DHQ: Digital Humanities Quarterly
2016
Volume 10 Number 2

Digital Humanities in the 21st Century: Digital Material as a Driving Force

Niels Brügger <nb_at_dac_dot_au_dot_dk>, The Centre for Internet Studies, and NetLab Aarhus University

Abstract

In this article it is argued that one of the major transformative factors of the humanities at the beginning of the 21st century is the shift from analogue to digital source material, and that this shift will affect the humanities in a variety of ways. But various kinds of digital material are not digital in the same way, which a distinction between digitized, born-digital, and reborn-digital may help us acknowledge, thereby helping us to understand how each of these types of digital material affects different phases of scholarly work in its own way. This is illustrated by a detailed comparison of the nature of digitized collections and web archives.

1. The Humanities – What's the Problem?

For decades the challenges and crises faced by the humanities have had an impact on debates about their status and future. In his seminal article "Beneath and Beyond the 'Crisis in the humanities,'" Geoffrey Galt Harpham even talks about "the perennial crisis in the humanities" that has characterized the humanities "over the past half-century" [Harpham 2005, 21], and he continues: 1

> Sometimes the crisis — whose dimensions can be measured by declining numbers of enrollments, majors, courses offered, and salaries — is described as a separate, and largely self-inflicted, catastrophe confined to a few disciplines; sometimes it is linked to a general disarray in liberal education, and sometimes to the moral collapse and intellectual impoverishment of the entire culture. But one point emerges with considerable regularity and emphasis: humanistic scholars, fragmented and confused about their mission, suffer from an inability to convey to those on the outside and even to some on the inside the specific value they offer to public culture; they suffer, that is, from what the scholar and critic Louis Menand calls a "crisis of rationale." [Harpham 2005, 21–22]

Most of the signs of crisis highlighted by Harpham have been widely debated [Olmos-Peñuela 2014], [Belfiore 2014], [Parker 2007], [Nussbaum 2010], and other signs have been added, for instance the contraction in public funding [Pascoe 2002], the lack of interdisciplinarity [Bassnett 2002], and the impact of the so-called inhuman on the humanities [Barnett 2014]. 2

These characteristics of the humanities may be very true and relevant, but they tend to overlook a major shift that has slowly affected the humanities since the late 1960s and that is now emerging rapidly on a large scale, namely the introduction of "the digital" in the humanities. 3

Within the last decade debates about the digital computer and the humanities have by and large taken place under the umbrella term "Digital Humanities." The humanities have become digital by making the objects of study available in digital form, by introducing digital analytical tools, and by establishing digital means of communication for collaborating during the research process, for discussing and disseminating research results, and for interacting with society at large. A more self-reflexive approach to the Digital Humanities has also emerged, ranging from manifestos such as the *Digital Humanities Manifesto* [Digital Humanities Manifesto] and edited volumes, monographs, and articles that map the field 4

[Screibman], [Berry 2012], [Gold 2012], [Burdick et al 2012], [Warwick 2012], [Borgman 2009], [Svensson 2010], [Deegan 2014], [Svensson 2012a], [Liu 2011], [Jones 2014], to special issues of journals [DHQ e-Science 2009], [Arts and Humanities in Higher Education 2011], [Historical Social Research 2012], [MedieKultur 2014] and reports [American Council of Learned Societies 2006], [European Commission 2011], [Holm 2015]. However, there have not been many attempts to systematically identify (some of) the driving force(s) behind the introduction of the digital in the humanities: if the digital is actually becoming more and more predominant within the humanities, what could be the reason for this?

5 Undoubtedly, a major and irreversible shift has taken place at the very heart of the humanities, since the sources and the data that are studied in many of the humanistic disciplines have changed from analogue to digital. A few figures may illustrate this. In 2000, 75% of all stored data was analogue (paper, film, photographic prints, vinyl, magnetic cassette tapes, etc.), but in 2007 this had shrunk to 7%, and in 2012, to 2% [Mayer-Schönberger 2013, 8–9]. From 1453 to 1503, after the introduction of the printing press, eight million books were printed, thus doubling the amount of written material in the world in 50 years. Today the volume of digital data doubles a little more than every third year, whereas the volume of analogue data hardly grows at all [Mayer-Schönberger 2013, 8–10]. Google processes more than 24PB of data per day, thousands of times the quantity of all printed material in the Library of Congress. Facebook gets +10 million photos uploaded every hour, and over an hour of video is uploaded on YouTube every second [Mayer-Schönberger 2013, 8].

6 Although these figures may be questioned, the tendency is undoubtedly clear: within the last decade we have been witnessing a major shift from analogue to digital material, and we have probably only seen the beginning of this. Thus, it can be argued that one of the major transformative factors of the humanities in the 21st century is the move from analogue to digital source material (obviously, other factors also play a role, cf. [Svensson 2011, 42]). Before the year 2000 the introduction of the digital in the humanities in the form of digitized collections, tools and communication platforms may have been an additional choice, but this may not be the case in the future. Since more and more source material comes in digital form — and in more and more cases in digital form only — there is no opting out of the digital, and the fundamental question for the humanities is probably not *if* the digital should be introduced in the humanities, but rather *how*. And it is worth noting that as a function of the transformation of source material from analogue to digital, the analytical, tools and the means of communicating about scholarly activities will in many cases also have to change, since the use of digitally- supported methods that were previously just a possibility tend to become a necessity when interacting with digital objects of study.

7 If this characteristic of the humanities at the beginning of the 21st century is correct, we must ask ourselves how this shift should be conceptualized, how it will affect the humanities, and to what degree the humanities will become digital.

8 Taking the figures above as a point of departure the figures above, this article will address these general questions by focusing on the core question, namely the nature of digital source material. The main argument is that all kinds of digital material are not digital in the same way, just because they are digital. To support this argument a distinction is introduced between digitized, born-digital, and reborn-digital sources. With this distinction as a stepping stone, the major differences between two examples of digitized and reborn-digital material are illustrated by a detailed comparison of the nature of digitized collections and web archives, and how these differences affect the scholarly use of these sources in various ways.

9 But before going into greater detail about this, it is necessary to reflect briefly on what could be understood by “the digital” and how it relates to the humanities when combined to produce “the Digital Humanities.”

2. Digital, Humanities — Digital Humanities

10 One of the main challenges when investigating the questions above is not only that the “Digital Humanities” is a battlefield of competing definitions — is it a “big tent” [Svensson 2012b, 36] or a “trading zone and meeting place”? ([Svensson 2011, 50–56], — this is also the case with the digital as well as with the humanities. Attempts to define the Digital Humanities often tend to forget what the humanities were before they apparently became “digital,” which adds to the confusion and complexity.

With a view to introducing the understanding of the Digital Humanities that will be used to frame and support the article’s

arguments about digital sources, this section will briefly reflect on “the digital computer” as well as on the main themes that definitions and mappings of the humanities must address, followed by a combination of these two lines of reflection into an understanding of the Digital Humanities. 11

2.1. Digital

It is striking that in the literature about the Digital Humanities not much attention has been focused on “the digital” (e.g. [Laue 2004], [Deegan 2014, 30]. References to the digital tends to focus on the historical development of computers as artefacts and of their use within the humanities, without systematic reflections about “the digital” and the digitality of the computer, which is what characterizes its way of being digital. 12

One exception is Evens, who regards the binary code of 0s and 1s as the point of departure for the development of digital artefacts and culture: “The common element in all digital technologies, the unifying aspect of the cultures, arts, and media that we call *digital*, is the discrete, binary code.” [Evens 2012, 7]. Thus, the binary code is a condition of possibility for technological artefacts (Barry also stresses the digital by focusing on software and drawing on work in software and critical code studies), but without defining the digital [Berry 2012, 4–6]. Evens understands the binary code of 0s and 1s as numbers that can be calculated [Evens 2012], but another possibility would be to understand 0s and 1s as letters, thus displacing the digital from being a matter of mathematics to being a matter of text. The latter approach can be found in Finnemann’s definition of the digital computer. 13

In his conceptualization the digital computer is characterized by three invariant traits: 1) a mechanical alphabet composed of a finite number of letters, each of which is void of meaning, 2) an algorithmic syntax, and 3) an interface which determines the semantics of the syntax [Finnemann 1999], see also [Brügger and Finnemann 2013, 68–69]. The idea that the 0s and 1s can be understood as letters and not numbers is based on the assumption that numbers are not void of meaning — 0 means “zero,” 1 means “one” — but in order to be able to combine them via the algorithmic syntax they must be void of meaning, like other kinds of letters that are combined into words. All three invariant traits are necessary as such, but they are not necessary in any specific form. Today the mechanical alphabet composed of a finite number of letters is the binary alphabet with the two letters 0 and 1, the algorithmic syntax comes in a number of forms, from machine to user level, and the interface can be any kind of device that can handle the digital alphabet and its algorithmic syntaxes, ranging from PCs and handheld devices to mainframe and networked computers. 14

The consequence of this approach to the digital is that on a very fundamental level the digital computer can be understood as a “writing machine,” with the letters being combined to form “texts” on all the levels of the computer, and these texts are editable at any time, down to the individual bit. In line with the point made above by Evens, it is important to include the fundamental digitality of computers in the understanding of the Digital Humanities since it constitutes the condition of possibility of all the concrete forms of the Digital Humanities. 15

2.2. Humanities

It is worth noting that the humanities always come in the plural. Of course, the plural refers to the many humanistic disciplines, but one could also see this as an indication of the fact that “the humanities” is a very varied field, and that there is no common definition of what it is. Nevertheless, it is possible to identify a number of themes that are often addressed in definitions of the humanities. 16

First, the question of boundaries has to be addressed: what is in-/outside the humanities? Is anthropology a social science or part of the humanities? Should, for instance, law be considered part of the humanities or part of the social sciences [Howarth 2004], and is communication a humanistic or social science discipline [Gronbeck 2005]? And how should we perceive new disciplines such as literature and medicine [Bolton 2008], or studies in the borderland between the humanities and the cognitive sciences [Beňuš 2010]? 17

Second, discussions of the humanities tend to revolve around the issue of whether a single, clearly delimited object of study can be identified, as well as a set of predominant theoretical and methodological approaches to be used in all disciplines. 18

Third, the aim of the humanities is often regarded as a defining feature of this field, for instance whether the humanities should contribute to the advance of mankind and of human culture, or be of a more descriptive nature, or be transformed into individual, societal or industrial value. 19

And fourth, the humanities can be delimited based on purely administrative divisions of higher education and research, which is a more formal argument with institutional affiliation determining what the humanities are (the humanities are simply what belong to an institutional entity such as a faculty of humanities). 20

Although there is no precise and concurrent definition of the humanities, these four themes constitute recurring issues in most attempts to define the humanities. Thus, any such definition constitutes a specific constellation of answers to the questions raised within each of the four themes: what are the boundaries, the object(s) of study, the theories and methods, the aims, and the institutional affiliation? And these constellations of answers vary throughout history and may depend on the national setting. Therefore, it is difficult to identify “the humanities” in an unequivocal manner with a view to using this definition as a stepping stone in an understanding of the Digital Humanities. Instead, one must content oneself with the idea that the humanities constitute a fuzzy and complex field, constantly under construction, and that the humanities vary a good deal. And therefore any understanding of the Digital Humanities should start by being very specific about how the humanities that become digital are understood. 21

2.3. Digital Humanities

As mentioned above, one of the greatest challenges facing any attempt to define the Digital Humanities is that neither “Digital” nor “Humanities” comes with a clear-cut definition, and in particular it is worth noting that the fuzziness and complexity of the “humanities” does not disappear or dissolve just because the term is combined with “digital.” On the contrary, the many differences remain and may even multiply, and any attempt to define the Digital Humanities without having defined the humanities will rebound into discussions of what is understood by the humanities. 22

Therefore, it may be a better solution to start with “the digital,” for instance, based on the fundamental definition of the digital outlined above. Such an approach does not solve the complexities of the humanities, but it gives a minimal common ground for debating how the interface between the humanities and the digital may be understood. 23

Thus, the following understanding of the Digital Humanities is suggested as the overall framing of the following discussion of the specific nature of digital sources: Humanities (regardless of what is understood by “humanities”) which to some extent use digital computers and thereby shape the invariant traits of the computer to fit specific scholarly needs, based on how they define themselves as “humanities,” their research questions and their practices. The advantage of this definition is that on the one hand it offers a shared common ground for any detailed definition of the Digital Humanities — that is, the conceptualization of digitality — while on the other hand it is open and flexible enough to incorporate the variety of detailed understandings of the humanities that have been established in theory or by the concrete scholarly practices, and instantiations of the use of the computer. It is important to stress that the interplay between “the digital” and “the humanities” is not regarded as a one-way deterministic logic from “digital” to “humanities.” Instead, it is understood as a dialectical interplay of shaping and re-shaping between “digital” and “humanities” — the two are interdependent. 24

With this understanding of the Digital Humanities, let us now have a closer look at some of the concrete ways in which the invariant traits of the computer — the digital alphabet, the algoritmic syntax and the interface — are (and have been) shaped by the scholarly practice within the humanities. 25

3. Three Types of Digital Material within the Digital Humanities

The nexus between the humanities and the digital is related to the material to be studied in a very fundamental way, in the main because the use of digital analytical tools does not make any sense if the object of study does not somehow exist in digital form. What follows in this section are three approaches to understanding the importance of the different digitalities of digital material. First, a general typology of three types of digital material is presented, based on the assumption that although all kinds of digital material share the same digital alphabet, they are not identical in all 26

respects just because they are digital — digital material is not just digital material. Second, a systematic approach is introduced as to where exactly we can locate “the digital” in the humanistic scholar’s research process. And, third, a brief outline of the historical development of the interplay between the humanities and the digital is presented.

3.1. Digitized, Born-digital, and Reborn-digital Material

27 On the one hand, all kinds of digital material share the same feature, namely that they are digital. But on the other hand, this digitality is always already embedded in a set of semantic, technical and academic structures, which implies that digital material is not just digital, but that it is digital in a variety of ways (a similar argument of embeddedness in already existing structures is put forward in relation to digital tools in [Collins 2011]). In the following section a general typology of digitality is suggested, based on the *provenance* of the digital material, and on a distinction between three main types of digital material: digitized, born-digital, and reborn-digital material.

28 Digitized material is analogue material that has been digitized, for instance, written documents on paper, parchment etc., or electronic media such as radio and television, and even pictures or 3D-models of artefacts. The process of digitization is any form of transformation of analogue material into digital form, be it the laborious keyboarding of written documents to punched cards, the more easily performed scanning of documents to image files, or the digital recording of sound and moving images. The main characteristic of digitized material is that its “becoming digital” is based on an original that was not digital — an original that can in many cases still be retrieved and thus function as a baseline (for reflections on digitized collections, see [Hockey 2000, 11–23], [Terras 2012].

29 Born-digital material is digital material that has never existed in any other form than digital. This includes all types of material on digital media such as CD-ROMs, DVDs, or the internet and the web. This type of digital material does not have an analogue original to go back to, we only have the digital original ([Berry 2012, 4] as well as [Jones 2014] and [Kirschenbaum 2013] also use the term born-digital).

30 Reborn-digital material is the term that is suggested to characterize born-digital material that has been collected and preserved, and that has to a large extent been changed in the process of collecting and preserving. Examples of this are emulated computer games or material in a web archive, the latter being presented in greater detail in section 4 below.

31 Each of the three general types can be subdivided, as the examples above illustrate, and as seen with (for instance) digitized documents such as newspapers that are different from digitized temporal media such as radio or television; just as born-digital material such as computer games on DVDs is very different from online web or apps on iPads (cf. also section 3.2).

3.2. The Digital in the Humanities Scholar’s Research Process — a Systematic Approach

32 The specific nature of each of the three types of digital material has an impact on how each of them can be used and approached by the scholar in the research process. But before going into more detail about how this unfolds in the digital world, let us have a look at a schematic representation of how the research process usually unfolded before the advent of the digital (Figure 1). This very general model does not necessarily fit all parts of the humanities, but it provides a good idea of where the digital can later be located within the research process of most disciplines. In addition, it has to be stressed that what are presented as two distinct worlds — an analogue and a digital — are in practice in today’s digital world very often intermixed in the sense that even scholars claiming to be Digital Humanists switch between analogue and digital in various phases of their research. But for the sake of argument, they are presented as two distinct scholarly environments.

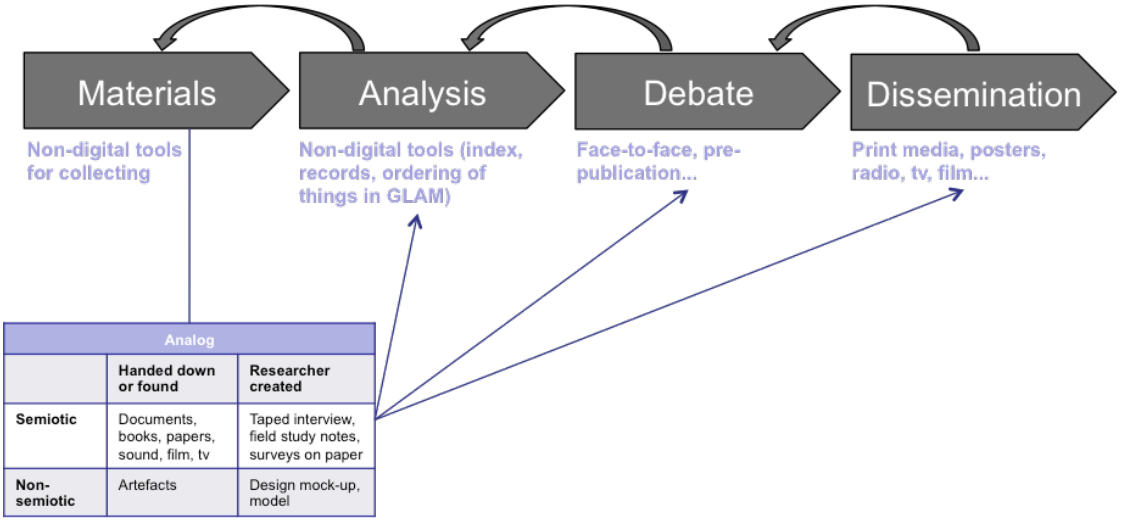


Figure 1. A schematic representation of the research process, before the advent of the digital.

33 The research process is boiled down to four main phases: the material of study is collected, it is analyzed and the results may be debated and later disseminated. This may be an iterative process, for instance, the analysis may mean that new material has to be procured, the debate may lead to corrections of the analysis, and the dissemination may affect all the previous steps (the iterative nature is indicated by the arrows looping back).

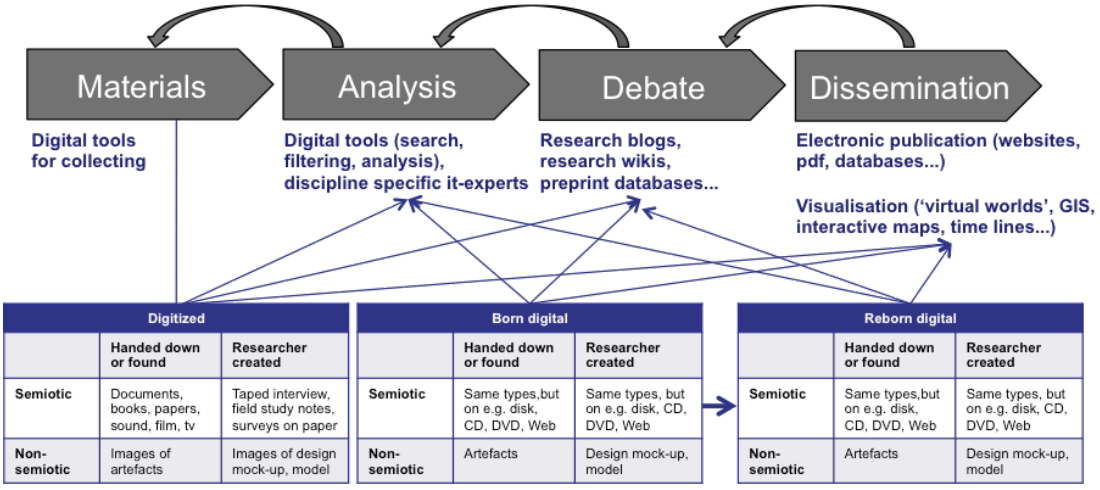
34 The key argument in this approach is that the nature of the material to be studied to a large degree determines — or at least sets up a range of possibilities for — each of the four steps. Analogue material has to be collected in an "analogue" way, which may imply that the scholar has to visit the collections or objects by moving physically to where they are; the analysis must rely on ancillary tools such as printed indexes, records on index cards, and on how things are ordered in some of the physical buildings and institutions holding the collections (e.g. in "GLAM" institutions, which stands for galleries, libraries, archives and museums). Discussions generally have to take place either face-to-face, for instance at conferences, or in pre-publications such as conference papers, with the process of dissemination taking place in print media such as books, journals or newspapers, or possibly in electronic media (radio, television).

35 But not only does the analogue nature of the objects of study affect each of the four phases of study. Their nature also has an impact on a more detailed level. Therefore, a systematic distinction is introduced between, on the one hand, semiotic and non-semiotic objects of study, and on the other hand material that is either handed down to, found or created by the scholar. Semiotic documents that are handed down from the past could, for instance, be objects carrying semiotic systems, based on discrete units (e.g. documents, books, newspapers with writing or still images, or film/television with writing, sound and moving images), whereas non-semiotic objects that are handed down are any kind of artefact. As to the objects of study that have been created by the scholars themselves, this type of material can also be either semiotic (for instance a taped interview, field study notes, surveys on paper, or the like), or non-semiotic, such as design mock-ups, or reconstructions of artefacts from the past (e.g. experimental archaeology). It has to be stressed that the two sets of distinctions are analytical distinctions that are made with a view of identifying the most predominant feature of the different types of objects of study. Thus, the clear distinction between semiotic and non-semiotic plays down the fact that semiotic materials are also artefacts (books are also artefacts), and that artefacts also mean something in themselves, although they do not "carry" a semiotic system, such as written letter for instance (an ancient weapon can be a sign of cultural power); and with regard to the latter distinction, material handed down from the past is never approached unbiased but is always selected by the scholar with a view to creating his or her own object of study, whereas the scholar's own creation of the material of study here and now often relies on material that was handed down.

36 As mentioned above, the aim is to underline that the characteristics of each of the different types of objects of study has an impact on how they can be approached in the four phases of the research process (as indicated by the thin arrows). The analytical approach to the object of study varies depending on whether we are studying handed-down documents

or artefacts, or interviews or archaeological experiments that have been created by the researcher, just as the debate and dissemination of the results may vary for the same reasons.

Now let us have a look at how this unfolds in a digital world, where the scholar is confronted with the three main types of digital material outlined in section 3.1. 37



Digitized		
	Handed down or found	Researcher created
Semiotic	Documents, books, papers, sound, film, tv	Taped interview, field study notes, surveys on paper
Non-semiotic	Images of artefacts	Images of design mock-up, model

Born digital		
	Handed down or found	Researcher created
Semiotic	Same types,but on e.g. disk, CD, DVD, Web	Same types, but on e.g. disk, CD, DVD, Web
Non-semiotic	Artefacts	Design mock-up, model

Reborn digital		
	Handed down or found	Researcher created
Semiotic	Same types,but on e.g. disk, CD, DVD, Web	Same types, but on e.g. disk, CD, DVD, Web
Non-semiotic	Artefacts	Design mock-up, model

Figure 2. A schematic representation of the research process in the digital age.

The fact that the nature of the material to be studied sets up a range of possibilities for each of the four steps in the research process also applies to digital material. In contrast to analogue material, digital material can be collected at a distance if it is online, digital tools for searching, filtering, and analysing the objects of study can support the analysis, and digital material opens an array of new ways of debating (e.g. blogs, wikis, pre-print databases) and of disseminating and interacting (e.g. computer files, databases, websites, visualisations such as virtual worlds, GIS, interactive maps and time lines). 38

However, although these opportunities for interacting with digital material are by and large a function of the digitality of the material as such, on a more detailed level they are also a function of each of the three main digital types mentioned above. Digitized, born-digital, and reborn-digital material all have their own array of possible forms of scholarly use, although overlaps may occur: digitized material can be collected, analyzed, debated and disseminated in ways that are specific to this material's digitality, whereas born-digital material has to be approached in other ways, and reborn-digital material in other ways again. This point will be elaborated in greater detail in section 4, where the nature of digitized and reborn-digital material is compared. 39

The three main types of digital material affect the research process in various ways, but there are other questions worth raising here as well: is the material semiotic or non-semiotic, and has it been handed down/found or created? With regard to digitized material, one main issue is that analogue artefacts and objects created by researchers cannot become digital as such but have to be transformed into images of some kind. And with regard to born-digital material, the non-semiotic objects are probably not entirely digital, but only come with digital components. As to reborn-digital material, each of the four subtypes of material will probably differ because they are most likely to be transformed in different ways when they are moved from the born-digital area to an archive. 40

3.3. Interplay between the Digital and the Humanities — A Brief Outline of a Historical Development

The systematic approach outlined above situates the digital within the scholarly process by introducing the lens of the three main types of digital material, based on their various provenances. But this analytical lens can also be used to give a brief outline of the historical shifts in the interplay between the humanities and the digital. As a working hypothesis, three waves of the nexus between the humanities and the digital can be identified (for other types of 41

periodisation, see for instance [Berry 2012, 3–6], [Thaller 2012, 14–17], and [Finnemann 2014, 95–100].

42 The first wave starts in the 1960s, when new research fields such as humanities computing, literary computing, computational linguistics, and later digital history emerge [Screibman], [Cohen and Rosenzweig 2005], [Finnemann 2014], and [McCarty 2005]. These new disciplines are initially based on the use of stand-alone mainframe computers; later, proprietary computer networks are introduced, and later again, the open internet. In terms of digital material, these traditions are in the main focused on digitized material.

43 The second wave starts in the mid-1990s, when two partly overlapping new research fields emerge, namely “New Media Studies” and “Internet Studies.” New Media Studies as well as Internet Studies are often not considered part of the Digital Humanities, but probably they should be, if the focus is on how the humanities — or at least a number of humanistic disciplines — have integrated the digital as an object of study. New Media Studies and Internet Studies are very broad fields of study, including a range of disciplines, but they all converge in the study of new media and of the internet (cf. [Brügger 2005, 106–107] for a brief account on the Digital Humanities and internet studies). In contrast to the first wave, the second wave has had a digital object of study from the very outset, and therefore it is focused on born-digital material.

44 The third wave is a further development and a subset of the second wave, since it adds a historical dimension to the study of born-digital material: historical studies of the internet within internet studies emerge in the early 2000s. One of the main objects of study is the archived web, but other types of preserved born-digital objects of study are also included; in this third wave the focus is on reborn-digital material (for an overview of web historiographies, see [Brügger 2010, 8–13]).

45 It is important to stress that the three waves have not replaced one another, but have continued to co-exist, which is why today we have the three main types of digital material identified above: digitized, born-digital, and reborn-digital.

4. Reborn-Digital Text — the Case of the Archived Web

46 As mentioned above, one of the main types of reborn-digital material is the archived web. In the following section, the case of the archived web will be introduced because it illustrates the importance of identifying the differences between digital material, despite its being digital, and consequently it may illustrate the importance of approaching digital materials in different ways. The archived web is compared to digitized material, since both types are transformations of already existing material, analogue and born-digital, respectively, but the material that is being transformed as well as the processes of transformation are very different, and so are the results. Thus, contrasting the two will highlight these differences (for a more detailed discussion of this topic, please see [Brügger 2005], [Brügger 2009], [Brügger 2011a], [Brügger 2011b], [Brügger 2012a], [Brügger 2013]; see also [Dougherty 2010], [Thomas 2010], [Foot and Scheider 2006], and [Masanès 2006]). However, it has to be stressed that since the process of digitization can range from keyboarding written documents to punched cards, text encoding, scanning documents to image files, or the digital recording of sound and moving images, digital collections are more complex and varied than the general presentation below indicates. Thus, a more detailed analysis of digitized collections in their own right should take into account the many differences characterizing these collections, including a focus on the provenance of as well as access to the collections.

4.1. Why Transform the Web from Born to Reborn Material?

47 Before going more into detail about the characteristics of the archived web, it is relevant to reflect on why the web is archived at all since it is a born-digital medium that is “out there” all the time. Even though we may have the impression that things can always be found on the online web, the web content changes at an unprecedented pace [Brügger 2005, 15], [Dougherty 2010, 8], [Brügger 2012a, 318]. The acknowledgement of this ever-changing nature of the online web, combined with an awareness of its growing importance for our societies and its importance as a historical source in the future, probably constitute the major impetus for cultural heritage institutions and individual scholars to transform the web into the archived web.

4.2. Macro and Micro Web Archiving

Web archiving refers to any form of deliberate and purposive preserving of online web material [Brügger 2011a, 25]. Two ways of web archiving can be distinguished: macro and micro web archiving [Brügger 2005, 10–11]. 48

Macro web archiving means web archiving carried out by professional archiving institutions, such as national libraries, with the aim of preserving the cultural heritage of, for instance, a nation state, and allowing as many different kinds of research projects as possible in the future. 49

Micro web archiving means web archiving carried out by people who are not professional web archivers, such as individual scholars or groups of scholars in relation to, for instance, a specific research project with the aim of preserving material of relevance for the research project in question. 50

What follows primarily applies to macro archiving, but much of it also applies to web archiving in general and to relatively small researcher-generated web collections. 51

4.3. Stable and Ephemeral Original

Digitization is based on the existence of an analogue original that in most cases is stable, whether it is a document on parchment, a newspaper, or a radio or television broadcast on tape. 52

In contrast, the web original to be archived is much more ephemeral in the strict sense of the word. As mentioned above, it is very likely to have changed or disappeared within a very short time interval, and there is thus no original to go back to (cf. [Schneider and Foot 2004, 115], [Masanès 2006, 1]). 53

4.4. What and How to Archive

Before starting to digitize an analogue collection, the main concern is *what* to digitize; whereas the *how* to digitize is in the main limited to questions about which software and hardware to use and how to arrange it. 54

The web can be archived in a variety of ways that will each result in genuinely different versions. For instance, it has to be decided where the archiving should start and stop, if specific file types should be included/excluded, if the archiving should be allowed to retrieve material on other servers, etc. Thus, the question of what to archive also applies to a web archive, but what is even more important is *how* to archive the web. The consequences of this are (1) that strictly speaking the archived web does not exist as such before it is archived, but is only created in the process of archiving; and (2) that if two archiving institutions decide to archive the same website at the same point in time, the results are very likely to be different, due to the different archiving settings. Thus, in contrast to digitized collections, where differences as to *how* to perform the digitization of the same analogue object will probably only lead to minor differences in the two collections, the *how* to perform the archiving in web archives often creates two unique, but different versions (it is also necessary to add that differences in assembling the archived bits and pieces in the archive may even complicate these differences). 55

4.5. Transparency and Opacity

The process of digitization is to a large extent transparent: the archiving institution has an overview of the collection to be digitized and of what is happening during the digitization process. 56

This is not the case with the process of web archiving. For a variety of reasons web archiving is a much more opaque process, mainly because one never knows exactly what happens “out there,” once the archiving software has been launched. Technical problems may occur, for instance the archiving process may simply stop, it may get off track, for instance, when encountering crawler traps such as calendars, or bot traps such as pages generating new links (both cases create an infinite circuit of requests to the webserver), or it may encounter file or software formats that cannot be archived (streamed video, java scripts, etc.). And a phenomenon that is specific for web archiving may add to the opacity, namely what can be termed “the dynamics of updating” (cf. [Brügger 2005, 22–23]), that is the fact that the 57

website being archived may change during the very process of archiving, and we do not know if, when, and where this may happen — as if the pages of a newspaper were continuously being edited while the newspaper was being digitized. And we cannot rely on going back and checking the original, since it may be either gone or changed.

4.6. Point(s) in Time and Continua

In general, the analogue media being digitized are usually related to either one point in time (for instance the publication date of printed media) or periods of time with clear-cut starting and stopping times (as with radio and television). 58

The web has a totally different “publication cycle.” Entire websites are published neither at a specific point in time nor within a clearly delimited time span, but are a continuous publication with no clear-cut starting and stopping time(s). In addition, the publication time is often not mentioned on the web by the “publisher.” The consequence of this is that the temporal subdivisions of the archived web material (in days, for instance) is added by the archiving institution, and is thus not an inherent part of the archived material — and therefore it is random and editable. 59

4.7. From Copies to Versions

The different ways of archiving the web combined with the opacity of the process means that the result of the web archiving process cannot be considered a copy on a 1:1 scale, as can a digitized object that is close to an identical copy of an original. Instead, the archived web must be regarded as a unique version — a version of a (probably) lost original, and very likely one version among other versions, none of which can be identified as the original. 60

4.8. From One Copy of Each to Too Few and Too Many Versions

In a digitized collection there is usually only one copy of each document, be it a handwritten manuscript, a newspaper, or a radio broadcast, basically because there is only one original and there is no good reason to create more identical copies. 61

In contrast, in a web archive there are very often both too few and too many versions of “the same” thing. On the one hand, it is very unlikely that any website has been archived all the time, for instance every day, which means that the scholar will be missing some of the material that was initially online (in addition, elements on a website (images, video, adds, etc.) may be missing too). On the other hand, some websites (or parts of websites) may have been archived several times at very short intervals, so there may be several versions from (almost) the same point in time. Thus, it is very likely that the scholar who wants to study material in a web archive does not only have one copy of each document. Instead, the processes and strategies of archiving the web and the preservation of the web in the web archive entail that in some cases a lot of what was once online is missing; whereas in other cases there may be too much material that is partly identical, but not exactly a duplicate. Finding one’s way through this network of (possibly) overlapping versions is especially difficult because the various extra versions of “the same” may not be versions of the same. There may be partial overlaps between, for instance, a webpage on a website archived in the morning and in the afternoon on the same day — as if a newspaper existed in several versions from the same day, with some pages (or elements on pages, e.g. images) being different in each version, or some being with fewer/more pages. 62

4.9. The Absence of a Register

Digitized collections are in the main based on an original analogue collection of which the archiving institution has a systematic overview, for instance there may be a register or catalogue. 63

This may also to a certain degree be the case in a web archive, namely if the web archive creates metadata for each archived website (as does, for instance, the Australian web archive Pandora, pandora.nla.gov.au). But this is not always the case, especially in web archives that use the so-called bulk archiving strategy, where vast amounts of websites are archived, either based on a list of all the domain names to archive (e.g. the Danish Netarkivet) or on following links from what has already been archived (e.g. the US-based Internet Archive). In these cases the mere number of domain names makes it impossible to manually register the websites being archived. For instance, the UK country code top level domain .uk consists of approximately 10 million domain names. In these cases the only types of register are either 64

the list of archived domains or the technical crawl log, where information about the archiving can be found, although the first of these two does not tell much, and the latter has to be transformed into information that makes sense to a scholar, and manually added information is still not an option.

For the reasons mentioned in the sections above, a web archive may not exactly know what is in the archive, and this state of affairs is even aggravated by the absence of a systematic register (and even the making of a register of a small number of websites is a challenge because of the nature of the archived web [Brügger 2011b]). 65

4.10. From Hyperlinks as Add-ons to Hyperlinks as Inherent

The advent of the hyperlink in the mid-1980s opened up new ways of searching and connecting documents in digitized collections. For instance, hyperlinks can be added to digitized newspapers or radio broadcasts, making it possible to get from one entity to another. 66

In contrast to digitized collections, where the hyperlink is an optional add-on, the hyperlink is an inherent part of the online web: in the main “no hyperlink” is the same as saying “no web.” And for the very same reason it is not optional to include the hyperlink in a web archive. At first sight one may not consider the necessity of including hyperlinks in a web archive a problem, but on a closer inspection it becomes evident that the inclusion of the online web’s hyperlinks in the web archive may challenge the scholarly use of the archive in ways that are unseen in digitized collections, basically because in digitized collections there is only one copy of each entity, and the publication of the material has already taken place in the past, at a specific and identifiable point in time. 67

The necessity of including hyperlinks in a web archive is a challenge, partly because archiving takes time (e.g. it takes two months to archive the entire Danish web on .dk), and partly because all web entities are not necessarily archived in the same depth below the front page (due to deliberate choices in the archiving settings, or to unexpected problems). Therefore, the presence of hyperlinks creates two sorts of inconsistencies in the web archive as a whole. First, a temporal inconsistency, because there may not be temporal coherence between the link source and the link target, that is the webpage where the link is found and the webpage to which the link points — they may have been archived with an interval of several days, weeks or even months. Second, when comparing the hyperlinks from a subset of archived websites, for instance in a network analysis, this hyperlink network may be inconsistent as such if the websites in question were not all archived in the same depth. Some may appear to have many links whereas others do not, as if we were comparing digitized newspapers which only had a front page in some cases but which had several sections in others; or comparing radio programmes which only had the first 30 seconds in some cases but only 10 seconds every minute in others. All in all, the compulsion to include hyperlinks in a web archive tends to make the collection temporally and spatially inconsistent when focusing on the software-based relations between websites or other web entities. 68

5. The Scholarly Use of the Web Archive

Compared to a digitized collection that tends to be more homogenous, web archives are by and large heterogeneous, messy and opaque at the outset. And this is true of web archives in ways that are much more fundamental than other types of collections, partly because the online web that was initially archived is in many ways also heterogeneous, messy and opaque; and partly because the archiving process itself as well as the preservation and access to the web in the web archive adds yet another layer of heterogeneity and messiness to the archived web. In short, the fundamental heterogeneity cannot be removed or resolved by technical means because it is a constitutive part of the web and of the web being archived. And to add insult to injury, the heterogeneity of the web archive also has a history of its own because the web and the web archive continue to develop, thus accumulating previous heterogeneities. 69

In continuation of the argument put forward above — the digital is not simply digital just because it is digital — let us now examine how the differences between a digitized collection and an example of a reborn-digital collection — a web archive — determine the range of possible ways in which they can be used by scholars. 70

5.1. Creating the Digital Object of Study

The first step when setting out to create the digital object of study in either a digitized collection or a web archive is to search the collection. This is one of the tasks where the advantages of the binary digital text (cf. section 2.1) become evident. 71

In digitized collections searches can usually be performed in a variety of ways, depending on the preferences and technical possibilities of the collection. But in many cases searches by name and date as well as some kind of free text searches are possible, often combined with various sorts of filtering of the results. 72

Within a web archive, searching for the name of a website may not be possible, basically because websites usually do not come with names [Brügger 2011b], but also because very often the only way to search the archive is to search for a specific URL, which is the web domain name which can be found in the location bar on the online web. However, some web archives have recently started to support free text searches, which are a great help for the scholar. But this leap forward also comes with a number of challenges which highlight the difference compared with digitized collections. 73

The well-known and relatively simple act of presenting search results may be slightly more challenging, partly because it is unclear what entity to show, partly because there may be many overlapping versions of “the same” thing in terms of time and space. Imagine a search for the term “humanities” in a collection of digitized newspapers: the result page will show all the pages where the word appears, and the results may be filtered further by the name of the newspaper, the date range or other parameters. But no matter what, there is only one copy of each of the pages on which the word is found; in other words, there is no temporal depth of each page. If we search for the term “humanities” in a ten-year-old web archive, the result page may show the individual webpages on which the term was found, and we may also be able to filter by the URL of the websites to which each page belongs, by date range, etc. But the results page also has to handle the existence of several — often overlapping — versions of “the same” webpage, since in the web archive “one found page” does not necessarily equal “one copy from each point in time,” but rather, as outlined above, something like “one webpage” equals “several versions.” In other words, each webpage may come with its own temporal and spatial depth in the form of partly overlapping versions, overlapping in both time and space (size). In addition, if the web archive does not have any metadata to display, it can be very difficult for the scholar to evaluate different versions. 74

Thus, the messiness of the web archive is replicated in the search results in such a way that the simple fact of displaying the results of a free-text search may prove to be much more complicated in a web archive than in a digitized collection. And there is no easy technical solution to this, because the core of the problem is the nature of the archived web itself. 75

5.2. Identifying, Evaluating and Selecting Versions

The next step after having searched the collection is usually to select and isolate the material that has been found for later study. This is also the case in a web archive, but before doing this one has to perform a step that is usually not necessary in digitized collections: deciding which of the many possible versions of “the same” should be included in the object of study. 76

As mentioned above, there are a number of reasons why a web archive may contain a large number of versions of “the same” web material; but it is important to stress that these versions are usually not identical duplicates, but genuinely different versions, and the differences may range from being very important to being insignificant. Therefore, the scholar using a web archive has to perform the tasks of identifying, evaluating and selecting the possible versions to include in the study. An example: we may want to study the BBC website (bbc.com), for instance the front page and the two sub-sections News (bbc.com/news) and Sport (bbc.com/sport). The front page may have been archived eight times per day, the front page of News and Sport three times every day, and the sub-pages at varying intervals (once every day, every second day, etc.). We then have many versions of some of the pages, fewer versions of others, and no versions of others; so the task is to decide which of these versions to choose to combine to form our object of study, even though one should be aware that because of the temporal and spatial inconsistencies of the archived web elements, one cannot reconstruct the website’s front page and the two sub-sections exactly as they looked at a specific point in time, but only as they may have looked at several points in time. Identifying, evaluating and selecting versions can be quite difficult, and it is therefore important that tools are developed to help the scholar to perform this task in such a way that 77

his or her choices are as informed as possible. This will include tools that can provide as much information about each of the versions as possible (e.g. starting/stopping time of the archiving), along with visualization tools to make it possible to display the different versions in a manageable manner, and tools to help select — or maybe even combine — which version should be used as the object of study.

5.3. Isolating the Versions Found

Once the scholar has chosen which versions to study, it is time to return to the step mentioned above, which involves isolating the material for later study. In a digitized collection this is in general a trivial task, mainly because each document is clearly identifiable as one (or more) file(s), and in general it is not linked to other documents (and if it is, this is usually done in a very unequivocal manner). For instance, digitized newspapers are usually kept as pdf files, whereas radio programmes are often preserved as avi files. Therefore, the relevant files can either be taken out of the collection and analyzed outside the archive, or bookmarked, put into a folder etc for further study. 78

Again, things work differently in a web archive, mainly because of the existence of hyperlinks as an inherent part of most web documents, since the hyperlink makes it slightly more complicated to remove the files from the archive. An example: in the study of bbc.com we have identified the versions to be studied, but if we then remove them from the archive, for instance by exporting them to our own computer, we also remove them from the communicative environment in which they were initially embedded in the archive, and suddenly things may appear either not to work anymore or to behave strangely. Links may no longer work because the link target is not part of our corpus, and this is true of the links that are immediately visible on the webpage and the more invisible links that make the webpage look as it does, for instance links to adds, images, or embedded videos located on other servers. Or links may work, but in strange ways if the material is removed from a web archive with no access to the live web and placed on a computer with online access. For instance, if a script on the front page of the archived version of bbc.com requests today's weather from the server of the national meteorological service, this server-side information may still be displayed if the server is still online, thus displaying today's weather and patching the website of the past with today's web (cf. [Brügger 2008, 159]). 79

Thus, the best way of isolating archived web material for study is to do this within the archive, for instance by generating an index of what has been chosen, and then limiting the study to this index. However, for technical reasons this may not be straightforward. For instance, the so-called deduplication procedure has to be taken into consideration; that is the fact that the web archive continually checks if a file is already in the archive (e.g. a pdf file), and if it is, it is not archived again with a view to saving disk space. But if a scholar later delimits a corpus to a certain date range, and the file was archived before the starting date of this date range, the file will not be displayed, even if it is in the archive. There are technical solutions to this, but the phenomenon illustrates that a web archive is a rather complex collection because of the existence of hyperlinks. 80

5.4 Analyzing, Debating, Disseminating

Once the object of study has been delimited and isolated, the scholar can start analyzing it. It is not possible to go into much detail about the great variety of analytical tools that can be applied in analyses of archived web material, but the main challenge for all these tools is that their possible use is a function of how the web archive has been constructed as well as how the concrete construction of the object of study within the web archive has been performed, and with what result. Thus, in a very fundamental way the nature of a reborn-digital collection such as a web archive constitutes the condition of possibility of the analysis. 81

5.5 Solutions? — The Need for a Web Philology

When faced with the web archive's challenges with regard to its scholarly use, are there any possible solutions for the scholar to address these challenges? 82

Since the heterogeneous, messy and opaque nature of the web archive is an inherent part of the archiving and of the archive itself, there are no easy solutions to overcome the challenges. However, this should not leave the humanities 83

scholar paralysed, because one of the most important tools in the traditional humanities toolbox is the skill of dealing with sources and their provenance. These skills just have to be reinterpreted and translated to fit a new digital environment. For instance, the task of comparing versions outlined above is in many ways similar to traditional philology, which means that a future "web philology" does not have to start from scratch but can be developed to fit the specific nature of the archived web by identifying and focusing on the differences between texts written on parchment or paper and texts archived on the web (for an outline of a web philology, see [Brügger 2008, 161–171].

84 The first steps towards a web philology would be to acknowledge the specific digitality of archived web, and to recognize that access to the digital code may even be a great help to the scholar, since this layer of invisible text beneath the visible text holds a lot of important information and metadata, among others about the provenance. Thus, what is needed is a new web philological toolbox that can help the scholar gain as much information as possible about the object of study. Such an approach will not solve the challenges, but it will help scholars to make as informed choices as possible.

6. Conclusions

85 Taking as a point of departure that one of the major transformative factors of the humanities at the beginning of the 21st century is the shift from analogue to a great variety of digital source material, it has been argued that this shift will affect the humanities in a variety of ways. To conclude, I shall briefly return to the question asked at the outset: to what degree will the humanities become digital?

86 On the one hand, the humanities are tending to become increasingly "digital." The reason for this is the above-mentioned shift from analogue to digital material, and the possibility of using digitally supported methods that this material allows and increasingly renders necessary. Thus, this spread of the digital will probably push the humanities closer to becoming the Digital Humanities.

87 On the other hand the Digital Humanities must by and large remain the humanities. The fundamental questions, theories, methods and aims of the humanities largely remain unchanged, despite the widespread digitality and the challenges that this raises for the humanities. The Digital Humanities do not in themselves constitute an entirely new paradigm. Instead, they open up an array of possibilities either for doing what was previously done in new ways, or for rethinking well-known practices of the humanities, for instance by integrating software-supported methods and by using digital research infrastructures.

88 What is needed is a well-balanced Digital Humanities allowing for the fact that the foundation of the scholarly activities of the humanities is fundamentally changed, but also acknowledging that this fact does not necessarily change the humanities fundamentally. All the humanities need not become Digital Humanities, but most humanities will not remain unchanged since they are challenged by the digitality of the object of study in itself or of the digitally supported methods and tools to be used within the humanities — or both. Therefore, in the 21st century the difference between the humanities and the Digital Humanities is quantitative rather than qualitative: all parts of the humanities will become digital to some extent, but not all will become digital to the same extent. Thus, the main question is not one of being digital or not, but rather one of being more or less digital.

89 And no matter where one wants to strike the balance between more or less digitality, the introduction of the digital in the humanities is a game changer that has already sparked relevant and important discussions about what the (Digital) Humanities could be.

Works Cited

American Council of Learned Societies 2006 American Council of Learned Societies. *Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. New York: American Council of Learned Societies, 2006.

Arts and Humanities in Higher Education 2011 *Arts and Humanities in Higher Education* 11, no. 1-2 (2011), special double issue: Digital humanities, digital futures.

Barnett 2014 Ronald Barnett. “Imagining the Humanities – Amid the inhuman.” *Arts & Humanities in Higher Education* 13, no. 1-2 (2014): 42-53.

Bassnett 2002 Susan Bassnett. “Is There Hope for the Humanities in the 21st Century?” *Arts & Humanities in Higher Education* 1, no. 1 (2002): 101-110.

Belfiore 2014 Eleonora Belfiore. “‘Impact,’ ‘value’ and ‘bad economics’: Making sense of the problem of value in the arts and humanities.” *Arts & Humanities in Higher Education* OnlineFirst, April (2014).

Beňuš 2010 Stefan Beňuš. “Building interfaces between the humanities and cognitive sciences: The case of human speech.” *Arts & Humanities in Higher Education* 9, no. 3 (2010): 353-374.

Berry 2012 David M. Berry. “Introduction: Understanding the Digital Humanities.” In *Understanding Digital Humanities*. Edited by David M. Berry. New York: Palgrave Macmillan, 2012, pp. 1–20.

Bolton 2008 Gillie Bolton. “Boundaries of Humanities: Writing Medical Humanities.” *Arts & Humanities in Higher Education* 7, no. 2 (2008): 131-148.

Borgman 2009 Christine L. Borgman. “The Digital Future is Now: A Call to Action for the Humanities.” *Digital Humanities Quarterly*, 3, no. 4 (2009). http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Brügger 2005 Niels Brügger. *Archiving Websites: General Considerations and Strategies*. Aarhus: The Centre for Internet Studies, 2005.

Brügger 2008 Niels Brügger. “The Archived Website and Website Philology: A New Type of Historical Document?” *Nordicom Review* 29, no. 2 (2008): 155-175.

Brügger 2009 51. Niels Brügger. “Website history and the website as an object of study.” *New Media & Society* 11, no. 1-2 (2009): 115-132.

Brügger 2010 Niels Brügger. “Introduction: Web history, an emerginrg field of study.” In *Web history*. Edited by Niels Brügger. New York: Peter Lang, 2010, pp. 1-25.

Brügger 2011a Niels Brügger. “Web archiving — Between past, present, and future.” In *The handbook of Internet studies*. Edited by Mia Consalvo, and Charles Ess. Oxford: Wiley-Blackwell, 2011, pp. 24-42.

Brügger 2011b Niels Brügger. “Digital history and a register of websites: An old practice with new implications.” In *The long history of new media: Technology, historiography, and contextualizing newness*. Edited by David W. Park, Nick W. Jankowski, and Steve Jones. New York: Peter Lang, 2011, pp. 283-298.

Brügger 2012a Niels Brügger. “Web History and the Web as a Historical Source.” *Zeithistorische Forschungen*, 9, no. 2 (2012): 316-325.

Brügger 2012b Niels Brügger. “When the Present Web is Later the Past: Web Historiography, Digital History, and Internet Studies.” *Historical Social Research*, 37, no. 4 (2012): 102-117.

Brügger 2013 Niels Brügger. “Historical network analysis of the Web.” *Social Science Computer Review*, 31, no. 3 (2013): 306-321.

Brügger and Finnemann 2013 Niels Brügger and Niels Ole Finnemann. “The Web and Digital Humanities: Theoretical and Methodological Concerns.” *Journal of Broadcasting and Electronic Media*, 57, no. 1 (2013): 66-80.

Burdick et al 2012 Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp. *Digital_Humanities*. Cambridge, MA: MIT Press, 2012. http://lab.softwarestudies.com/2012/11/digitalhumanities-book-is-out.html

Cohen and Rosenzweig 2005 Daniel J. Cohen, and Roy Rosenzweig. *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia, Penn.: University of Pennsylvania Press, 2005.

Collins 2011 Ellen Collins, Monica E. Bulger, and Eric T. Meyer. “Discipline matters: Technology use in the humanities.” *Arts & Humanities in Higher Education* 11, no. 1-2 (2011): 76-92.

DHQ e-Science 2009 *Digital Humanities Quarterly*, 3, no. 4 (2009), Special Cluster: e-Science for the Arts and Humanities. http://digitalhumanities.org:8081/dhq/vol/3/4/000079/000079.html

Deegan 2014 Marilyn Deegan. “‘This ever more amorphous thing called Digital Humanities’: Whither the Humanities Project?” *Arts & Humanities in Higher Education* 13, no. 1-2 (2014): 24-41.

Digital Humanities Manifesto *A Digital Humanities Manifesto*. 2008, http://manifesto.humanities.ucla.edu/2008/12/15/digital-humanities-manifesto/

Dougherty 2010 Meghan Dougherty, Eric T. Meyer, Christine Madsen, Charles van den Heuvel, Arthur Thomas, and Sally Wyatt. *Researcher Engagement with Web Archives: State of the Art.* London: JISC, 2010.

European Commission 2011 European Commission. *The New Renaissance: Report of the "Comité des Sages" on bringing Europe's cultural heritage online*. Luxembourg: Publications Office of the European Union, 2011.

Evens 2012 Aden Evens. "Web 2.0 and the Ontology of the Digital." *Digital Humanities Quarterly*, 6, no. 2 (2012). http://www.digitalhumanities.org/dhq/vol/6/2/000120/000120.html

Finnemann 1999 Niels Ole Finnemann. "Modernity Modernised: The Cultural Impact of Computerisation." In *Computer, Media and Communication*. Edited by Paul A. Mayer. Oxford: Oxford University Press, 1999, pp. 141–159.

Finnemann 2014 Niels Ole Finnemann. "Digital humanities and networked digital media." *MedieKultur. Journal of media and communication research* 57 (2014): 94-114.

Foot and Scheider 2006 Kirsten A. Foot, and Steven M. Schneider. *Web campaigning*. Cambridge, Mass.: MIT Press, 2006.

Gold 2012 Matthew K. Gold. *Debates in the Digital Humanities*, ed. Minneapolis/London: University of Minnesota Press, 2012.

Gronbeck 2005 Bruce E. Gronbeck. "Is Communication a Humanities Discipline? Struggles for academic identity." *Arts & Humanities in Higher Education* 4, no. 3 (2005): 229-246.

Harpham 2005 Geoffrey Galt Harpham. "Beneath and Beyond the 'Crisis in the Humanities'". *New Literary History* 36, no. 1 (2005): 21–36.

Historical Social Research 2012 *Historical Social Research/Historische Sozialforschung* 37, no. 3 (2012), special theme: Controversies around the Digital Humanities.

Hockey 2000 Susan Hockey. *Electronic texts in the Humanities*. Oxford: Oxford University Press, 2000.

Holm 2015 Poul Holm, Arne Jarrick, and Dominic Scott. *Humanities World Report 2015*. New York: Palgrave Macmillan, 2014.

Howarth 2004 David Howarth. "Is Law a Humanity (or is it more like Engineering)?" *Arts & Humanities in Higher Education* 3, no. 1 (2004): 9-28.

Jones 2014 Steve E. Jones. *The emergence of the digital humanities*. New York & London: Routledge, 2014.

Kirschenbaum 2013 Matthew Kirschenbaum. "The .txtual Condition: Digital Humanities, Born-Digital Archives, and the Future Literary." *Digital Humanities Quarterly*, 7, no. 1 (2013). http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html

Laue 2004 Andrea Laue. "How the computer works." In *A Companion to Digital Humanities*. Edited by Susan Schreibman, Ray Siemens, and John Unsworth. Oxford: Blackwell, 2004, pp. 145–160.

Liu 2011 Alan Liu. "The state of the digital humanities: A report and a critique." *Arts & Humanities in Higher Education* 11, no. 1-2 (2011): 8-41.

Masanès 2006 Julien Masanès. *Web Archiving*, ed. Berlin: Springer, 2006.

Mayer-Schönberger 2013 Viktor Mayer-Schönberger and Kenneth Cukier. *Big Data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt Publishing Company, New York, 2013.

McCarty 2005 Willard McCarty. *Humanities computing*. New York: Palgrave Macmillan, 2005.

MedieKultur 2014 *MedieKultur. Journal of media and communication research* 57 (2014), special issue: Digital Humanities: Now and Beyond

Nussbaum 2010 Martha C. Nussbaum. *Not For Profit: Why Democracy Needs the Humanities*. Princeton: Princeton University Press, 2010.

Olmos-Peñuela 2014 Julia Olmos-Peñuela, Paul Benneworth, and Elena Castro-Martínez. "Are sciences essential and humanities elective? Disentangling competing claims for humanities' research public value." *Arts & Humanities in Higher*

Education OnlineFirst, May (2014).

Parker 2007 Jan Parker. “Future Priorities of the Humanities in Europe: What Have the Humanities to Offer? Report of a round table conference held to draft a manifesto for the European Commissioner and working papers for the EC Working Party on Future Priorities for Humanities Research.” *Arts & Humanities in Higher Education* 6, no. 1 (2007): 123-127.

Pascoe 2002 Robert Pascoe. “An Australian Perspective on the Humanities.” *Arts & Humanities in Higher Education* 2, no. 1 (2002): 7-22.

Schneider and Foot 2004 Steven M. Schneider, and Kirsten A. Foot. “The Web as an Object of Study.” *New Media Society* 6, no. 1 (2004): 114–122.

Screibman Susan Schreibman, Ray Siemens, and John Unsworth. *A Companion to Digital Humanities*, eds. Oxford: Blackwell, 2004.

Svensson 2010 Patrik Svensson. “The Landscape of Digital Humanities.” *Digital Humanities Quarterly*, 4, no. 1 (2010). http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html

Svensson 2011 Patrik Svensson. “The digital humanities as a humanities project.” *Arts & Humanities in Higher Education* 11, no. 1-2 (2011): 42-60.

Svensson 2012a Patrik Svensson. “Envisioning the Digital Humanities.” *Digital Humanities Quarterly*, 6, no. 1 (2012). http://www.digitalhumanities.org/dhq/vol/6/1/000112/000112.html

Svensson 2012b Patrik Svensson. “Beyond the Big Tent.” In *Debates in the Digital Humanities*. Edited by Matthew K. Gold. Minneapolis/London: University of Minnesota Press, 2012, pp. 36–49.

Terras 2012 Melissa Terras. “Digitization and digital resources in the humanities.” In *Digital Humanities in practice*. Edited by Claire Warwick, Melissa Terras, and Julianne Nyhan. London: UCL/Facet Publishing, 2012, pp. 47–70.

Thaller 2012 Manfred Thaller. “Controversies around the Digital Humanities: An agenda.” *Historical Social Research* 37, no. 3 (2012): 7-23.

Thomas 2010 Arthur Thomas, Eric T. Meyer, Meghan Dougherty, Charles van den Heuvel, Christine Madsen, and Sally Wyatt. *Researcher Engagement with Web Archives: Challenges and Opportunities for Investment.* London: JISC, 2010.

Warwick 2012 Claire Warwick, Melissa Terras, and Julianne Nyhan. *Digital Humanities in practice*, eds. London: UCL/Facet Publishing, 2012.