have gathered under the signifier of “digital history”. They experiment with tools, concepts and methods from other disciplines, mostly computer science and computational linguistics, to benefit the discipline of history, constituting methodological interdisciplinarity.¹³

In digital history, therefore, historians collaborate with computational experts to try and adjust tools and methods from other disciplines to fit the needs of historians. The ambition is that, at some point, such digital methods might eventually diffuse to the broader field of history and be adopted by historians who do not collaborate with computational experts. These cross-disciplinary interactions are what interests me in this book. Historians in digital history try to innovate historical research in a way that is methodologically and epistemologically acceptable to the values and norms of their discipline.¹⁴ At the same time, computational experts are interested in what is computationally feasible when confronted with the heterogeneous, imperfect and incomprehensive collections that historians have been working with for centuries. Computational methods are not yet adapted to such issues, and how to extract valuable information from historical datasets is a matter of active research.

Digital history thus creates uncertainty for both sides; historians are uncertain how they as historians should use digital methods, and computational experts are

<https://doi.org/10.2752/147800413X13515292098070>; Frank M. Bischoff and Kiran Klaus Patel, “Was Auf Dem Spiel Steht. Über Den Preis Des Schweigens Zwischen Geschichtswissenschaft Und Archiven Im Digitalen Zeitalter,” *Zeithistorische Forschungen – Studies in Contemporary History* 17, no. 1 (2020): 145–56.

11 Bob Nicholson, “The Digital Turn,” *Media History* 19, no. 1 (2013): 59–73, <https://doi.org/10.1080/13688804.2012.752963>.

12 Adrian Bingham, “The Digitization of Newspaper Archives: Opportunities and Challenges for Historians,” *Twentieth Century British History* 21, no. 2 (2010): 225–31, <https://doi.org/10.1093/tcbh/hwq007>.

13 Julie Thompson Klein, *Interdisciplining Digital Humanities: Boundary Work in an Emerging Field*, online (University of Michigan Press, 2014), <https://doi.org/10.3998/dh.12869322.0001.001>.

14 Wolfgang Kaltenbrunner, “Reflexive Inertia: Reinventing Scholarship through Digital Practices” (PhD thesis, Leiden University, 2015).

uncertain how digital methods should work with historical datasets. The opportunity that arises from this uncertainty is that historians and computational experts need to negotiate the methods and concepts under development. Historians need to adapt their practices to what is computationally feasible, but the methods that are being developed need to be adapted to what is of interest to historians. How historians can influence the development of digital methods, and how digital methods affect the methodology and epistemology of the historical discipline, has thus far been underexplored.¹⁵ In this book, I explore these issues by following digital history scholars and understanding their practices, responding to the call to action from the information scientist Christine Borgman.¹⁶ As such, this book is inspired by the well-known social studies of science, applied to digital history in practice.¹⁷ Through a mixed-methods, multi-sited ethnographic approach, I provide a critical view on digital history grounded in how it is conducted and negotiated.

To support this analysis, I develop a model to analyse digital history collaborations as trading zones. This concept was developed by the historian of science Peter Galison to describe how two communities with vastly different practices and discourses can interact and negotiate a joint enterprise. He defined a trading zone as “an arena in which radically different activities could be *locally*, but not globally, coordinated.”¹⁸ While historians and computational experts in general employ different discourses and practices, and publish in different formats and venues, locally it is possible to coordinate practices toward a shared objective. Through such coordination a trading zone emerges which I analyse according to three dimensions. First, engagement as the extent to which the two communities come together to meet and interact. That is, a trading zone where historians and computational experts share an office and meet daily is different from one where communication is done per email once a month. Second, power relations as the extent to which

15 Hinke Piersma and Kees Ribbens, “Digital Historical Research: Context, Concepts and the Need for Reflection,” *BMGN – Low Countries Historical Review* 128, no. 4 (2013): 78–102; Bernhard Rieder and Theo Röhle, “Digital Methods: Five Challenges,” in *Understanding Digital Humanities*, ed. David Berry (Palgrave Macmillan, 2012), 67–84; Gerben Zaagsma, “On Digital History,” *BMGN – Low Countries Historical Review* 128, no. 4 (2013): 3–29, <https://doi.org/10.18352/bmgn-lchr.9344>.

16 Christine L. Borgman, “The Digital Future Is Now: A Call to Action for the Humanities,” *DHQ: Digital Humanities Quarterly* 3, no. 4 (2009).

17 Karin Knorr Cetina, *Epistemic Cultures: How the Sciences Make Knowledge* (Harvard University Press, 1999); Bruno Latour and Steve Woolgar, *Laboratory Life: The Social Construction of Scientific Facts* (SAGE Publications, 1979).

18 Peter Galison, “Computer Simulations and the Trading Zone,” in *The Disunity of Science: Boundaries, Contexts, And Power*, ed. Peter Galison and David J. Stump (Stanford University Press, 1996), 119, emphasis in original.

one community has a stronger negotiating power to decide goals and practices than the other community. For example, computational experts may push a tool for historians while historians remain unable to adapt the tool to their needs. Finally, changing practices as the extent to which the trading zone remains an interaction of distinct communities, or merges into a singular community of shared practices. That is, whether these trading zones remain distinct historians and computational experts, or blend into a community of digital historians.

Following arguments that digital history is to be positioned between the traditionally historical and the computational or digital, I focus on practices between these two ends.¹⁹ I argue that digital history does not occupy a singular position between the digital and the historical. Instead, historians continuously move across this dimension, choosing (or finding themselves in) different positions as they construct different trading zones through cross-disciplinary engagement, negotiation of research goals and individual interests.

This book is thereby aimed at scholars interested in digital history and its relations to the historical discipline and to digital humanities. At the heart of my investigation are the processes of negotiating and exchanging of disciplinary practices, and how such trading affects the way historians practice historical research. Furthermore, this book will be of interest to scholars working on interdisciplinary collaborations towards digital research infrastructures.

Structure of the Book

In the rest of this chapter, I contextualise digital history by discussing its relationship to digital humanities and exploring its origins in histories of library sciences, archival sciences and historiography. I argue that for many decades historians have been able to trust librarians and archivists to facilitate historical research, without deeply engaging with these communities. However, with digital infrastructures, the structure of databases affects what historians can do and what questions can be pursued. I argue that this change of infrastructure for historical research is what necessitates cross-disciplinary collaborations, so that historians steer these infrastructures into directions suitable for historians.

In Chapter 2, I develop a theoretical model for analysing cross-disciplinary collaborations, basing my work on the concept of trading zones. I elaborate this

¹⁹ Jennifer Edmond, “The Role of the Professional Intermediary in Expanding the Humanities Computing Base,” *Literary and Linguistic Computing* 20, no. 3 (2005): 367–80, <https://doi.org/10.1093/lc/fqi036>; Patrik Svensson, “The Digital Humanities as a Humanities Project,” *Arts and Humanities in Higher Education* 11, no. 1–2 (2011): 42–60, <https://doi.org/10.1177/1474022211427367>.

concept according to the three aforementioned dimensions. Based on the work of the sociologists Harry Collins, Robert Evans and Michael Gorman, trading zones are conceptualised according to the first two dimensions of changing practices and power relations. The dimension of changing practices describes the extent to which practices in a trading zone remain heterogeneous, conducted by two distinct communities, or homogeneous, conducted by a unified community without distinction. The dimension of power relations describes who is in control of practices in trading zones, where I build upon the work of the philosopher Michel Foucault. I extend this two-dimensional model of trading zones with the framework of communities of practice by the educational theorist Étienne Wenger to better describe how the communities in trading zones engage with one another. Following the elaboration of the theoretical model, I discuss how the concept of trading zones has been applied in digital humanities literature thus far, noting that this literature has not sufficiently considered local variations in digital humanities practices. Finally, I elaborate how my method of research is based in ethnographic work as described by Clifford Geertz.

In Chapter 3, I explore the first dimension of engagement by analysing how historians in digital history collaborations engage with historical peers and cross-disciplinary collaborators. Such engagements include interdisciplinary boundary crossing, where historians cross disciplinary boundaries to engage with computational experts. Yet by doing so historians may develop new practices and vocabularies that hinder discussion with historical peers, leading to intradisciplinary boundary construction where historians become separated. Finally, I explore how digital history collaborations may cross institutional boundaries, through collaborations between different institutes, or construct such boundaries, through the institutionalisation of centres and labs. I elaborate such mechanisms of engagement by an analysis of ethnographic observations and interviews about digital history as conducted at the University of Luxembourg. This university established the Luxembourg Centre for Contemporary and Digital History (C²DH) in 2016 and the Digital History Lab in 2015.

In Chapter 4, I explore the second dimension of power relations by analysing how participants in digital history collaborations coordinate tasks and goals. I describe four case studies of digital history collaborations of which I have interviewed multiple participants to gain differing perspectives of the goals of the collaborations. I analyse these interviews building upon the work of the information scientist Judith Weedman who described incentives for collaborating as related to 1) reasons for joining a project, 2) individual goals for a project, and 3) expected effects of participation after the project has ended. Through this analysis I identify six categories of incentives: 1) funding, 2) digital history/humanities, 3) data, 4) tool development, 5) historical research and 6) computational research.

In the second half of the chapter, I juxtapose these incentives to analyse how incentives conflicted in collaborations and how such conflicts were resolved, leading to power asymmetries and detachment of individual practices from the collaboration.

In Chapter 5, I explore the final dimension of changing practices. In this concluding chapter I analyse the extent to which the historians in my studies adopted new practices that altered how they conducted historical research. By reviewing the findings of chapters 3 and 4 I show that the changes of practices are not uniform for all historians participating in digital history. Instead, the professors in history who led the institutional units (chapter 3) and collaborations (chapter 4) served as what I call digital history brokers who connect and translate between the historical and computational communities in digital history trading zones. I argue that digital history brokers are essentially performing infrastructuring to resolve the tensions that arise when digital infrastructures are developed and negotiated between historians and computational experts. I conclude this chapter with a set of recommendations for future digital history collaborations.

Positioning Digital History

In studying practices and negotiations of digital history it is necessary to demarcate which practices and negotiations count as examples of digital history. Digital history and digital humanities more broadly are underdefined, and volumes have been dedicated to questions of whether “digital humanities” refers to a discipline, field, or something else, who is part of it, and how it must further be defined.²⁰ The website <https://whatisdigitalhumanities.com/> demonstrates this in an ironic fashion by providing a different definition from a scholar every time the visitor refreshes the page. The scholar of digital media Smiljana Antonijević tellingly groups her discussion of the terminology, boundary work and communities of digital humanities under the section *Controversies in Digital Humanities*.²¹

Rather than definitions of what should and should not count, the digital humanist Roopika Risam suggests “accents” to recognise and respect that practices are localised, and may be different between geographical, linguistic, or

²⁰ For example, see Melissa Terras, Julianne Nyhan and Edward Vanhoutte, eds., *Defining Digital Humanities* (Ashgate, 2013).

²¹ Smiljana Antonijević, *Amongst Digital Humanists: An Ethnographic Study of Digital Knowledge Production*, pre-print (Basingstoke New York, NY: Palgrave Macmillan, 2015), 16–29.

disciplinary communities.²² Rather than a singular global model of digital humanities that highlights certain practices at the expense of others, digital humanities may be considered a global field of diverse, bordering areas where no area is central to all.²³ Since my study is empirically based within the context of the Netherlands, Belgium and Luxembourg, my accent of digital history emphasises practices present in this region and its geographical, linguistic or epistemic neighbours. Therefore, my positioning of digital history is not meant as a global definition, but as a characterisation of the practices that I investigate.²⁴

To start from a broader view, I see digital history within the scope of digital humanities. This view is not uncontested. Some authors argue that the two have different topical emphases.²⁵ Furthermore, digital humanities is commonly traced to the Italian Jesuit priest Roberto Busa, while digital history is traced to quantitative history and public history.²⁶ Yet the terms overlap in several significant ways. Digital history is arguably one of the dominant strands within digital humanities and is strongly represented at digital humanities conferences.²⁷ Both furthermore

22 Roopika Risam, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy* (Northwestern University Press, 2018), <https://doi.org/10.2307/j.ctv7tq4hg>.

23 Amy E. Earhart, “Digital Humanities Within a Global Context: Creating Borderlands of Localized Expression,” *Fudan Journal of the Humanities and Social Sciences* 11, no. 3 (2018): 357–69, <https://doi.org/10.1007/s40647-018-0224-0>.

24 Vered Amit, “Introduction,” in *Constructing the Field: Ethnographic Fieldwork in the Contemporary World*, ed. Vered Amit (Routledge, 2000), 1–18.

25 Stephen Robertson, “The Differences between Digital History and Digital Humanities,” in *Debates in the Digital Humanities* (University of Minnesota Press, 2016).

26 For origins of digital humanities, see Susan Hockey, “The History of Humanities Computing,” in *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens and John Unsworth, online (Blackwell, 2004), 3–19; Steven E. Jones, *Roberto Busa, S. J., and the Emergence of Humanities Computing: The Priest and the Punched Cards* (2018); for origins of digital history, see Edward L. Ayers, “The Pasts and Futures of Digital History,” *History News* 56, no. 3 (2001): 5–9; Stephen Brier, “Confessions of a Premature Digital Humanist,” *The Journal of Interactive Technology & Pedagogy*, no. 11 (2017); Shawn Graham, Ian Milligan and Scott Weingart, *Exploring Big Historical Data: The Historian’s Macroscope* (Imperial College Press, 2015); the historian Jane Winters, however, draws the origins of digital history to Busa, arguing that his work was “very clearly an exercise in historical research” Jane Winters, “Digital History,” in *Debating New Approaches to History*, ed. Marek Tamm and Peter Burke (Bloomsbury Academic, 2018), 277.

27 A number of analyses of DH conferences show history as a strong strand within the field. Scott Weingart has analysed submissions to the ADHO DH conference in 2017, with historical studies as the fifth discipline “Submissions to DH2017 (Pt. 1),” *The Scottbot Irregular* (blog), November 10, 2016, <http://scottbot.net/submissions-to-dh2017-pt-1/>; Eetu Mäkelä and Mikko Tolonen analysed submissions to DHN2018, finding historical studies as the top discipline,

share important commonalities. Digital humanities and digital history emerge in the meeting between computational approaches to historical or humanistic subjects.²⁸ Both share dispositions towards texts.²⁹ Therefore, I regularly place digital history in the wider context of discussions about digital humanities, providing a much wider ground for what constitutes digital humanities and how it affects practices within the humanities.

In characterising digital history, several authors have argued that it involves approaching (preferably big) data with tools to create a narrative or other representation of the past.³⁰ It has furthermore been argued that digital history is also about the reflection on these practices and understanding how the digital changes the way historians work.³¹ Finally, digital history has been said to be an interdisciplinary collaboration, not only using available datasets and tools but developing them.³² This emphasis on development is resonated in debates about the digital humanities, with scholars emphasising practices such as modelling, building, or even creating infrastructures for large datasets.³³ What these authors share is that

“DHN2018 – an Analysis of a Digital Humanities Conference” (Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference, CEUR-WS, 2018), 1–9; in my own analysis of DHBenelux 2019 submissions, I found “history” and “historical” to be among the top words in abstracts “DHBenelux 2019 Submissions,” *Max Kemman* (blog), September 3, 2019, <http://www.maxkemman.nl/2019/09/dhbenelux-2019-submissions/>.

28 Edmond, “The Role of the Professional Intermediary in Expanding the Humanities Computing Base”; Svensson, “The Digital Humanities as a Humanities Project.”

29 Antonijević, *Amongst Digital Humanists*; Erik M. Champion, “Digital Humanities Is Text Heavy, Visualization Light, and Simulation Poor,” *Digital Scholarship in the Humanities* (2016), <https://doi.org/10.1093/llc/fqw053>; Kasper Risbjerg Eskildsen, “Leopold Ranke’s Archival Turn: Location and Evidence in Modern Historiography,” *Modern Intellectual History* 5, no. 3 (2008): 425–53, <https://doi.org/10.1017/S1479244308001753>.

30 Graham, Milligan and Weingart, *Exploring Big Historical Data: The Historian’s Macroscope*; Jo Guldi and David Armitage, *The History Manifesto*, online (Cambridge: Cambridge University Press, 2014), <https://doi.org/10.1017/9781139923880>; Toni Weller, “Introduction: History in the Digital Age,” in *History in the Digital Age*, ed. Toni Weller (Routledge, 2013), 1–20.

31 Zaagsma, “On Digital History.”

32 Daniel J. Cohen et al., “Interchange: The Promise of Digital History,” *The Journal of American History* 95, no. 2 (2008): 452–91; Andreas Fickers, “Towards A New Digital Historicism? Doing History In The Age Of Abundance,” *VIEW Journal of European Television History and Culture* 1, no. 1 (2012): 19–26.

33 John Unsworth, “What Is Humanities Computing and What Is Not?,” in *Defining Digital Humanities*, ed. Melissa Terras, Julianne Nyhan and Edward Vanhoutte, Digital Research in the Arts and Humanities (Routledge, 2002), 51–63; Stemphen Ramsay, “On Building,” *Stephenramsay.us* (blog), January 11, 2011, <https://web.archive.org/web/20170704144620/http://stephenramsay.us:80/text/2011/01/11/on-building/>; Richard Rogers, *Digital Methods* (MIT Press, 2013), 259.

simply using digital means in humanities scholarship by itself does not constitute digital humanities work.³⁴

My focus is on the cross-disciplinary collaboration through which digital history development is performed. Yet, some scholars might argue that digital history is most interesting when conducted by individuals, when historians learn how to write software code themselves. Reflections on how understanding code shapes practices exist as well for the digital humanities more broadly.³⁵ Yet, here too, historians depend on utilising a language developed by computational experts. Furthermore, programming largely depends on importing packages developed by others and combining these in appropriate flows.³⁶ There is thus still an indirect interaction as historians import concepts and tools developed by computational experts. Since my interest is in how the import of methods and practices affects historians, an individual view of digital history makes these methodological and epistemological tensions internal. It is not a coincidence that essays that consider how historians are affected as users of technology have taken the form of reflective pieces of internal tensions.³⁷ By studying collaborations instead, I aim to make these tensions, the uncertainty of digital history and the process of negotiation explicit and observable.

But what is it that is being developed in these digital history collaborations? I argue that this can be characterised as the development of infrastructures, where the goal is that the product of the collaboration may underlie historical research, during the project or in the future. This future historical research then need not be conducted through cross-disciplinary collaborations, nor does it demand advanced technical proficiency. I understand these infrastructures as the constellation of technologies and practices required to access, collect and analyse sources for historical research. Now that more and more aspects of historical scholarship are becoming digital, the need for digital infrastructures that

34 Antonijević, *Amongst Digital Humanists*; Anne Burdick et al., *Digital Humanities* (MIT Press, 2012).

35 Joris van Zundert and Ronald Haentjens Dekker, “Code, Scholarship, and Criticism: When Is Code Scholarship and When Is It Not?,” *Digital Scholarship in the Humanities* 32, no. suppl_1 (2017): i121–33, <https://doi.org/10.1093/llc/fqx006>.

36 Églantine Schmitt, “Des Humains Dans La Machine: La Conception d’un Algorithme de Classification Sémantique Au Prisme Du Concept d’objectivité,” *Sciences Du Design* 2, no. 4 (2016): 83–97.

37 Lara Putnam, “The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast,” *The American Historical Review* 121, no. 2 (2016): 377–402, <https://doi.org/10.1093/ahr/121.2.377>; Julia Laite, “The Emmet’s Inch: Small History in a Digital Age,” *Journal of Social History* (2019), <https://doi.org/10.1093/jsh/shy118>.

facilitate the scholarly cycle becomes increasingly urgent.³⁸ The development of digital infrastructures depends on collection experts, computational linguists and computational researchers to collaborate on physical technology, digital technology and user interfaces. Historians have started to become aware of this, with some historians criticising their profession for their silence on the impact of digitisation.³⁹ Others furthermore called for historians to become actively involved:

It was previously enough to take a thing – a printed volume, or an archival box – and place it upon a scholar’s desk; there was no need to know what was being done with it in order to deliver it correctly. Now, as material is delivered digitally, every design decision taken when building new user interfaces allows some kinds of use but may exclude others. [. . .] This is then a call to historians to be there at the beginning of that process, to help design those systems to meet our needs.⁴⁰

As a result, the hidden infrastructures underlying historical practices have become visible. The historians I study in this book have joined collaborations to shape the infrastructures to their disciplinary needs, so that other historians may benefit from the new technological means without the requirement of learning how to code or collaborate with computational experts themselves. My study of how digital history affects historical practices thereby follows the approaches developed in the field of social construction of technology:

Technology is not an independent, non-social variable that has an ‘impact’ on society or culture. On the contrary, any technology is a set of social behaviours and a system of meanings. To restate the point: when we examine the ‘impact’ of technology on society, we are talking about the impact of one kind of social behaviour on another.⁴¹

The “impact of one kind of social behaviour on another” in my case is the impact of collaborative negotiations of digital history on practices of the wider history discipline.

In short, I position digital history in the negotiations and practices between historians and computational experts in the development of digital infrastructures to the benefit of historical research more broadly. Yet infrastructures are

38 Jennifer Edmond et al., “Springing the Floor for a Different Kind of Dance – Building DARIAH as a Twenty-First-Century Research Infrastructure for the Arts and Humanities,” in *Digital Technology and the Practices of Humanities Research*, ed. Jennifer Edmond (Open Book Publishers, 2020), 207–34, <https://doi.org/10.11647/obp.0192.09>.

39 Fickers, “Veins Filled with the Diluted Sap of Rationality: A Critical Reply to Rens Bod.”

40 Peter Webster, “Digital Contemporary History: Sources, Tools, Methods, Issues,” *Temp: Tidsskrift for Historie* 14 (2017): 37.

41 Bryan Pfaffenberger, “Fetishised Objects and Humanised Nature: Towards an Anthropology of Technology,” *Man* 23, no. 2 (1988): 42, <https://doi.org/10.2307/2802804>.

not a new phenomenon to historical scholarship. In the next section I therefore position digital history in historiography, by tracing its roots in developments in historical research and its infrastructures of archives and libraries.

Origins of Digital History

As I argue above, historians have been called to use computers since the 1940s. This raises the question of what makes digital history different from earlier periods. In order to understand the current state of digital history, it is useful to consider the debates that led to what is now called digital history. These debates surround how to search, collect and analyse source material, especially when confronted with overabundant source material.⁴² To provide insights into the shifting practices and arguments, I start from the 1940s, the period in which modern computers and practices of computing were invented. From there on, I consider several developments in the history of history as a profession, and its relationship with closely related professions. History as a community of historians with shared practices and concepts cannot be described without considering the archives and libraries that are central to historical research. As such, this history considers the “inside”, the history of the historical discipline, as well as the “outside”, the history of developments in work practices and infrastructures in general over many fields.⁴³ I thus synthesise the “inside” historiography of historical practices, with the “outside” developments of archives, libraries and information technology as infrastructural to historical practices.

The historian Ernst Breisach distinguishes between two forms of historiography.⁴⁴ The first approach is to provide an overview of perspectives and debates, without assuming historiography has a certain direction or that historical research improves over time. The second approach, in contrast, is to discuss historiography as the development of history as a science, giving preference to historians who aided that development, while neglecting arguments that did not endure. I take

⁴² E.g., Fickers, “Towards A New Digital Historicism? Doing History In The Age Of Abundance”; Ian Milligan, *History in the Age of Abundance?: How the Web Is Transforming Historical Research* (2019); Roy Rosenzweig, “Scarcity or Abundance? Preserving the Past in a Digital Era,” *The American Historical Review* 108, no. 3 (2003): 735–62, <https://doi.org/10.1086/529596>.

⁴³ Geoffrey C. Bowker, “The History of Information Infrastructures: The Case of the International Classification of Diseases,” *Information Processing and Management* 32, no. 1 (1996): 49–61, [https://doi.org/10.1016/0306-4573\(95\)00049-M](https://doi.org/10.1016/0306-4573(95)00049-M).

⁴⁴ Ernst Breisach, *Historiography: Ancient, Medieval & Modern*, 2nd ed. (Chicago: University of Chicago Press, 1994), 3–4.

the latter approach, focusing on the development of the boundaries and boundary practices of history as a community of practice. For ease of reading, I thereby describe a more or less linear path of developments leading towards the current state of digital history.

Continuing the view of disciplines as communities sharing practices and concepts, I describe the historiographical developments as boundary work.⁴⁵ That is, there is a continuous debate about what it means to be an academic historian, what a good historical analysis is and what the role of sources must be in historical research. Historians thus draw boundaries within which a historian must operate to remain recognisable as a historian. To fall outside of that boundary would mean their work is no longer recognised as historical scholarship. Such boundary work is similarly prevalent in the archival and library professions, as I show. Yet in order to develop this boundary work, the three communities simultaneously cross boundaries in their interdependency on one another. I, therefore, show how archives and libraries are infrastructural to historical scholarship.

1940s–1970s: Expansion & Automation

As noted in the introduction, perhaps the first historian who argued historians should use computers for scholarship was Murray Lawson, who presented his paper at the annual meeting of the American Historical Association in 1946.⁴⁶ In this paper, he described how a combination of punched cards and microfilm would enable historians to counter the abundance of source material. His vision of historical research using machines was based on the earlier writing of the engineering scientist Vannevar Bush who published his famous *As We May Think* in 1945, in which he proposed a hypothetical *memex* device which combined microfilms and punched cards to store literature and provide quick access to the individual scientist faced with an abundance of publications.⁴⁷ In a similar problem statement, the librarian Fremont Rider (1885–1962), a student of Melvil Dewey, published *The Scholar and the Future of the Research Library, a Problem and Its Solution* in 1944.⁴⁸ In this book, Rider extrapolated the growth of libraries

45 Thomas F. Gieryn, “Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists,” *American Sociological Review* 48, no. 6 (1983): 781–95, <https://doi.org/10.2307/2095325>.

46 Lawson, “The Machine Age in Historical Research.”

47 Vannevar Bush, “As We May Think,” *The Atlantic Monthly* 176, no. 1 (1945): 101–8.

48 Fremont Rider, *The Scholar and the Future of the Research Library, a Problem and Its Solution* (N.Y.: Hadham Press, 1944).

to predict unmanageable amounts requiring vast library storage space.⁴⁹ Moreover, the Library of Congress provided the catalogue information for other libraries by selling catalogue cards, yet was itself falling behind in its efforts so that catalogue cards were created with a delay, or not at all.⁵⁰

While concerns of abundant collections were not new to their field, archivists of the 1940s too deemed existing approaches no longer sufficient.⁵¹ The archivist and scholar of archival theory Terry Cook described the development at the US National Archives as follows:

When the National Archives in Washington was created in 1934, it inherited an awesome backlog of about one million metres of federal records, with a growth rate of more than sixty thousand metres annually. By 1943, under the expansion of the state to cope with the Great Depression and World War II, that growth rate had reached six hundred thousand metres annually.⁵²

In short, the US National Archives saw the number of new to be added records rise by tenfold within a decade. Librarians and archivists developed diverging ideas about how to confront these problems. Some archivists rejected the increasingly impractical ideas of the archivist Hilary Jenkinson, who had argued in 1922 that archivists must not perform any interpretation, but keep all records produced by archived administrations, so that the archive would remain as objective evidence.⁵³ The archivist Margaret Cross Norton, who co-founded the Society of American Archivists in 1936, stated that “it is obviously no longer possible for any agency to preserve all records which result from its activities. The emphasis of archives work has shifted from preservation of records to selection of records for preservation”.⁵⁴

Archivists increasingly needed to select what should be archived, and what should otherwise be discarded. The appraisal of documents, earlier rejected by Jenkinson on the grounds of it being a subjective exercise tainting the objectivity of the archive, was inevitable, but needed to be systematised so as to retain a

49 Rolland E. Stevens, “The Microform Revolution,” *Library Trends* 19, no. 3 (1971): 379–95.

50 Barbara B. Tillett, “Catalog It Once for All: A History of Cooperative Cataloging in the United States Prior to 1967 (Before MARC),” *Cataloging & Classification Quarterly* 17, no. 3–4 (1994): 3–38, https://doi.org/10.1300/J104v17n03_02.

51 Cf. Ann Blair, *Too Much to Know: Managing Scholarly Information before the Modern Age* (New Haven, Conn: Yale University Press, 2010).

52 Terry Cook, “What Is Past Is Prologue: A History of Archival Ideas Since 1898, and the Future Paradigm Shift,” *Archivaria*, no. 43 (1997): 26.

53 Hilary Jenkinson, *A Manual of Archive Administration Including the Problems of War Archive Making* (The Clarendon Press, 1922); Cook, “What Is Past Is Prologue.”

54 Quoted in Cook, “What Is Past Is Prologue,” 26.

professional status. The archivist Theodore Schellenberg synthesised the rules for appraisal, and thus became “the father of appraisal theory in the United States”.⁵⁵ In 1956 Schellenberg published *Modern Archives: Principles and Techniques*, in which he argued documents had primary and secondary values.⁵⁶ The primary value referred to the value for the original creator of a document. The secondary value referred to the unforeseen use in the future by others, due to evidential or informational values. Evidential value is the historical value for researchers, as a trace of the functioning of the organisation in which the document was created. Informational value is the research value of the contents of a document as traces of the societal context in which the document was created. Of interest to note here is the close relationship between Schellenberg’s principles of appraisal, and the historical profession:

Since Schellenberg’s generation also coincided in its upbringing with the widespread professionalization of academic history in the universities, it is also not surprising to find in his work the close identification of archivists with historians, and archival “informational value” with historical themes and interpretations.⁵⁷

Future use by historians consequently became a central criterion for the selection of documents for American archivists. After Schellenberg there was thus arguably a true mutual dependency between historians and archivists; where historians had been dependent on archivists since the historian Leopold von Ranke had emphasised the systematic study of archival sources, archivists were now becoming dependent on historians to determine what should be archived in the first place.⁵⁸ However, this “use-based approach” to appraisal was also criticised for being non-transparent, as well as for introducing a theory of appraisal dependent on contexts unrelated to the creation and use of the original document.⁵⁹ Although appraisal would thus become a core practice for archivists, how to appraise documents remained a matter of debate.

Librarians in contrast did not debate the extent to which documents could be selected or discarded. With respect to the problem of cataloguing running

⁵⁵ F. Gerald Ham, *Selecting and Appraising Archives and Manuscripts* (The Society of American Archivists, 1993), 7.

⁵⁶ Theodore R. Schellenberg, *Modern Archives. Principles and Techniques* (F.W. Cheshire, 1956).

⁵⁷ Cook, “What Is Past Is Prologue,” 29.

⁵⁸ Rens Bod, *De Vergeten Wetenschappen: Een Geschiedenis van de Humaniora* (Prometheus, 2010); Georg G. Iggers, “The Crisis of the Rankean Paradigm in the Nineteenth Century,” *Syracuse Scholar* (1979–1991) 9, no. 1 (1988); Georg G. Iggers, “The Professionalization of Historical Studies and the Guiding Assumptions of Modern Historical Thought,” *A Companion to Western Historical Thought*, 2007, 225–42, <https://doi.org/10.1002/9780470998748.ch12>.

⁵⁹ Cook, “What Is Past Is Prologue.”

behind the addition of books, rules for standardisation were suggested in 1940 by the Library of Congress (LoC) with the aim of simplifying the process. On the one hand, librarians needed to develop social practices to create trust in the co-operation between LoC and other libraries:

As for accepting the work of others, [Andrew Osborn] noted catalogers cannot even accept uncritically the cataloging from the Library of Congress, but must add, subtract, and modify the records, until they might as well have cataloged it themselves. He said not just large libraries, but also small libraries did this.⁶⁰

On the other hand, librarians relied on technological solutions to combat the growth of the collection. Two technologies are of interest in this history; microphotography and punched cards.

In the 1940s, microphotography was hardly an innovative technology, having been invented in 1839 by John Dancer. Yet, at that time there was no clear use case for microphotographs. It took until the 1920s before microfilms, the most commonly used format of microphotography, became prevalent, and until the 1930s before it was used seriously by historians and librarians.⁶¹ From 1935–1942 American historians participated in the *Historical Records Survey*, a New Deal program in which historians surveyed records in archives and libraries of historical value. While inventorying, historians were asked to microfilm these records. The main goals for microfilm were preservation of fragile material, as well as to provide wider access, as microfilms could be copied and distributed more widely than the original documents. Yet these two goals ultimately failed within the programme, as microfilms were of poor quality, unreadable and the microfilms themselves ended up being as inaccessible as the original records.⁶² However, as an experimental trial, it was successful in innovating the methods of microfilming, as well as in proving the utility of the technology.

Librarians were soon convinced of the wonderful promises of microfilm, with some hailing it as the most important innovation since Gutenberg's printing press. Microfilms were discussed to such an extent that it seemed almost an end in itself.⁶³ Several reasons drove the enthusiasm for microfilming, notably access (to obtain rare books), preservation (to replace items on deteriorating paper), usability (to replace large volumes such as newspaper volumes that were difficult to handle) and

⁶⁰ Tillett, "Catalog It Once for All," 29.

⁶¹ Stevens, "The Microform Revolution."

⁶² Clifton D. Foster, "Microfilming Activities of the Historical Records Survey, 1935–42," *The American Archivist* 48, no. 1 (1985): 45–55.

⁶³ Susan A. Cady, "The Electronic Revolution in Libraries: Microfilm Déjà Vu?," *College & Research Libraries* 51, no. 4 (1990): 374–86.

saving space (to replace print material with much smaller microforms).⁶⁴ However, despite the warnings of Fremont Rider about unmanageable collections, librarians did not initially deem saving space a primary reason for microfilming. When they did so, it was mainly done for large bound volumes such as newspapers. Many libraries held newspapers, creating a large market for microfilms. Books in contrast only rarely ended up being microfilmed.⁶⁵ Instead, preservation and access were the main reasons for microfilming. Yet, not everyone was as enthusiastic about microfilm. Readers found it difficult to use the microfilms, as documents were put on microfilm at a higher pace than the development of usable microfilm readers. Cost-savings were a reason for microfilming, and the quality of the images was not always considered as much as should have been.⁶⁶ Microfilm then failed to fulfil the promise of greater usability. The enthusiasm of librarians for microfilm was consequently not shared by scholars and historians. Once beyond the peak of the technology's hype in the 1970s, the primary purpose moved from preservation and access to saving storage space, but with less excitement than during the hype of the 1940s–50s.⁶⁷ Microfilm thus did not end up transforming the library. The technology's potential advantages did not convince the community to accept the disadvantages.

It could, however, be argued that microfilm did end up transforming archives. Although seemingly not as heavily debated in literature, microfilm introduced a significant possibility for archives. Many of the collections in archives are unique to that archive, contrary to most library collections. Historians need to visit a specific archive to read a unique document. With microfilm, archives could duplicate parts of their collections and make these available in archives elsewhere, even on other continents. The access that microfilm could provide to source material located elsewhere was a significant transformation of infrastructure. Historians gladly accepted a cumbersome microfilm reader if this allowed them to find a piece of information they could not have consulted otherwise. Overall, it is of interest that with microfilm, the first attempts at transforming the collections of infrastructures are seen; first for reasons of access and preservation, and later for financial reasons of cheap material and saving storage space. This process would later be repeated with the digitisation of collections.

64 Stevens, "The Microform Revolution."

65 Stevens.

66 Cady draws a parallel between the advantages and disadvantages of microfilms and of modern-day digital texts, in saving space but introducing issues of readability: "The Electronic Revolution in Libraries."

67 Cady.

The other technology of interest to librarians was punched cards. This, too, was hardly an innovative technology, invented in 1887 and first used on a large scale for computing the results of the 1890 US Census.⁶⁸ This census was one of the early situations in which there was too much data to handle, requiring technological innovations, and as such punched cards became the “big data” technology of the late nineteenth and early twentieth century. Yet, it took several decades before scientists and librarians became invested in the technology. Librarians started experimenting with punched cards in the 1930s, but throughout the 1940s were more excited about microfilm instead. However, in the 1940s, scientists became increasingly interested in punched cards for the use of information processing and retrieval.⁶⁹ It nonetheless took until the 1960s before punched cards became systematically used in libraries. With the advent of computer systems, libraries became increasingly invested in using these systems, with punched cards as the input mechanism for entering data. At first, automation of library systems focused on library circulation and keeping track of inventories, but later on libraries turned to computer systems for the creation and maintenance of catalogues. Libraries had already grown accustomed to using catalogue cards to describe and maintain their collections.⁷⁰ Therefore, it was a small step to recreate these cards as punched cards.

In 1965, the Library of Congress started investigating the use of computer systems for library processes, leading to the MARC (Machine-Readable Cataloging) format in 1968, then named MARC II.⁷¹ It took another decade for the MARC format to be fully recognised by libraries, but then became such an important standard that machine-readable formats, in contrast with microfilms, arguably did transform libraries, laying the groundwork for later digitisation projects. Punched cards were not an end in itself the way microfilm was perceived to be, but was a necessary medium in the first steps to moving library catalogues to computer systems. The librarian Sally McCallum concluded there were three reasons why MARC became a central piece of libraries.⁷² First, it was innovative. Second, it was developed collaboratively, by engineers with participation of librarians. And third, the

68 Robert V. Williams, “The Use of Punched Cards in US Libraries and Documentation Centers, 1936–1965,” *IEEE Annals of the History of Computing* 24, no. 2 (2002): 16–33, <https://doi.org/10.1109/MAHC.2002.1010067>.

69 Williams.

70 Markus Krajewski, *Paper Machines: About Cards & Catalogs, 1548–1929*, trans. Peter Krapp (MIT Press, 2011), <https://doi.org/10.7551/mitpress/9780262015899.001.0001>; Tillet, “Catalog It Once for All.”

71 Sally H. McCallum, “MARC: Keystone for Library Automation,” *IEEE Annals of the History of Computing* 24, no. 2 (2002): 34–49, <https://doi.org/10.1109/MAHC.2002.1010068>.

72 McCallum.

Library of Congress, which already was a central institute in the creation and distribution of catalogue cards in the USA, adopted the format immediately for its catalogues, so that other libraries could benefit from this.

During the 1960s, historians became invested in computer systems as well. In this decade, universities established computer centres featuring a mainframe computer, which required punched cards for both input and output. Despite historians being called conservative by Lawson, there are several examples of historians who quickly adopted mainframes for research. For example, the historian Tito Orlandi already experimented with punched cards for the creation of a critical edition during his doctoral research in 1960.⁷³ More generally for the humanities, this is the period where the founding of “humanities computing”, a forerunner of digital humanities, is traced back to. While the founding myth of digital humanities starts in 1949 when the Jesuit priest Father Roberto Busa approached IBM with the request to collaborate on an automated concordance of the works of St. Thomas Aquinas on punched cards, it took until the 1960s before humanities computing became more established.⁷⁴ In 1963 the University of Cambridge founded the Centre for Literary and Linguistic Computing, in 1966 the University of Tübingen appointed a research officer for computer applications in the humanities and in that same year the journal *Computers and the Humanities* was founded.⁷⁵

Historians in this period engaged with university computer centres, but not under the flag of humanities computing. After 1945, historians became more involved with the social sciences for the adoption of theoretical theories as well as methods. In Europe, and especially France, this happened mainly within the *Annales* school, while in the USA this happened under the flag of quantitative history, also known as cliometrics. This latter movement became more fully established in the 1960s, with the founding of journals such as *Historical Methods* and the *Journal of Social History* in 1967, the *Journal of Interdisciplinary History*

73 Julianne Nyhan and Andrew Flinn, “Hic Rhodus, Hic Salta: Tito Orlandi and Julianne Nyhan,” in *Computation and the Humanities*, Springer Series on Cultural Computing (Cham: Springer International Publishing, 2016), 75–86, <https://doi.org/10.1007/978-3-319-20170-2>.

74 Hockey, “The History of Humanities Computing”; Jones, *Roberto Busa, S. J., and the Emergence of Humanities Computing*; this founding myth of digital humanities is increasingly debated and contested, c.f. Rachel Sagner Buurma and Laura Heffernan, “Search and Replace: Josephine Miles and the Origins of Distant Reading,” *Modernism/Modernity* 3, no. 1 (2018).

75 Hockey, “The History of Humanities Computing”; Julianne Nyhan and Andrew Flinn, “The University Was Still Taking Account of Universitas Scientiarum: Wilhelm Ott and Julianne Nyhan,” in *Computation and the Humanities*, Springer Series on Cultural Computing (Cham: Springer International Publishing, 2016), 55–73, <https://doi.org/10.1007/978-3-319-20170-2>.

in 1970, and *Social Science History* in 1976.⁷⁶ In 1967, a conference was held in the USA to discuss the then current state of quantitative history with three aims: 1) to present notable findings of earlier scholars, 2) to survey material that could be used for quantitative research, and 3) to raise hopes for the future.⁷⁷ Participants succeeded largely in the final part, with the (in)famous conclusion from the historian Emmanuel Le Roy Ladurie that “the historian will be a programmer or he will be nothing” in 1968.⁷⁸ The historian Theodore Rabb enthusiastically wrote that “[n]ot since the days of Leopold von Ranke and his followers has there been such joy and excitement about the discovery or the inventive new use of documentary evidence”.⁷⁹

Although this may in hindsight have been an exaggeration, it is fair to say quantitative history was a step further in the professionalisation of the field started by Von Ranke. It brought attention to the accumulation of datasets, making history a more cumulative science. Due to the emphasis on methods of statistics, analysis arguably became more transparent and open to debate.⁸⁰ One of the most prominent works in the field is *Time on the Cross* by the economic historians Robert Fogel and Stanley Engerman, published in 1974.⁸¹ Fogel and Engerman studied the economics of slavery in southern states in the USA and tested the then commonly agreed belief that slavery was economically inefficient. After investigating economic and social factors, they concluded that slavery was economically viable and states with slavery were actually more efficient than states without. Its reception was generally positive at first, but their study was later denounced for containing too many errors to support the conclusions and questions were raised about whether a numerical view of a moral issue such as slavery was valid.⁸² Yet the explicit methods, datasets and statistics allowed for a scholarly debate to emerge that would have been difficult otherwise.

76 John F. Reynolds, “Do Historians Count Anymore?: The Status of Quantitative Methods in History, 1975–1995,” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 31, no. 4 (1998): 141–48, <https://doi.org/10.1080/01615449809601196>.

77 Theodore K. Rabb, “The Development of Quantification in Historical Research,” *Journal of Interdisciplinary History* 13, no. 4 (1983): 591–601.

78 Quoted in Stone, “The Revival of Narrative,” 13.

79 Rabb, “The Development of Quantification in Historical Research,” 596.

80 Morgan Kousser, “Quantitative Social-Scientific History,” in *The Past before Us: Contemporary Historical Writing in the United States*, ed. M. Kammen (Cornell University Press, 1980).

81 Robert W. Fogel and Stanley L. Engerman, *Time on the Cross: The Economics of American Negro Slavery* (1974; repr., New York; London: W.W. Norton, 1995).

82 Michiel Leezenberg and Gerard de Vries, *Wetenschapsfilosofie Voor Geesteswetenschappen*, 5th ed. (Amsterdam University Press, 2001).

This transparency was, however, simultaneously one of the weaknesses of the field. It opened studies to criticism, leading to more rejections from reviewers and thus fewer successful publications.⁸³ Another problem was that quantitative history did not always fit within the boundary work of historians, argued to consist of:

[A] concern for the understanding and explanation of situations, processes, or events, *more* than for the theoretical means by which such understand and explanation are reached [. . .]; a willingness to relate one's findings to the classic questions of history [. . .]; an emphasis on temporal causation.⁸⁴

Quantitative history arguably strayed too far from these characteristics. It consequently did not maintain momentum after the mid 1980s.⁸⁵ Quantitative history separated from the dominant branch of the historical community, yet it did not disappear.⁸⁶ In contrast, with the advent of computers and online sources, as I detail in the next section, in recent years quantification has steadily increased not as a goal in itself but as a part of historical analysis.⁸⁷

Instead, the narrative method was revived, starting in the 1950s among a small group of historians, gaining prominence in the 1970s and arguably becoming the dominant form of history from the 1990s onward. In this revival, the methods of sociology and economics were replaced with methods of anthropology, with which historians would study the culture of a time. This movement is consequently regularly referred to as the cultural turn. Historians shifted their efforts to the analysis of power relations, mentalities and presenting these results in narrative form. The movement furthermore included the investigation of the meaning of words and ideas in their historical context, and as such led to the so-called linguistic turn.⁸⁸ According to the historian Lawrence Stone, the revival of narrative marked the end of the attempts to “produce a coherent

83 Reynolds, “Do Historians Count Anymore?”

84 Rabb, “The Development of Quantification in Historical Research,” 598, emphasis in original.

85 Reynolds, “Do Historians Count Anymore?”

86 Robert Whaples, “Is Economic History a Neglected Field of Study?,” *Historically Speaking* 11, no. 2 (2010): 17–20, <https://doi.org/10.1353/hsp.0.0109>.

87 Pat Hudson and Mina Ishizu, *History by Numbers: An Introduction to Quantitative Approaches*, Second edition (Bloomsbury Academic, 2017); Steven Ruggles, “The Revival of Quantification: Reflections on Old New Histories,” *Social Science History* 45, no. 1 (2021): 1–25, <https://doi.org/10.1017/ssh.2020.44>.

88 James Vernon, “Who’s Afraid of the ‘Linguistic Turn’? The Politics of Social History and Its Discontents,” *Social History* 19, no. 1 (1994): 81–97, <https://doi.org/10.1007/s13398-014-0173-7.2>.

scientific explanation of change in the past”.⁸⁹ History instead gained a renewed attention toward the role of interpretation in historical research, and thus reinforced hermeneutics as the core method.⁹⁰

The perceived downfall of quantitative history leads to an interesting problem of self-identification for digital history as a profession. Within the digital humanities, some historians argue that digital history has a long tradition parallel to the literary-oriented digital humanities, starting with social and quantitative history, as well as public history, in the 1970s.⁹¹ In contrast, outside of the digital humanities, historians emphasise that digital history is not a continuation of quantitative history, but actually embedded in the cultural turn.⁹²

Notable with the cultural turn is the attention towards the general people, and the required new sources to investigate these people (reasons 1 and 3 above). This signifies another step in the shifting attention from elites to the general population, which had arguably started with the economic theories of Karl Marx and was refined with quantitative history with social and economic models and sources, and thus continued as part of the narrative method with anthropological interpretation of sources.⁹³

This shift in attention by historians coincided with a shift in the archival profession, where from the 1960s onward archivists too became more concerned with records of the general population.⁹⁴ This meant that not only the traditional records from governments or institutions should be considered. Already in 1944, one of the *Annales*' most prominent historians, the historian Marc Bloch, contended that in the pursuit of a historical account of a society, all types of sources are relevant for study.⁹⁵ However, this exponentially enlarged the problem of archival overload. Not only had the amount of traditional archival documents increased, now too the number of institutions, organisations, or

89 Stone, “The Revival of Narrative,” 19.

90 Hans-Georg Gadamer, *Truth and Method* (1960; repr., Bloomsbury Academic, 2014); Hayden White, “The Question of Narrative in Contemporary Historical Theory,” *History and Theory* 23, no. 1 (1984): 1–33, <https://doi.org/10.2307/2504969>.

91 Brier, “Confessions of a Premature Digital Humanist.”

92 Ayers, “The Pasts and Futures of Digital History.”

93 Bod, *De Vergeten Wetenschappen*; Lynn Hunt, “French History in the Last Twenty Years: The Rise and Fall of the *Annales* Paradigm,” *Journal of Contemporary History* 21, no. 2 (1986): 209–24, <https://doi.org/10.1177/002200948602100205>; Leezenberg and de Vries, *Wetenschapsfilosofie Voor Geesteswetenschappen*.

94 Patrick M. Quinn, “Archivists and Historians: The Times They Are a-Changin’,” *The Midwestern Archivist* 2, no. 2 (1977): 5–13.

95 Marc Bloch, *The Historian's Craft*, trans. Peter Putnam, Repr (Manchester: Manchester Univ. Press, 2004).

individuals from which to select documents increased. This required new methods of appraisal, and from the 1980s on several theorists argued for a “societal approach” of appraisal to replace the “use-based approach”. Where the latter meant archivists selected records that ought to be of importance for future historical research, the “societal approach” meant archivists should select records that best reflect the society in which they were created.⁹⁶ In the terminology of Schellenberg, this meant archives shifted emphasis from social value, the historical value to researchers, to informational value, the research value to investigate the societal context. The scope of expertise required by archivists thus broadened even further. Not only would an archivist need to know about archival practices, but they also needed to understand historical practices to know how to provide records of importance to historical research, and now they required knowledge of sociological practices to reflect society in their archives. Furthermore, from the 1980s onward, they would have to learn new skills related to information technology.

1980s–2010s: Digitalisation

Although computers had been under development for several decades, the 1970s saw the first examples of personal computers. These were aimed at hobbyists, as they required assemblage by the owner, but in the 1980s the computer industry was transformed by computers from IBM and Apple that worked out of the box.⁹⁷ Many scholars soon had a computer standing on their desk and learned how to use this device. Scholars who previously wrote their articles and books by hand or with typing machines moved to word processing software.⁹⁸ Research too increasingly required a computer, as archives and libraries started moving their collections to digital formats.

Libraries had prepared for this “digital revolution”. The aforementioned MARC standard meant libraries already had much of their catalogue available in machine-readable form. A major actor in moving libraries into the digital period was the American Ohio College Library Center, founded in 1967, which was later renamed the Online Computer Library Center as it broadened its services outside of Ohio College, and is nowadays more commonly known simply as the OCLC. In the 1980s and 1990s, libraries started digitising their collections and

⁹⁶ Cook, “What Is Past Is Prologue.”

⁹⁷ Walter Isaacson, *The Innovators: How a Group of Hackers, Geniuses and Geeks Created the Digital Revolution* (Simon & Schuster, 2014).

⁹⁸ Matthew Kirschenbaum, *Track Changes: A Literary History of Word Processing* (Harvard University Press, 2016).

publishing these in “digital libraries” with the aim of providing access.⁹⁹ By the 1990s, most libraries had moved from card catalogues to digital systems for item retrieval in the form of OPACs (Online Public Access Catalogues). While the card catalogue was maintained for existing items, new items would only be added to the OPAC.¹⁰⁰ In 1998, OCLC launched WorldCat, an online catalogue where anyone can find items in any library connected to the WorldCat system.¹⁰¹ Digitisation of library collections thereby followed similar arguments as those around microfilms, as I discuss above. At first, catalogues and documents were digitised to provide access. In libraries with fragile materials digitisation occurred for preservation. Other libraries, such as the digital research library JSTOR, which was established in 1995, digitised with the aim of saving storage space; if libraries could provide access to a digital copy of a journal, they would be able to discard the physical copy.¹⁰²

Archivists saw a rougher transition to the new medium. At first, some wondered anxiously whether archivists would be replaced by computer specialists or information managers.¹⁰³ Early digital archives emphasised what was digitally available and could be put into the databases of the time. This mainly concerned statistical data, coinciding with the developments toward quantitative history. Yet in the mid-1980s this changed, as relational databases became available that were more compatible with existing non-digital archival practices. As archivists became more involved with the digital medium, these digital archives were also organised increasingly according to the rules of the profession. The earlier so-called “library-oriented, discrete-item approach” came under discussion from archivists that demanded more context and provenance to be embedded in the systems.¹⁰⁴ In other words, although at first some form of technological determinism provided the conditions within which archival material could be digitally stored, later on the boundary work of archivists became more active to structure digital archives according to the norms of the profession. A remaining challenge for archivists is to

99 Christine L. Borgman, “What Are Digital Libraries? Competing Visions,” *Information Processing and Management* 35, no. 3 (1999): 227–43, [https://doi.org/10.1016/S0306-4573\(98\)00059-4](https://doi.org/10.1016/S0306-4573(98)00059-4); Marilyn Deegan and Kathryn Sutherland, *Transferred Illusions: Digital Technology and the Forms of Print* (Ashgate, 2009).

100 Deegan and Sutherland, *Transferred Illusions*.

101 “WorldCat”, accessed May 12, 2021, <https://www.worldcat.org/>.

102 Deegan and Sutherland, *Transferred Illusions*.

103 Cook, “What Is Past Is Prologue.”

104 Terry Cook, “Easy To Byte, Harder To Chew: The Second Generation of Electronic Records Archives,” *Archivaria* 33 (1991): 206.

develop practices for the digital society, i.e. to create infrastructures for archiving born-digital material such as websites or social media.¹⁰⁵

Relational databases also enabled historians to employ qualitative research digitally. In 1980, the historian Manfred Thaller released the relational database management system CLIO.¹⁰⁶ This software has been argued to have initiated “history and computing” as a precursor to digital history, as it was the first database system specifically designed for historical sources and research.¹⁰⁷ The 1980s subsequently saw the establishment of the *Association for History and Computing* in 1983, of the *Nederlands Historisch Data Archief* (Dutch Historical Data Archive) in the Netherlands in 1988, and the initiation of the *Vereniging voor Geschiedenis en Informatica* (Association for History and Informatics) between Belgium and the Netherlands in 1987.¹⁰⁸ Yet, while several history programmes started including computation in their curricula, history and computing remained a small community. Practices hardly diffused to the wider discipline, despite the activities within a group of enthusiasts.¹⁰⁹

In the wider digital humanities, similar groups of enthusiasts established research centres to allow sustainable interactions between computational experts and humanities scholars, supported by third-party funding.¹¹⁰ From the mid-1980s onward, the Netherlands saw a field called *alfa-informatica* (alpha-informatics) enjoy a short peak, in which humanities students learned how to use computers and write code. However, alfa-informatics was deemed a mere support service for helping scholars use powerful but complex computers. With the advent of more usable software the field’s potential to establish humanities computing widely in the Netherlands soon drifted away in budget cuts.¹¹¹ Rather

105 Kimberly Barata, “Archives in the Digital Age,” *Journal of the Society of Archivists* 25, no. 1 (2004): 63–70, <https://doi.org/10.1080/0037981042000199151>; Christine L. Borgman, *Big Data, Little Data, No Data* (MIT Press, 2015); Niels Brügger and Niels Ole Finnemann, “The Web and Digital Humanities: Theoretical and Methodological Concerns,” *Journal of Broadcasting & Electronic Media* 57, no. 1 (2013): 66–80, <https://doi.org/10.1080/08838151.2012.761699>; Milligan, *History in the Age of Abundance?*

106 Manfred Thaller, “Automation on Parnassus Clio – a Databank Oriented System for Historians,” *Historical Social Research* 5, no. 3 (1980): 40–65.

107 Boonstra, Breure and Doorn, “Past, Present and Future of Historical Information Science.”

108 Boonstra, Breure and Doorn.

109 Speck, “History and Computing.”

110 Peter Robinson, “Digital Humanities: Is Bigger, Better?,” in *Advancing Digital Humanities*, ed. Paul Longley Arthur and Katherine Bode (London: Palgrave Macmillan UK, 2014), 243–57, https://doi.org/10.1057/9781137337016_16.

111 Joris van Zundert and Karina van Dalen-Oskam, “Digital Humanities in the Netherlands,” *H-Soz-Kult* (2014).

than historians developing software themselves for their specific purposes, they moved to generally available commercial software such as Microsoft Access. In contrast with Le Roy Ladurie's claim that historians would need to become programmers, thanks to database software, they only needed to learn to press the right buttons.¹¹² However, this move was not uncontested:

[A] wonderful big lie, with respect to the complexities of database design. It was wonderful because of its user-friendly interface. It rapidly swept away its stubborn predecessors like dBASE and Paradox. If a historical dataset was not too complicated, database design and querying were easy. Finally, the computer seemed to have reached the stage of development of the modern car: the mechanic with his oilcan was no longer needed. Built-in 'wizards' compensated for lack of theoretical knowledge and querying a database could be as simple as searching for words in a text processor. One could even successfully complete certain tasks without knowing exactly what had happened.¹¹³

Besides a move away from custom humanities software to generic commercial software, Manfred Thaller noticed a wider move away from using the computer for historical research in the 1990s. He reflected on this, perhaps somewhat cynically, as follows in an interview:

[T]he more serious disappointment, which I still think is something which has damaged parts of the Humanities, is that in the 1990s there was a move away from working with formalised results. And I have a strong suspicion that that simply relates to the fact that if you want to study a phenomenon formally – I do not say quantitatively because my own work had moved far away from quantification by the late 1980s – computers have the obnoxious habit of telling you time and time again that your data may contain errors, while what may actually be going on is that your data contains something that does not fit your hypothesis. So, it's a long and painstaking process. However, it is much, much faster, and much less frustrating to go into an archive and find a document with a human appeal and publish it and add a clever interpretation to it. Historical research has certainly fallen into what I consider a trap by getting away from doing the types of research that are harder to do.¹¹⁴

While his interpretation is debatable, it signifies that even while computers could do more than quantification, computational approaches were difficult to align with the cultural turn.

This coincides with to the so-called archival turn in the 1990s, in which historians and other scholars started to consider the archive not just as a provider

112 Speck, "History and Computing."

113 Boonstra, Breure and Doorn, "Past, Present and Future of Historical Information Science," 27.

114 Julianne Nyhan and Andrew Flinn, "It's Probably the Only Modestly Widely Used System with a Command Language in Latin: Manfred Thaller and Julianne Nyhan," in *Computation and the Humanities* (Springer International Publishing, 2016), 205, https://doi.org/10.1007/978-3-319-20170-2_13.

of research material, but as an object of study in itself.¹¹⁵ The first step towards this archival turn among historians was taken by the philosopher Michel Foucault, who argued that archives should not only be considered as physical spaces containing documents. Instead, archives constitute structures of power that keep documents in a particular order, thereby structuring what can be said about the past.¹¹⁶ The second step, which truly started this archival turn, was taken by the philosopher Jacques Derrida, who built upon the work by Foucault. He argued that archives are structures of power that determine what is preserved and what is destroyed, so that the past is not just preserved but constructed by archives.¹¹⁷ On the one hand, these arguments led to an acknowledgement of archival work that was not prevalent before. Historians consequently became interested in studying the ethnography of archives, leading to critical, postcolonial and feminist perspectives on archives as political actors.¹¹⁸ On the other hand, historians developed methods to counter the construction of the past by archives, by focusing on individuals that did not fit the general narrative of their time. In accordance with the quote from Manfred Thaller, this method has been described as follows:

[To] search the archive for eccentric anecdotes and enigmatic fragments as the basis for constructing counterhistories that interrupt the homogenizing forces of previous grand historical narratives and archival order by grounding themselves in the contingent and “the real,” all the while acknowledging that “the real” is never accessible as such.¹¹⁹

Newly developed narratives are thereby set in contrast with the narrative of the archive, while still requiring archival sources on which counternarratives are based.

Yet the historical archival turn has been criticised for not engaging with archivists. The work of archivists still remained invisible to many historians.¹²⁰

115 Eric Ketelaar, “Prolegomena to a Social History of Dutch Archives,” in *A Usable Collection: Essays in Honour of Jaap Kloosterman on Collecting Social History*, ed. Aad Blok, Jan Lucassen and Huub Sanders (Amsterdam University Press, 2014), 40–55.

116 Michel Foucault, *The Archaeology of Knowledge And the Discourse on Language*, trans. A.M. Sheridan Smith (Pantheon Books, 1972).

117 Jacques Derrida, “Archive Fever: A Freudian Impression,” *Diacritics* 25, no. 2 (1995): 9–63.

118 Alexandra Walsham, “The Social History of the Archive: Record-Keeping in Early Modern Europe,” *Past & Present* 230, no. suppl 11 (2016): 9–48, <https://doi.org/10.1093/pastj/gtw033>; Elizabeth Yale, “The History of Archives: The State of the Discipline,” *Book History* 18, no. 1 (2015): 332–59, <https://doi.org/10.1353/bh.2015.0007>.

119 Jaimie Baron, *The Archive Effect: Found Footage and the Audiovisual Experience of History* (Routledge, 2014), 3.

120 Terry Cook, “The Archive(s) Is a Foreign Country: Historians, Archivists, and the Changing Archival Landscape,” *The American Archivist* 74, no. 2 (2011): 600–632, <https://doi.org/10.17723/aarc.74.2.xm04573740262424>.

While historians studied the archive as a structure of power, archivists were not included and consulted. Arguably, although historians thus became more aware of archives, making archival structures visible in their work, some features of archival practices remained hidden from view in an infrastructural role.

Despite an apparent move away from the computer for more advanced tasks than word processing software, the 1990s brought another technology that nonetheless established the computer as an indispensable tool. In 1990, the computer scientist Tim Berners-Lee developed the HTTP protocol that laid the foundation for the World Wide Web.¹²¹ With the HTTP protocol, a document could create a link to another document, so that related documents could easily be retrieved. Reminiscent of the *memex* device by Vannevar Bush, the web transformed scientific communication and communication in general, by making it much easier to quickly retrieve documents from anywhere, as well as disseminate documents to others.¹²² In 1993, Tim Berners-Lee published the first proposal for a specification for HTML, while the software developer Marc Andreessen announced the Mosaic browser as a first easy to install and easy to use web browser with support for images.¹²³

One of the earliest examples of disseminating historical research via the web is work by the historian Edward Ayers and his collaborators. Their *The Valley of the Shadow* project on the American Civil War was published as a web page in 1993 containing maps, letters and other documents.¹²⁴ While another notable project of historical publishing including multimedia, the *Who Built America?* project, used CD-ROM for dissemination, the web page of *The Valley of the Shadow* facilitated a form of public access that proved more advantageous.¹²⁵ Although at first the project was criticised, historians soon recognised “that the digital medium allowed Ayers to create a thoroughly captivating, technically savvy, and wholly unexpected comparative approach to the Civil War, one so complex and interconnected that such a thing seemed impossible in more linear media such as film and books.”¹²⁶

121 Tim Berners-Lee and Mark Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor* (Harper San Francisco, 1999); Isaacson, *The Innovators*.

122 Bush, “As We May Think.”

123 Tim Berners-Lee and Daniel Connolly, “Hypertext Markup Language (HTML) A Representation of Textual Information and MetaInformation for Retrieval and Interchange,” *w3.org* (W3, 1993); Isaacson, *The Innovators*.

124 This project is still available online: “The Valley of the Shadow: Two Communities in the American Civil War”, accessed May 12, 2021, <http://valley.lib.virginia.edu/>.

125 William G. Thomas, “Computing and the Historical Imagination,” in *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens and John Unsworth, online (Blackwell, 2004), 116–32.

126 Thomas, 62.

As such, the project showed the first example of a historical publication online, providing easy access to sources and updates, as well as a rich media offering of images, maps and different ways of browsing the publication. From this web project came the first explicit notion of digital history, when Edward Ayers and William G. Thomas III founded the *Virginia Center for Digital History* in 1998, later defining the term as “an approach to examining and representing the past that works with the new communication technologies of the computer, the Internet network, and software systems.”¹²⁷ Interestingly, this use of the term digital history thereby pre-dates the starting point for the popularisation of the term digital humanities with the book *A Companion to Digital Humanities* published in 2004.¹²⁸

A difficulty of the early 1990s web was that it could prove difficult to find information of interest. Although there is a history of web search engines or web portals with collections of links, of interest to my discussion is the founding of Google in 1998.¹²⁹ Google started as a digital library project and is interesting for several reasons.¹³⁰ First, Google provides a single point of access to all kinds of information, originating from libraries, archives, governments etc. This is what the media scholar Siva Vaidhyanathan calls the Googlization of everything, so that Google is “the lens through which we view the world.”¹³¹ This raises the question to what extent Google provides the lens to the past when historians use Google’s services to explore libraries and read books or articles.¹³² Second, Google demonstrates the importance of the physical technical infrastructure, i.e. the machines underlying the digital infrastructure, for providing access: “they deployed far more bandwidth, processing power, and storage capacity to the task than any rival.”¹³³ Although this book does not focus on these physical, technical infrastructures, it is important to note that digital libraries are not “virtual” intangible entities. Digital libraries are embedded in physical infrastructures that introduce a power relation, as not every institution will have the funds to deploy such a technical infrastructure.¹³⁴ Finally, apart from the famous Google Search that

127 William G. Thomas in Cohen et al., “Interchange: The Promise of Digital History.”

128 Susan Schreibman, Ray Siemens and John Unsworth, eds., *A Companion to Digital Humanities*, online (Oxford: Blackwell, 2004); for a history of the digital humanities terminology, see Matthew Kirschenbaum, “Digital Humanities As/Is a Tactical Term,” in *Debates in the Digital Humanities*, ed. Matthew K. Gold, online (University of Minnesota Press, 2012).

129 For a more elaborate history of the web, see Isaacson, *The Innovators*.

130 David Hart, “On the Origins of Google,” National Science Foundation (2004).

131 Siva Vaidhyanathan, *The Googlization of Everything (And Why We Should Worry)* (Berkeley: University of California Press, 2011), 7.

132 Kemman, Kleppe and Scagliola, “Just Google It.”

133 Isaacson, *The Innovators*, 476.

134 Vaidhyanathan, *The Googlization of Everything*.

now dominates web search for billions of people, Google provides two services of interest that have successfully infiltrated the work of historians. These services provide an interface to the digital infrastructures of archives and libraries: Google Scholar and Google Books.

Google Scholar launched in 2004 as a search engine specifically for academic literature. Although to my knowledge no comprehensive study has been undertaken on how exactly historians use Google Scholar and how this impacts their usage of secondary literature, it has been shown that many historians frequently use Google Scholar.¹³⁵ Google Scholar has proved such a successful search tool that the discovery tools that were provided by university libraries have come under pressure. The Utrecht University Library was notably the first to remove their own discovery tool in 2013, instead pointing users to online search tools such as Scopus and Google Scholar.¹³⁶ Interestingly, although the library did not receive major complaints, especially scholars from the humanities were disappointed as they did not consider search tools such as Google Scholar to be apt solutions. By helping these users find specific databases for the humanities, such as JSTOR, these complaints were alleviated. This indicates that humanities scholars did not appreciate a generic, catch-all search tool, but demanded specific discovery systems tailored to their disciplines. In 2018, Utrecht University Library completed the next step to fully integrate their book catalogue in WorldCat.¹³⁷

Kortekaas and Kramer state they believe that “the OPAC is dead.”¹³⁸ This means that the library is essentially receding from the front-end, retiring the search systems developed in-house, to a back-end task of ensuring the collections are discoverable in other search systems. Moreover, journals are increasingly consulted online rather than in print, so that users are sent directly from the search tool to the journal website.¹³⁹ In other words, while the library was visible in the search user interface or in the collection, it increasingly takes on an infrastructural role of invisibly connecting other search interfaces to licensed online material, including journals and ebooks. As before with Microsoft Access, here too we might speak of a trading zone, including scholars, librarians and commercial technology firms,

135 Kemman, Kleppe and Scagliola, “Just Google It.”

136 Simone Kortekaas and Bianca Kramer, “Thinking the Unthinkable – Doing Away with the Library Catalogue,” *Insights: The UKSG Journal* 27, no. 3 (2014): 244–48, <https://doi.org/10.1629/2048-7754.174>.

137 Interview with Coen Wilders, head of academic services Utrecht University Library, April 15, 2020.

138 Kortekaas and Kramer, “Thinking the Unthinkable,” 248.

139 Peter Boyce et al., “How Electronic Journals Are Changing Patterns of Use,” *The Serials Librarian* 46, no. 1–2 (2004): 121–41, https://doi.org/10.1300/J123v46n01_14.

notably Google. The technology firms introduce a power asymmetry here, as Google is providing search tools on which librarians and scholars have no influence, but that are so enticing that they push out the existing search systems. Within this trading zone then, librarians are challenged to take on new roles.

Google Books also launched in 2004, then under the name of Google Print. Google Books shifted the practice and purpose of mass digitisation to such an extent that one might ask whether the earlier efforts could rightly be called mass digitisation. Before Google, digitisation efforts emphasised precision and preventing duplication. Librarians were able to complete projects in which a million pages were digitised, but Google promised to digitise 4.5 billion pages in a period of six years.¹⁴⁰ In other words, “it took the most aggressive and technologically advanced library digitizers a decade to scan less than what Google was able to scan each week.”¹⁴¹ Preferring speed over precision, Google Books contains a lot of books, but with a lot of errors. Users in general, and historians specifically, were critical (and still are) about technical issues of quality in the scans, the metadata, or wrongly stated copyrights.¹⁴² Research based on the digitised sources has consequently been characterised as investigating a historical record that never existed.¹⁴³ Furthermore, Google Books was found to contain a surplus of academic rather than popular literature, diminishing its value in representing a society or time period.¹⁴⁴

Google Books was, moreover, criticised for socio-political issues. Robert Darnton, the director of the Harvard University Library between 2007 and 2016, criticised Google Books for establishing a monopoly, since Google was the only one to possess the means for such large-scale digitisation and copyright related trials in court.¹⁴⁵ Furthermore, Google did this as a for-profit company, not as a library whose purpose is to provide access to knowledge. In Europe, Jean-Noël Jeanneney, the president of the *Bibliothèque nationale de France* between 2002

140 Deegan and Sutherland, *Transferred Illusions*.

141 Alissa Centivany, “The Dark History of HathiTrust,” *Proceedings of the 50th Hawaii International Conference on System Sciences* (2017), 2361.

142 E.g., Robert Townsend, “Google Books: What’s Not to Like?,” American Historical Association blog (2007).

143 Johan Jarlbrink and Pelle Snickars, “Cultural Heritage as Digital Noise: Nineteenth Century Newspapers in the Digital Archive,” *Journal of Documentation* 73, no. 6 (2017): 1228–43, <https://doi.org/10.1108/JD-09-2016-0106>.

144 Eitan Adam Pechenick, Christopher M. Danforth and Peter Sheridan Dodds, “Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution,” ed. Alain Barrat, *PLOS ONE* 10, no. 10 (2015), <https://doi.org/10.1371/journal.pone.0137041>.

145 Robert Darnton, *The Case for Books: Past, Present, and Future* (New York: PublicAffairs, 2009).

and 2007, criticised Google Books for imposing Anglo-Saxon cultural values and knowledge.¹⁴⁶ He argued European cultural heritage should not depend on American industries for preservation and access. These authors consequently pushed for public competitors to Google Books, respectively the Digital Public Library of America and Europeana.¹⁴⁷

Some of these criticisms can be explained by the observation that these mass digitisation efforts were pushed mainly by computer scientists and engineers, whereas the efforts around microfilm 60 years earlier were pushed by librarians. As such, these efforts arguably constitute trading zones including technologists, librarians and expected users. I might again speak of a power asymmetry with powerful technology firms, as other projects “are overshadowed by mass digitization, whose intoxicating claims appear to fuel our voracious appetite for digital media, making us ever more impatient of obstacles to the seamless integration of content with commercial search engines – and ever more reluctant to engage closely and critically with what we find electronically.”¹⁴⁸ While the boundary work of librarians traditionally emphasised precision of material and metadata and carefulness to prevent duplication, the values of speed and efficiency prevailed in mass digitisation projects. Furthermore, the speed of mass digitisation limited the material that could be digitised. For example, medieval manuscripts required much more careful handling, requiring different practices of digitisation, thereby creating a bias for certain types of library sources.¹⁴⁹

Libraries had incentives to participate though, as Google did not keep the books to themselves. Participating libraries received digital copies of the scanned books that they were free to distribute for non-commercial use. The University of Michigan Library was the first to join Google’s efforts, following several reasons that together led to the decision to collaborate.¹⁵⁰ One reason was that Google would cover the costs and would return digitised books within an alluring time frame. Second, collaborating with Google was deemed to possibly increase the university’s reputation. Third, digitisation was deemed moral, to make the collection accessible to society. The final two reasons show that while I might speak of a trading zone of librarians and technologists, librarians were already rather aligned

146 Jean Noël Jeanneney, *Google and the Myth of Universal Knowledge: A View from Europe*, trans. Teresa Lavender Fagan (Chicago: University of Chicago Press, 2007).

147 Nanna Bonde Thylstrup, *The Politics of Mass Digitization* (Cambridge, MA: The MIT Press, 2018).

148 Deegan and Sutherland, *Transferred Illusions*, 160.

149 Andrew Prescott and Lorna Hughes, “Why Do We Digitize? The Case for Slow Digitization,” *Archive Journal*, 2018.

150 Centivany, “The Dark History of HathiTrust.”

with the aims of Google: mass digitisation was deemed inevitable, as something that libraries were just supposed to do. Finally, librarians wanted to make a statement regarding copyright. The University of Michigan Library therefore set out to digitise its entire holdings, while other participating libraries digitised material that was out of copyright.

Issues surrounding copyright eventually led Google into several court cases. While Google ultimately won the legal battle, the Books project had lost its momentum and currently does not seem to receive significant attention for further development anymore, leaving “a database containing 25-million books and nobody is allowed to read them.”¹⁵¹ Yet the feeding back of digitised material challenged librarians to develop their own digital infrastructures. The librarians at the University of Michigan soon recognised they needed to collaborate with other institutes to be able to develop and maintain a sufficiently powerful digital infrastructure. Consequently, while Google Books’ development stalled, libraries formed national infrastructures such as HathiTrust in the US (established in 2008) and Delpher in the Netherlands (established in 2013).

Some worried early on about the sustainability of Google Books. As an alternative, the Internet Archive, established in 1996, announced the Open Content Alliance (OCA) in 2005 as a consortium effort, including Microsoft amongst others. Despite this different institutional structure, the two efforts ended up being not too dissimilar in procedure and results.¹⁵² Both services “black boxed” the exact procedures of digitisation. Both permitted the libraries that provided the original works to redistribute the digitised material for non-commercial use. Finally, both offered a web interface to read the books. Yet comparing Google Books and the OCA provides insight into the flexibility digitisation allows for functionality. As both projects focused on scanning books, the procedures for handling and scanning were consequently similar. However, the databases of the two projects are very different, leading to significantly different practices.

Both Google Books and the OCA provide full-text search within a book in the web interface, but only Google Books provides full-text search on the entire collection of books.¹⁵³ While the OCA maintained a relatively classic model of searching by metadata and reading a book by flipping through the pages, Google

151 James Somers, “Torching the Modern-Day Library of Alexandria,” *The Atlantic* (2017). Interestingly, after years of silence, the Google Ngram Viewer received an update in 2020 to include data up to 2019; “Google Ngram Viewer”, accessed May 12, 2021, <https://books.google.com/ngrams/info>.

152 Kalev Leetaru, “Mass Book Digitization: The Deeper Story of Google Books and the Open Content Alliance,” *First Monday* 13, no. 10 (2008).

153 Leetaru.

enabled entirely new forms of interacting with books. First and foremost, Google Books allows a user to search for a specific question and find a single passage in a book that answers this question, without needing to read the rest of the book. At this level, some scholars have criticised Google Books, and digital libraries in general, as providing something that is similar to libraries, but worse. It is debatable whether it is desirable that people search for bits and pieces within books, rather than consulting a book as a comprehensive work in itself.¹⁵⁴ Moreover, keyword search assumes a scholar already knows what they are looking for and only retrieves those relevant parts. This leaves historians to worry about the loss of the context of the library, as well as of serendipity as an important factor in knowledge discovery.¹⁵⁵

Since Google keeps the full-texts of collected books in a database, in contrast with the OCA, this furthermore allowed new forms of research on the entire corpus. This has famously been demonstrated with the Google Ngram Viewer.¹⁵⁶ In this approach, the full-texts of books are used to investigate the texts through n-grams, where “n” refers to the length in number of words that follow one another in a text. For example, “archive” is a 1-gram, “digital history” a 2-gram, and “the history of infrastructures in” a 5-gram. This way the development of a specific term, or a combination of terms, can be analysed over a long period of time, and compared with the evolution of other terms (see Figure 1). This user interface

154 Vaidhyanathan, *The Googlization of Everything*; in this sense, the understanding of books as being valuable for containing information has been argued to reflect a scientific rather than humanistic perspective. The information scientist Ronald E. Day argues: “The contrast between science and humanities scholarship, when it does exist today, at least in terms of hermeneutics, is between documents as containers of information, which are consulted for the information that they representationally contain, and texts, understood through close readings and a type of understanding that involves both a bridging of hermeneutic horizons and a critical and sometimes formally performative questioning of their topics by the style of these very texts”. Google Books then elevates a scientific understanding of books. *Indexing It All: The Subject in the Age of Documentation, Information, and Data*, History and Foundations of Information Science (Cambridge, Massachusetts: MIT Press, 2014), 24.

155 Kim Martin and Anabel Quan-Haase, “Are E-Books Replacing Print Books? Tradition, Serendipity, and Opportunity in the Adoption and Use of e-Books for Historical Research and Teaching,” *Journal of the American Society for Information Science and Technology* 64, no. 5 (2013): 1016–28, <https://doi.org/10.1002/asi.22801>; Kim Martin and Anabel Quan-Haase, “The Role of Agency in Historians’ Experiences of Serendipity in Physical and Digital Information Environments,” *Journal of Documentation* 72, no. 6 (2016): 1008–26, <https://doi.org/10.1108/JD-11-2015-0144>.

156 Erez Aiden and Jean-Baptiste Michel, *Uncharted: Big Data as a Lens on Human Culture* (Riverhead Books, 2013); “Google Ngram Viewer”, accessed May 12, 2021, <https://books.google.com/ngrams>.

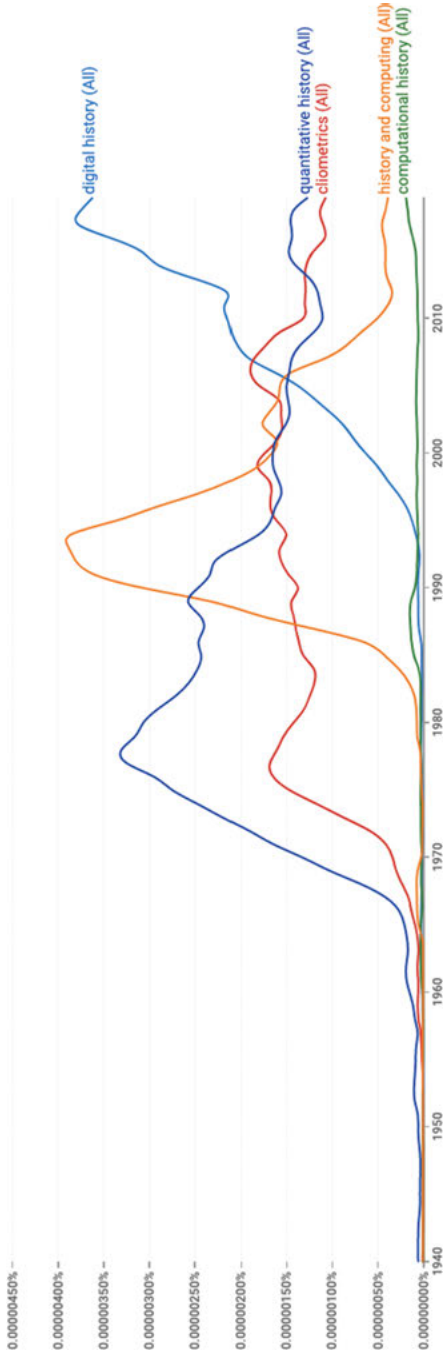


Figure 1: Google Ngram Viewer comparing occurrences of digital history, quantitative history, cliometrics, history and computing and computational history between 1940 and 2019. This chart shows in one simple overview the rise and fall of communities preceding digital history, notably quantitative history and history and computing.

has since been replicated for other text corpora, such as the Dutch National Library's newspaper corpus.¹⁵⁷ These differences in functions are notably not because the OCA could not offer similar functions to Google Books in theory, but because the OCA chose not to offer such functions; a decision that may have been informed by considerations of technological feasibility or path-dependency.

Mass digitisation contains interesting similarities as well as contrasts to the previously described practices of microfilming. Both efforts seemingly have similar goals and require a person in combination with a machine to transform a document into another format. This new format then requires, or allows, new practices for consultation. As a contrast, where microfilm was a good technology for preservation, but not optimal for distribution and access, Google Books and the OCA are instead good for distribution and access, but essentially bad at preservation.¹⁵⁸

Yet Google Books took digitisation even further, into datafication.¹⁵⁹ That is, Google Books offered functionality beyond a digital surrogate of the original object. Google aggregated the collections of books from multiple libraries and turned this into one big dataset of words (or n-grams), which then facilitated new practices. Such datafication of humanities sources underlies much large-scale analysis in the digital humanities, with prominent scholars calling for macroanalysis or the more commonly used term distant reading.¹⁶⁰ In these approaches, scholars are challenged not to “close read” the sources one by one, but to provide an overview of the corpus, and with statistical analyses contextualise data points. In the terminology of Schellenberg, Google transformed books to give informational value; not as information containers in themselves, but as containers of language that signify the societal context within which they were written, published and maintained.¹⁶¹

This is not to say Google Books caused these approaches. The term “digital humanities” was coined in 2004, and “distant reading” was coined by the literary historian Franco Moretti in 2000 already, four years before Google Books and ten years before the Google Ngram Viewer.¹⁶² Still, Google Books fit in what, in the

157 “PoliticalMashup KB ngramviewer”, accessed May 12, 2021, <http://ngramviewer.kbresearch.nl/>.

158 Deegan and Sutherland, *Transferred Illusions*; Leetaru, “Mass Book Digitization.”

159 Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt, 2013).

160 Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History* (University of Illinois Press, 2013); Franco Moretti, *Distant Reading* (Verso Books, 2013).

161 The digital humanist Frédéric Kaplan critically argued that Google not only datafied but essentially commodified linguistic expression, using linguistic data to develop and improve sales of advertisements. “Linguistic Capitalism and Algorithmic Mediation,” *Representations* 127, no. 1 (2014): 57–63, <https://doi.org/10.1525/rep.2014.127.1.57>.

162 Franco Moretti, “Conjectures on World Literature,” *New Left Review*, no. 1 (2000): 54–68.

terminology of the philosopher of science Thomas Kuhn, has been called the fourth paradigm: research based on data-intensive computing.¹⁶³ While Google Books is not the cause of this turn to data-intensive humanities research, it did arguably make it more prominent.

With such large-scale datasets and digital methods, it has been argued that historians can return to *longue durée* historical investigations.¹⁶⁴ For example, the historian Jo Guldi experimented with the Google Ngram Viewer and other databases to investigate the history of walking over three centuries, and especially the apparent rise of walking between 1800–1850.¹⁶⁵ Such investigations require counting of terms over long periods of time, introducing issues of concept drift.¹⁶⁶ That is, the meaning of terms may change over time and context to describe different concepts, or other terms might be used to describe the same concept. Linguistics is consequently an important subject of digital history. This is arguably a continuation of the linguistic turn that started with the cultural turn described above, embedded in computational technologies.

The computational linguistic approach to large unstructured datasets requires expertise that is not part of the work of historians and, therefore, invites collaboration with computational linguists.¹⁶⁷ Furthermore, the subsequent systems required to store and provide access to this data and the user interfaces to retrieve and consult this data in whatever form require computational expertise in knowledge modelling, database design, user interface design and human-computer interaction. The digital infrastructures of digital history thus require cross-disciplinary collaborations on a level not seen before in archives and libraries.

163 Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 2nd ed., International Encyclopedia of Unified Science Foundations of the Unity of Science (Chicago University Press, 1994); Tony Hey, Stewart Tansley and Kristin Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, 2nd ed. (Microsoft Research, 2009).

164 Guldi and Armitage, *The History Manifesto*.

165 Jo Guldi, “The History of Walking and the Digital Turn: Stride and Lounge in London, 1808–1851,” *The Journal of Modern History* 84, no. 1 (2012): 116–44.

166 Shenghui Wang, Stefan Schlobach and Michel Klein, “Concept Drift and How to Identify It,” *Web Semantics: Science, Services and Agents on the World Wide Web* 9, no. 3 (2011): 247–65, <https://doi.org/10.1016/j.websem.2011.05.003>.

167 Michael Piotrowski, *Natural Language Processing for Historical Texts*, ed. Graeme Hirst, *Synthesis Lectures on Human Language Technologies*, vol. 5 (Morgan and Claypool, 2012), <https://doi.org/10.2200/S00436ED1V01Y201207HLT017>; Barbara McGillivray, Thierry Poibeau and Pablo Ruiz Fabo, “Digital Humanities and Natural Language Processing: Je t’aime . . . Moi Non Plus,” *Digital Humanities Quarterly* 14, no. 2 (2020).

As the case of Google Books demonstrates, the way the database is structured affects what a historian can do with the data. A full-text search on the level of a book is significantly different from a full-text search on the level of the entire library. How data is processed may introduce biases or limitations not immediately recognisable to historians.¹⁶⁸ On top of this, the user interface affects even further what a historian can do with the data; a search box returning a list of results is a significantly different tool than an Ngram Viewer, leading to different questions. The archival turn led to an understanding of archives as structuring the perspective on the past. Likewise, the user interfaces of archives and libraries act as an interface to the past, shaping perspectives on the past.¹⁶⁹ The infrastructures directly influence the possible practices of historians and the knowledge that may be generated. Therefore, historians are collaborating in digital history projects with the goal of steering these infrastructures into directions suitable for historians. It is through these collaborations that historians, computational linguists, computer scientists, archivists and librarians negotiate how digital infrastructures will facilitate future practices of historians.

To understand how these negotiations takes place, I develop a model to critically examine collaborations as trading zones in which concepts, methods and tools are shared and exchanged. In the next chapter, I elaborate this model and discuss how I apply the trading zones concept to digital history collaborations. Readers who prefer to skip directly to my studies of digital history collaborations may instead prefer to move on to Chapter 3. In the third chapter I examine digital history collaborations at the University of Luxembourg by means of the first dimension of my model, engagement, to consider how historians engage with one another and with cross-disciplinary collaborators.

168 Antske Fokkens et al., “BiographyNet: Methodological Issues When NLP Supports Historical Research,” in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, ed. Nicoletta Calzolari et al. (Reykjavik, Iceland: European Language Resources Association (ELRA), 2014), 3728–35.

169 Margaret Hedstrom, “Archives, Memory, and Interfaces with the Past,” *Archival Science* 2, no. 1 (2002): 21–43, <https://doi.org/10.1007/BF02435629>.