When the steam printing press appeared in the early nineteenth century, using machine-made rolls of paper instead of handmade sheets, it sparked a wave of utopian expectations and visions. This technological innovation, contemporaries pointed out, was causing a revolution in communications, transforming the availability of knowledge. For as little as a penny, anyone could buy the most recent scientific knowledge and, ultimately, change the foundations of society. In Britain, the Society for the Diffusion of Useful Knowledge was founded in 1826 with the explicit aim to ‘leave nothing undone, until knowledge has become as plentiful and as universally diffused as the air we breathe’.1

The similarities between that communication revolution and the one we have experienced in the last twenty years or so are in many ways striking. The expectations and hopes invested in digital technology and in the new ways of disseminating knowledge are readily apparent in our daily lives.2 In academia, the innovations are starting to shape methodologies and theoretical approaches. Yet while many historians encounter digital resources, databases, visualizations, and narratives in their work, on the national scene the topic of digital history is still relatively unexplored, and digital history projects have been few and far between.3 This article therefore sets out to give an introduction to some of the major themes in the discussions about digital history, and to suggest some further reading.4

The digital humanities and digital history
Internationally, the field of the digital humanities has experienced rapid growth, to the point where for a while now it has been ‘the first “next big thing” in a long time’.5 In fact, some commentators even talk of a backlash, caused by a reaction to the shift in priorities, required skills, and award systems. One aspect of this concerns postgraduate students, who are less interested in the digital aspects and more passionate about the humanities, while another challenge stems from researchers who feel that budgets are decreasing due to the prioritization of digital humanities projects over more traditional research.6 Nevertheless, on a national level, the digital humanities
seem to be in a rather expansive phase, as is demonstrated by the rise of
digital humanities centres, programmes, and courses, and the promotion
of projects dealing with digital infrastructure.

Defining the digital humanities is an academic meta-genre in its own
right, and here I will limit myself to some brief introductory remarks.
Generally speaking, many scholars and professionals agree with the basic
assumption that the digital humanities exist because academia can no longer
isolate its practices and methodologies from the changes caused by what
has been called the computational socioeconomic revolution.7 The authors
of the pioneering *A Companion to Digital Humanities*, published in 2004,
similarly claim that the field grew out of an explicit interest in using all
available forms of information technology in the humanities. The goal was
twofold: to use information technology to shed light on the human record,
and to let the development and use of information technology be guided
by the human record.8

Originating as it did in the field of humanities computing, the digital
humanities have been seen as an experimental arena where coding meets
critical thought.9 Patrik Svensson, professor of the humanities and information
technology, has emphasized that the digital humanities are a ‘trading zone
and meeting place’, and that we need to critically assess what is included
in the ‘big tent’ and how it is ‘epistemologically textured’.10 Others point
to the tactical and populist nature of the term and how it is ‘unabashedly
deployed to get things done—“things” that might include getting a faculty
line or funding a staff position, establishing a curriculum, revamping a lab,
or launching a center’.11

Although cross-disciplinary aspirations are a trademark of the digital
humanities, opinions differ on the status of digital history. In many cases,
the digital humanities are associated with linguistics, languages, and literary
studies, deploying methodologies and practices that are more suited (and
actively adapted) to these disciplines. As scholars have rightly noted, the
digital humanities rarely include digital history, and digital history might even
benefit from distancing itself from the ‘big tent’ of the digital humanities.12

Although the digital humanities and digital history have a lot in common
in terms of their historical development, there are longstanding differences
in the ways studies are conducted in the various fields. The background to
digital history is sometimes sought in cliometrics and quantitative history,
and with the attempts by the social history movement in the sixties and
seventies to use computers for statistical analyses of demographic data.13
Others claim that digital history has evolved at the intersections between
‘gathering, preserving and presenting the past on the web’,14 meaning that
it has been less occupied with computing per se and more influenced by the
practices of collecting, presenting, and disseminating material, linking it to the (digital) efforts and practices of museums, libraries, and archives.\textsuperscript{15}

When trying to outline digital history, then, a couple of different factors need to be borne in mind. The first is the fact that the historian can be both a consumer and a producer of data, information, and knowledge.\textsuperscript{16} As a consumer, the historian is confronted with digitized online resources which must be assessed critically before they can be used; as a producer, the historian can use digital media to present and disseminate data, information, and knowledge to online audiences—which, of course, raises the question of how the medium itself affects the historical narrative and the production of knowledge, and how digital media can best be used for didactic purposes. The third factor is the use of computational methods to analyse historical data. Since the technical and methodological solutions used by different branches of history vary significantly, the present article will focus on some recent debates concerning big data and the challenges and opportunities presented by the way computational methods are envisaged by many funding agencies today.

Digital historical resources and their challenges

The past twenty years have seen an exponential rise in the number of resources that are digitally available. Thanks to mass digitization by both private and public actors around the world, we now have access to an immense amount of textual, visual, audio, and audiovisual data. Whether students, researchers, or university lecturers, everyone can turn to sources available online for their work, and many of the websites offering such materials are specifically addressed to an academic audience. In a comprehensive guide to history in the digital world, the historians Jessica Parland-von Essen and Kenneth Nyberg have pointed to the difference between digitized versions of primary sources and born-digital materials—a distinction between the main types of digital historical resources that is common among archivists, but more rarely used by historians.\textsuperscript{17}

Digitized materials come in a variety of forms and thus pose different challenges. They can be scanned images available in PDF format or through a viewer of some kind, (hopefully) human-readable. As some have pointed out, however, the mass digitization of source material by Swedish archival institutions has in many cases prioritized quantity and automation before quality. This means that it has been common for microfilms to be automatically scanned rather than the original source material. Recurring problems with some of these older digitized documents are their lack of sharpness and/or of greyscale, making them hard to decipher or tiring to read for longer
periods of time. More recent digitization projects, however, have aimed for higher standards when it comes to resolution, colour depth, and sharpness, making possible not only reading but also downloads and reproduction. A step further in making material available for research comes with the conversion of digitized text into machine-encoded text using optical character recognition (OCR) technology, which produces a searchable text that can then be copied and edited. Typically this applies to typed or printed texts, but can also be used for handwritten material.

A recurring issue is the searchability of the digitized corpora. Searchability, as Parland-von Essen notes, turns on the accuracy and comprehensiveness of the metadata and its contextual information. Metadata, which has been defined as ‘structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource’, is often called data about data, and can be compared with the information given by classic library card catalogues. When it comes to digital materials, the standard varies significantly——and trustworthiness and reliability with it.

Born-digital materials present a different kind of challenge to the historian than digitized ones. Born-digital sources are defined by the fact that they originate in digital form and have no analogue original. The obvious examples are websites, online forums and communities, and wikis, but online newspapers, e-books, Internet-disseminated television shows, electronic records, and digital sound recordings also fall into this category. Born-digital materials are often discussed in relation to issues of preservation and storing, but also, of course, in terms of availability and accessibility, given the ever-changing nature of the Web. These materials are often determined by the specific hardware that was used to make them, and their formats may not be sustainable. Equally, the metadata and contextual information of born-digital sources are often lacking or poor in quality. Responding to such challenges, archives and libraries are currently taking on a more active role in the preservation of born-digital materials.

A key question for all resources available through digital media is how they are digitally represented. When applying digital technologies to historical resources and narratives, it is necessary to represent the data of interest in a machine-readable form. In order to enable computers to identify, locate, and manipulate the data in its programs and applications, data and its relation to other data needs to be identified, defined, and named. When it comes to texts, there are several ways of making them machine-readable, whether by using OCR or one of the more sophisticated representations such as those specified by the Text Encoding Initiative (TEI), a non-profit international organization that provides encoding guidelines for machine-readable texts, developed with the humanities, social sciences, and linguistics in mind.
Inevitably, there are pitfalls for the historian using online historical resources, whether found through databases with digitized or transcribed records or in the form of contemporary websites, communities, social forums, or wikis.\textsuperscript{24} Can we then use the same methods and critical perspectives when it comes to digital resources, or is it necessary to develop specific digital competences to deal with this new medium? The historian Andreas Fickers has highlighted the challenge in making use of “the millions of interesting sources on the web without drowning in a flood of entertaining yet historically irrelevant information”.\textsuperscript{25} He singles out the importance of critically examining the metadata, which he argues is crucial for source criticism and the contextualization of the information provided by a source. A related question is that of authenticity. To his mind, the democratization of access has disrupted ‘long-evolved systems of trust and authenticity, ownership and preservation’, leaving the usefulness of digital historical resources reliant not only on the trustworthiness of the providers and the existence and accuracy of the metadata, but also on the quality and aptness of the databases, search engines, and algorithms used in the digital environment where the information is found.\textsuperscript{26} The critical assessment of digital sources is consequently a skill that now needs to extend to understanding both the origins and the structure of the data, information, and knowledge presented by the digital tools, online databases, and visualizations.\textsuperscript{27}

**Historical knowledge and the digital media**

Most historians are likely to use some form of digital media to present the results of their research, if only a word processing or presentation program or open access publishing. These forms are essential to academic writing today and do not require any substantial change in methodology or narrative style. When it comes to gathering data from online databases and presenting research findings, whether as visualizations or even hypertext, the methodological and narrative impact is of a different magnitude. The challenge of presenting histories in the digital world is really also the challenge of collaboration, not only with fellow historians, but with museums, archives, libraries, and, of course, with the developers needed to turn the historians’ ideas, text, and findings into code.

One aspect of this is the process of digitization and the subsequent creation of digital collections, hosted by either universities or archival institutions. Archivists frequently have to consider the implications of digitization as it touches upon the problems of conservation and preservation, but, as some point out, there has been a striking reluctance (and even resistance) on the part of professional historians to take part in these discussions.\textsuperscript{28} One might
also interpret the lack of interaction as resulting from the professional divide between historians and archivists/librarians, a divide that sometimes results in an unwillingness on both sides to invite the other to join the conversation. 29 To many observers, however, there is an obvious need for historians to take part in the process of digitization and the subsequent presentation and dissemination of digital resources.30

Another challenge stems from the aforesaid difference between the digital humanities and digital history. While the digital humanities, and especially linguistics and languages, have been at the forefront of adapting digital tools to their fields, historians have shown less interest. Johanna Drucker, a leading figure in digital aesthetics, underlines the importance of engaging in the design of digital environments and platforms that embody the theoretical and methodological premises of the humanities, because, as she says, ‘Humanistic methods are necessarily probabilistic rather than deterministic, performative rather than declarative’.31 The historian Edward Ayers touched on something similar in 1999, when he commented on the ‘Enhanced teaching, professional community building, experiments in hypermedia, and impressive digital archives’ that were the fruits of the first digital historians’ work. At the time, he felt that it was not proven that historians could ‘create forms of narrative and analysis that adequately exploit the possibilities’ offered by the new media and technology.32 Ayers question still holds good today.

One often-quoted suggestion for how to use the new media to improve historical narrative was made by the historian Robert Darnton. In ‘A Program for Reviving the Monograph’, Darnton admits to sounding utopian, although his aim was pragmatic. ‘The electronic space is out there, waiting to be filled by something more substantial than the junk produced by the consumer industries’, he wrote. ‘It has room for a new kind of publication, one that will not replace the book but that will revive it and send it into orbits beyond the galaxy of Gutenberg.’33 Darnton’s suggestion is to structure electronic publications into different layers, taking into account the readers’ various interests and needs for specialization. Thus the first layer might consist of the concise account of the subject while the second layer presents the arguments in greater depth; a third layer could include documentation; a fourth, historiographical information with discussions of previous scholarship; a fifth, a didactic component, with suggested topics for classroom discussion; and, as the sixth layer, a crowd-sourcing function with reader’s reports and comments.34

Following Darnton, Ayers suggests that digital history could help us overcome the limitations of simple narrative and monographic abstraction—‘We might try writing in more self-conscious ways, manipulating point of view,
chronology, and voice more than in our current practice.’ Digital history could thus be a way of engaging with the complexity of the past and exploring a more concise narrative style. Ayers goes on to identify further opportunities offered by the new media; for example, writing hypertextual narratives with interlinked texts and documents, offering more precise arguments, evidence claims, and associations. He concludes that this would permit more evolved and dynamic discussions of our narratives and references, allowing for new commentaries and new connections to be made.35

When exploring different kinds of narratives for purposes other than scholarly writing, a broader approach seems promising. Helyom Viana Telles, for example, highlights digital games as quasi-history works and argues that their relationship to the past is worth examining. ‘What can be learned about history from the interaction with digital games, and how?,’ she asks.36 Claudio Fogu has also emphasized the importance of digital games, claiming that they are ‘pushing the processes of de-temporalization and de-referentialization of history toward the formation of a new notion of the historical’. Evoking the classic Aristotelian paradigm, he goes on that ‘history has replaced poetry and philosophy as the realm of the possible’.37 Telles shares this view, and chooses to view the fictitious not as false, but as rooted in the sphere of the real and the likely. ‘Thus, for historians, the resource of fictional digital world creation may represent an effective form of expressing knowledge and representing history that offers a positive response to the challenges posed by the post-modern critique of historiography’, Telles concludes.38 Other scholars have underscored the role of games that enable gamers to take on active roles while navigating complex historical processes.39

Augmented reality (AR) games or apps, which give a live view of real-world environments where elements have been augmented by computer-generated input in the form of sound, video, graphics, or GPS data,40 have also been used as didactic tools. Karen L. Schrier is certain of the impact of AR games, and suggests that they can enhance the learning of (1) historical name, places, and themes; (2) historical methodology and the limits to representations of the past; and (3) alternative perspectives and challenges to “master” historical interpretations.’ If nothing else, the AR games motivated the participants of her survey to ‘gather, evaluate, and interpret historical information, devise hypotheses and counter-arguments, and draw informed conclusions.’41

Digital augmentation is, of course, not limited to games and applications that we consciously use; it also surrounds us in our everyday lives in urban spaces. Phone apps with geographical, social, and commercial information can accompany us wherever we go, producing a specific experience of place.
Given the impact of digital augmentation, scholars have therefore called attention to the power relations hidden in the digital and coded information that shape our experiences, highlighting how the geographically referenced content obtains ‘a persuasive rhetoric of authority’, mediated through the technological processes.42

The implications of big data

One of the more influential ways of envisaging computational approaches today is connected to what is known as big data. Data sets that are too big to be stored and processed using conventional methods generally go by the name big data, but the term also implies a certain kind of quantitative analytics, for they are sometimes used in predictive analytics—analyses that aim to predict the future using large sets of data on current and historical events. Big data has become a buzzword for many research programmes, not least the EU-funded Big Data Public Private Forum and the EU research and innovation programme Horizon 2020.43 The role of history as a discipline might seem marginal to these types of programmes; however, historical data are anything but peripheral, and the question should rather be which discipline is the most suited to processing and interpreting the data. Initiatives to address this, and the question of how big data can be used by historians, are now underway.44

The marginalization of historians when it comes to the processing of big data can be traced back to the diminishing role of quantitative social science history and the influence of the ‘linguistic turn’, which from the 1960s onwards has seen the historian’s focus shift away from sociology and towards anthropology. ‘Rather than SPSS guides and codebooks, innovative historians carried books of French philosophy and German literary interpretation’, Ayers writes, maintaining that the first computer revolution largely failed.45 The ‘linguistic turn’ still exerts a considerable influence over humanities research, while the questions associated with big data processing are generally raised from different viewpoints. However, as the historian James Grossman emphasizes, there is no need to recast historians as statisticians; it is sufficient to recognize that historical narratives offer a way of organizing and presenting big data as meaningful information.46

The historian of history teaching Thomas Nygren draws our attention to the dangers associated with these kinds of perspectives and their claims. The concept of data, not to mention visualizations such as tables, charts, and maps, may create the illusion of objectivity, thus ignoring or giving legitimacy to power relations.47 The sociologist Tressie McMillan Cottom, meanwhile, sees the inherent logic in these large-scale projects, claiming
that ‘Data-tizing literature at large scale becomes meaningful not because of its ontological superiority per se but because it rationalizes the hegemonic cultural imperative that all things (and beings) be data-tized’, so that the big data seem to exert ‘normative and economic power over the questions scholars ask and how they set out to answer them’.\textsuperscript{48} She also emphasizes that the data models inherited from commercial apps deal with power relations such as race and gender in an unsophisticated and simplistic way, making it hard to trust the results.

\textbf{Final reflections}

When it comes to new technology in general, both prophets and sceptics seem to be a self-evident part of the process. The dangers are frequently highlighted, as is the liberating potential of the innovation in question. The same kind of fears and hopes have been seen before—think only of the advent of the mass printing and distribution of books in the early nineteenth century. Yet although the similarities might seem apparent, the question remains of how best to interact with and use the new medium. How does it affect our research practices and how can we take advantage of its possibilities as producers of information and knowledge?

Digital history spans several very different activities, united by the fact that they all deal with digital media in one way or another. The consumption of digital history calls for an awareness of how digital processing and presentation affects the artefacts found online. Producing digital history is what the historian William G. Thomas referred to as creating ‘a framework, an ontology, through the technology for people to experience, read, and follow an argument about a historical problem.’\textsuperscript{49} This calls for an interest in the technological and didactic questions raised by the new media, and for a sincere collaboration with people who have the necessary competence to present and disseminate the historical narrative in digital form. The use of computational methods to analyse large amounts of data raises questions of an epistemological as well as methodological nature. Digital resources present us with several challenges, encouraging us to evaluate what we take for granted and to adjust our skills to the changing circumstances.

\textbf{Noter}

2 For Internet sceptics and evangelists, see Daniel J Cohen & Roy Rosenzweig, \textit{Digital History: A Guide to Gathering, Preserving and Presenting the Past on the Web} (open access}


9 Miller 2015.


14 Robertson 2014; Cohen & Rosenzweig 2015.


16 The conceptual trio of data, information, and knowledge (sometimes with the addition of wisdom, creating what is often referred to as the DIKW pyramid) is by no means unproblematic. These three concepts are frequently used in relation to one another, knowledge being defined by its relation to information, information by its relation to data. Data as a concept is usually taken to be the least abstract, used to describe the parts of which information is comprised. Actionable information might in turn be called knowledge. Although these concepts have been criticized on several accounts, I have chosen to use them here to indicate a hierarchical relationship. For a critical perspective, see Martin Frické, ‘The Knowledge Pyramid: A Critique of the DIKW Hierarchy’, *Journal of Information Science*, 35/2 (2009), pp. 131–42.


20 Nyberg & Parland-von Essen 2014, pp. 56 & 60.
28 Fickers 2012.
34 Darnton 1999; Ayers 1999.
35 Ayers 1999.
38 Telles & Alves 2015.
42 Mark Graham, Matthew Zook & Andrew Boulton, ‘Augmented reality in urban places: contested content and the duplicity of code’, *Transactions of the Institute of British Geographers*


45 Ayers 1999.


