SSOAR

Open Access Repository
www.ssoar.info

When the present web is later the past: web historiography, digital history and internet studies

Brügger, Niels

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:
Brügger, N. (2012). When the present web is later the past: web historiography, digital history and internet studies. *Historical Social Research*, *37*(4), 102-117. https://doi.org/10.12759/hsr.37.2012.4.102-117

Nutzungsbedingungen:
Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
https://creativecommons.org/licenses/by/4.0/deed.de

Terms of use:
This document is made available under a CC BY Licence (Attribution). For more Information see:
https://creativecommons.org/licenses/by/4.0

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-383785

When the Present Web is Later the Past: Web Historiography, Digital History, and Internet Studies

Niels Brügger [*]

Abstract: *»Wenn das Web Vergangenheit wird: Web-Geschichtsschreibung, Digitale Geschichte und Internet-Forschung«.* Taking as point of departure that since the mid-1990s the web has been an essential medium within society as well as in academia this article addresses some fundamental questions related to web historiography, that is the writing of the history of the web. After a brief identification of some limitations within digital history and internet studies vis-a-vis web historiography it is argued that the web is in itself an important historical source, and that special attention must be drawn to the web in web archives – termed reborn-digital material – since these sources will probably be the only web left for future historians. In line with this argument the remainder of the article discusses the following methodological issues: What characterizes the reborn-digital material in web archives, and how does this affect the historian's use of the material as well as the possible application of digital analytical tools on this kind of material?

Keywords: web, historiography, archiving, digital history, digital humanities, internet studies, e-research, analytical tools.

1. Introduction

Since the mid-1990s the world wide web has become the nexus of digital networked communication in the public domain by integrating former analog media as well as previously individual network applications (e.g. email, chat, newsgroups, listserv), and by developing its own software forms (blogs, wikis, video sharing). The world wide web – or simply: the web – has become the center of gravity of the digital networked communicative infrastructure, and in many ways also of our communicative infrastructure at large since many off-line activities are entangled in the web such as social, cultural, political, and commercial life.

Although new digital devices are constantly emerging, historians who in the future want to understand our time probably will have to understand the web.

[*] Niels Brügger, Centre for Internet Studies, Aarhus University, Helsingforsgade 14, Aarhus N, Denmark; nb@imv.au.dk.

Thus, they may be in the need of, on the one hand, a web historiography aiming at writing the history of the web, and, on the other, a 'web-minded' historiography, that is a historiography which pays attention to the role of the web in present day society.

Since knowledge about the web's history and about the use of the web in historical study is a condition for the web-minded historiography web historiography will be the focal point of the present article. Based on an identification of the web historiographical limitations within two related research traditions – digital history and internet studies – some of the fundamental methodological questions within web historiography are discussed, and it is argued that special attention has to be drawn to the web as a historical source, especially the web as it can be found in web archives.

1.1 Digital Academia

The web is not only an important object of study for historians, be that for a web historiography or for a web-minded historiography. For two decades the web has also been an integrated part of historiography since more and more documents are made available on the web, and the web is part of the historian's methodological toolbox.

In this respect, contemporary historiography is part of a wider movement within the humanities and the social sciences. For several years stand-alone and networked computers have been used in research, for instance within literary computing, humanities computing, and computational linguistics. The web is just the latest platform within these trends which in recent years have been known under umbrella terms such as 'cyberscience', 'digital humanities' and 'e-research', and which have been more and more closely related to the establishing of national and transnational digital research infrastructures (cf. ESFRI 2006; ESH 2011).[1]

1.2 Digital Sources and the Web

Digital humanities and digital research infrastructures are occupied with providing the material to be analyzed in a digital form as well as the digital tools to analyze the material and to convey the result. The combination of digital sources and digital tools allows for a number of studies which were not possible beforehand, for instance based on search queries in large amounts of digital material.

However, digital materials may have become digital in different ways each of which affects the nature of the material differently, and in many ways the

[1] About these traditions and recent debates cf. Berry 2012; Dutton and Jeffreys 2010; Gold 2012; Jankowski 2009; Schreibman et al. 2004; Svensson 2011; Thaller 2012.

analytical tools which can be used are a function of the form in which the digital sources are made available. In the following, a distinction between digitized, born-digital, and reborn-digital material is used (cf. Brügger 2012b).

Digitized material is previously analog material which has been digitized. This could be either semiotic sources (e.g. written or printed documents, images, audio, video) or artifacts (made available as photographs, film, or video).

Born-digital material is material that has never existed in any other form than digital. This could be material on diskette, CD-ROM, DVD, or in computer networks; in general the born-digital material is not created by the scholar, but in some cases it is, for instance in the form of online surveys or other kinds of web-based data-collecting.

Reborn-digital material is digitized or born-digital material which has been collected and preserved, and which to some degree has been changed in this process. This could be material in a web archive.

In this triadic distinction the web can play three different roles. First, as a platform for distributing digitized analog materials, second, as a born-digital source, and, third, as a reborn-digital source in a web archive.

2. Traditions Related to Web Historiography

Especially two research traditions with affinity to digital humanities and e-Research are of interest to web historiography: digital history and internet studies. Despite the fact that these two traditions have not been related to each other, it is fruitful to examine them together identifying how web historiography can break with as well as continue both of them.

2.1 Digital History

The spread of the web in the mid-1990s combined with increased digitization of analog collections of documents provided historians with new ways of finding, manipulating and analyzing the source material as well as of disseminating their studies. Out of these new opportunities emerged what in the late 1990s was to be known as 'digital history'.[2]

One of the most influential scholars within the field of digital history as related to the web has been Roy Rosenzweig. In his co-authored *Digital*

[2] About the term 'digital history', see Cohen et al. 2008, 453-4. Digital history is only the latest stage of a long tradition (from the 1940s) among historians for using computers and computer networks (Thomas 2004). One of the earliest works of historical scholarship on the web was Edward L. Ayer's 'Valley of the Shadow Project' about two communities in the American Civil War (Thomas 2004, 62-3).

History: A Guide to Gathering, Preserving, and Presenting the Past on the Web (Cohen and Rosenzweig 2006) digital history is understood as the use of digital media and digital networks to make historians "do our work as historians better" (ibid., 3). As the sub-title indicates, the web is mainly considered a platform for gathering, preserving and presenting the past. What is at center stage for digital history is the 'history web', that is the web used as a historiographical tool.

Only when discussing the collecting of history online – "gathering the history that was made online, or 'born digital'" (ibid., 161) – the web is touched upon as a historical source in its own right. In relation to collecting blogs and newspaper websites related to the attacks on the USA on September 11 2001 it is maintained that "a large percentage of this initial set of historical sources, unlike paper diaries or print versions, will likely be gone if we look for them in 10 years (…) Similarly, unlike the pages of their physical editions, newspaper websites change very rapidly" (ibid., 161). Unfortunately, these relevant ascertainments are accompanied neither by considerations as to how the process of preserving the blogs and newspaper websites in a web archive may affect the materials, nor by reflections about the characteristics of the archived web materials as historical documents.

However, these issues are touched upon very briefly in an earlier article about the possible use of digital sources and how such use challenges historiography. According to Rosenzweig the Internet Archive (an American non-profit internet archive, cf. Kimpton and Ubois 2006, 202-4) can be considered "an extraordinarily valuable resource" (Rosenzweig 2003, 751) but it is stressed that much of the material is characterized by various forms of incompleteness, basically because something is missing (images, pages). In addition, Rosenzweig mentions that the dynamic and hyperlinked nature of the web may pose a problem when preserving it (ibid., 742). But apart from these few sentences web archiving and web archives are not touched upon.

Thus, within the tradition of digital history of which Rosenzweig's texts are seen as representative, the web is mainly used for finding, searching and annotating digitized source material, for getting in contact with a large audience (fellow historians and the public), and for presenting the sources and the results of the historical studies in new and more interactive ways than in print, radio or television – for instance in time lines or by combining sources with geographical information (about 'historical GIS', see Thomas 2004, 66; ESF, 29-30). In this venture the main sources are digitized versions of analog materials, whereas the web is merely considered a database with and a dissemination platform for these documents.

Apparently, the enthusiasm among digital historians that the tiresome task of finding documents in archives and libraries is now made possible with a mouse click overshadows that the very medium for this activity – the web, online as well as archived – is in itself a valuable and valid source to contemporary

history, and that historiographical reflections are needed about the status, the archiving and the subsequent use of this source.

2.2 Internet Studies

Internet studies is a very broad and interdisciplinary field of study, and has to a large extent been dominated, first, by a number of common online themes – communities, games, news, politics, language, and privacy, among others – and, second, by the development of approaches originating from the social sciences (e.g. virtual ethnography, network analysis).[3]

As was the case with digital history, internet studies have mainly been based on digital sources, and the use of digital analytical tools has been widespread (e.g. hyperlink analysis tools, cf. De Maeyer 2012). However, in contrast to digital history, the primary digital source is born digital and not digitized, namely the internet in its many forms.

In addition, for more than a decade the many internet studies have had one thing in common: focus has mainly been on the internet as it looked at the time of the study, whereas the historical developments of the internet in the past has only attracted little attention. Thus, it comes as no surprise that the scarcity of historical internet studies is accompanied by a lack of theoretical and methodological reflections about doing internet and web history (an exception is Rosenzweig 2004).

This state of affairs is illustrated in some of the publications which set out either to discuss or to condensate the state of the art of the field. In 2005 a special issue of *The Information Society* was dedicated to internet studies (The Information Society 2005), but neither the history of the internet nor reflections on internet historiography was on the agenda. A few years later two handbooks on internet studies were published almost simultaneously (Hunsinger et al. 2010; Consalvo and Ess 2011), and in these two volumes only one chapter indirectly addressed web history (Brügger 2011a). There may be good and practical reasons for this: when establishing a new field of study one of the top priorities is often to study the present in order to legitimize the discipline, and with the object of study in mind the short life of the internet and especially the web may not have been considered a history.

However, historical studies of the internet have been made (e.g. Abbate 2000; Hauben and Hauben 1997; Henderson 2002; Naughton 2002; Poole 2005), whereas historical studies of the web are scarce, probably because the

[3] By 'internet studies' Consalvo and Ess understand the study of "the distinctive sorts of human communication and interaction facilitated by the Internet" (Ess and Consalvo 2011, 1), and in this sense "Internet studies may be traced to the early 1990s" (ibid.). See also Barry Wellman's brief account of the first years of internet studies (Wellman 2011). 'Internet studies' has largely been identified with the activities within the Association of Internet Researchers (Hunsinger et al. 2010, xxiii; Wellman 2011, 21).

internet has a history of five decades, whereas the web only dates back to 1991. But within the last couple of years the number of historical web studies has been growing (for an overview see Brügger 2010b), and edited volumes (Brügger 2010a; Brügger and Burns 2012) as well as journal articles continue to appear (e.g. Jacobson 2012; Weber 2012).

Apparently, the wish to undertake studies which keep up with a rapidly changing medium such as the web overshadows the past of the web on which the present web is based. In consequence, web historiography as a sub-field of study in its own right with specific methodological and theoretical challenges has not played any role within internet studies.

3. Web Historiography

The conception of web historiography argued for in this article emerges in continuation of – and in rupture with – digital history as well as internet studies. Web historiography is in continuation of digital history and internet studies in bringing into focus the use of digital sources and digital analytical tools.[4]

In addition, on the one hand, web historiography continues the historiographical approach of digital history which is almost absent within internet studies, and, on the other, it continues internet studies' interest for the web as an object of study and an important source, which is absent within digital history with its focus on 'history web' instead of 'web history'.

However, web historiography takes digital history and internet studies one step further in three ways. First, by bringing into focus the third type of digital sources mentioned above, namely reborn-digital material in the form of archived web, especially as it appears in broad web archives archiving the cultural heritage (cf. below). Second, by acknowledging that digital analytical tools used for analyzing digitized as well as born-digital online material cannot necessarily be applied as is to archived web material. Third, by insisting on undertaking web historiographical reflections as to methodology and theory. The argument for bringing archived web material in focus is that as the present becomes past the web continually disappears and the archived web will gradually become the only web source from the past; thus, in the future web history will to a large extent be written on the basis of archived web material.[5]

A first step in the methodological reflections is to raise the following questions: What characterizes the reborn-digital material in broad web

[4] Within web historiography, digital history, and internet studies the digital sources and digital analytical tools are in most cases supplemented with analog sources and tools (Brügger 2010c, 41-2).

[5] Weber 2012 is one of the very few web historical studies based on broad web archives.

archives, and how does this affect the historian's use of the material as well as the possible application of digital analytical tools on this kind of material?[6]

3.1 Web Archiving and Web Archives

With a view to examining what characterizes the reborn-digital material in broad web archives clarifications are needed as to what 'web archiving' and 'broad web archive' may signify.

Web archiving can be understood as "any form of deliberate and purposive preserving of web material" (Brügger 2011a, 25). The most widespread way of collecting the web is web harvesting, that is the use of web crawlers that contact web servers and download their files to the archive.[7]

One of the main characteristics of web archiving is that the process of archiving itself may change what is archived, thus creating something that is not necessarily identical to what was once online. The reasons for this are, first, that a number of choices have to be made as to what and how to archive (archiving strategy, archiving software, file types or parts of websites to in-/exclude, archiving depth below the front page, in-/exclusion of material on other web servers, and the like). And, second, that a website may be updated during the process of archiving, just as technical problems may occur whereby web elements which were initially online are not archived. Thus, it can be argued that the process of archiving creates the archived web on the basis of what was once online: the born-digital web material is reborn in the archive.

In the main web archiving can serve two different purposes. It can be performed by archiving institutions, such as national libraries with a view to preserving the cultural heritage of, for instance, a nation state, or it can be undertaken by scholars with a view to collecting and preserving an object of study for a specific research project.[8] In the first case the result is a very broad web archive with no specific future use in mind, whereas in the latter the result is a narrow web archive with an intended use. Both types of web archives are created on the basis of archiving strategies, and regarding the broad web archive, especially three strategies are commonly used: snapshot, selective, and event archiving. The snapshot strategy aims at archiving a large amount of web material, usually entire Top Level Domains such as .de, .uk or .dk which may take several months. The selective strategy archives a limited number of websites, selected individually in advance and usually archived frequently. And the event strategy sets out to archive the web activity in relation to a specified

---

[6] For further reflections about web historiography, see Brügger 2009, 2010a, 2010b, 2010c, 2012b.

[7] This section briefly relates Brügger 2011a, 25-9. For a brief history of web archiving see ibid., 29-32. About web archiving, see Brügger 2005, Brügger 2011a; Masanès 2006; Schneider, Foot and Wouters 2009.

[8] For an example of the latter see Foot and Schneider 2006, 28.

event (e.g. natural disasters, political elections, sport events, etc.) based on prior selection.

3.2 Characteristics of the Reborn Web

As a result of the combination of the many choices made by the archiving institution and the possible updating and technical problems during the process of archiving the archived web is not necessarily reborn as a copy on a 1:1 scale of what was initially on the live web at a given point in time. It is better characterized as a unique version of which the original is forever lost (cf. Brügger 2011a, 34). The process of web archiving as well as its result – versions and not copies – imply that the reborn-digital web material in broad web archives has the following three characteristics.[9]

First, the broad web archive is incomplete as well as too complete compared to what was once online. It is incomplete since something is probably missing. But it is too complete in the sense that more different versions may exist of, for instance, the same web page or website from (almost) the same – or from overlapping – point(s) in time (it may have been archived as part of a snapshot, a selective, and an event harvest).[10]

Second, the sheer size of a broad web archive makes it almost impossible to document in detail how each web element has entered the archive. General information about the archiving may exist (chosen strategies, used software, known deficiencies, planning of the archiving, used web addresses, etc.), just as automatically generated information about the execution of the archiving process can be available (e.g. logfiles), but most broad web archives do not provide access to this information. Thus, due to this lacking documentation it can be difficult to clarify the differences and similarities of the versions, and the result is a certain uncertainty about the status of the archived material.

Third, archived web material in broad web archives tends to be inconsistent in terms of time and space, compared to the online web. This is due to the fact, first, that all elements were not archived simultaneously and with the same intervals, and, second, that everything has not necessarily been archived in its totality, for instance that all websites are not archived at the same depth. In addition, these inconsistencies are aggravated by the too complete nature of the archive, that is the possible existence of more versions of the same, from each point in time as well as with different spatial extension. All in all the broad web archive constitutes a patchwork of overlapping, but not identical times and spaces, and it is therefore less consistent than the online web from which it is created.

[9] These characteristics and others are elaborated in Brügger 2012a.

[10] Being incomplete is an inherent part of almost any archive, but as shown in this article web archives are incomplete in different ways than traditional archives.

4. Methodological Challenges to Web Historiography

Each of these three characteristics of the broad web archive challenges the web historian's research process. In the following some of these challenges shall be discussed, first by identifying a few general challenges that are there, no matter what the historical study focus on, second by outlining challenges related to the part of the web on which the analysis focuses.[11]

4.1 General Challenges

The first general challenge is to find the relevant material in the web archive. The trivial task of searching today's online web is somewhat more complicated in a web archive. Most web archives do not support free text search, but only access to specific web sites and web pages by writing the correct web address.[12] Moreover, even if free text search was an option, it would be hard to re-make a search like it would have been performed in the past, and thereby to find the archive's versions of relevant material about, for instance, the terrorist attacks on the USA in September 2001 present on the web at that time. The reason for this is the incomplete and too complete nature of the web archive: the whole web to be searched may not be in the archive (most likely it is not), and any web page may exist in several versions from the same point in time. In both cases the result is a biased search result. Thus, the web archive challenges one of the most natural means of navigating the online web today.

If the relevant material has been found in the web archive – despite these limitations – the next general challenge is to delimit what has to be used for further study, in other words: to create a corpus. Delimiting the sources to be studied is only the first step, since in a broad web archive it has to be decided which of the different versions of the same that have to be included in or excluded from the corpus, in case more versions exist. Thus, creating a corpus in a web archive is very often a two-step enterprise, and the second step is complicated by the possible existence of overlapping, but not identical versions, and by the lack of documentation as to their provenance.

After having found and delimited the relevant web material a third general challenge may arise, if the web historian wants to make a list of the empirical material, for instance in the form of a register. In relation to, for instance, websites the major challenges are, first, that it can be difficult to determine when a website started since this is not usually communicated on the website, and the first time it appears in the archive does not tell much about when it was published for the first time; second, registering what happens after a website is

[11] The general challenges are discussed in detail in Brügger 2011b, 2012a, 2012b.
[12] However, the Australian Pandora offers subject entries, cf. <http://pandora.nla.gov.au>.

published is a challenge because temporal sub-divisions such as exact time of publication or updating are usually not communicated on a website, thus making the time after the first publication an undivided continuum; third, determining a fixed end of publication, based on material in a web archive, is a challenge since often websites remain on the web server even if no longer updated and they may thus have been archived years after they actually stopped; or the archive may have stopped archiving a website even though it has continued to be on the web. The apparently simple task of making a register of the material to study is not trivial.

4.2 Specific Challenges

In addition to these general challenges, the historian who sets out to use sources from a broad web archive is facing a number of specific challenges, all of which revolve around how the web is understood as an object of study. With a view to delimiting the web as an analytical object, a distinction between five analytical levels may be useful: a) the web element (e.g. an image or a video); b) the web page (what is seen in a browser window, e.g. the front page); c) the website (interrelated web pages); d) the web sphere (web activity in relation to a theme or an event, e.g. political election); and e) the web as a whole (phenomena transcending the web, e.g. the web's content in its totality) (cf. Brügger 2009, 122-5). It should be stressed that these analytical levels are knit together since they constitute each other's context: the web page is the background for the web elements, the website is composed of web pages with web elements, etc.

When studying archived web entities which are not constituted by smaller units related to each other by such technical means as hyperlinks – for instance a web element or a web page – the main challenge is related to the issues of completeness. But when the object of study is web entities which are composed of technically interrelated units – for instance a website composed of hyperlinked web pages, or a web sphere or the web as a whole composed of hyperlinked websites – the main challenge is not only the issues of completeness, but also that relations between the web entities may be inconsistent in the web archive. These two main types of specific challenges shall be illustrated.

4.3 Web Elements and Web Pages

For a web historian who sets out to study how video (web element) or page layout on news outlets (web page) has developed since 1995 the major challenge will probably be that things are missing in the archive, or that choices have to be made between versions. It is impossible to archive streaming video which is a problem in itself just as it affects all related uses of the video such as embedded video (the linking to and integration of a video from, for instance, a

video sharing service such as YouTube). Thus, the link reference to an embedded video may be in the web archive but not the link target, the video itself.

In the example of page layout of news outlets on the web the entire web page may not be in the web archive at all which is often the case, if it is a page far below the front page, or it may not have been archived at the desired points in time (hour, day, month). In addition, the page's web elements (video, images, sound) may be missing, or the entire style sheet keeping the elements in place is not archived and the result is that only running text is shown, without any placeholders.

If one wants to make quantitative studies of how the number of videos on specific websites has changed or qualitative studies of the video content or of the web pages (e.g. visual, argumentation, or rhetorical analysis) then the above mentioned incompleteness obviously constitutes a problem.

However, in some cases the digital nature of the archived web can help the web historian, and this is where digital analytical tools may be relevant to use. For instance, if the numerical changes in the use of video from 1995 to 2005 is to be studied, the videos themselves can be found by searching for specific file types (.mov, .avi, etc.), but even if the videos are not archived the link reference to an embedded video can be found, and this reference can be an acceptable source, since it testifies to the fact that a video was actually shown on the web page. In short, if the general access to the web archive through the web address is supplemented with sophisticated search queries searching for specific file types or strings of source code even an incomplete web archive may be of use to the web historian. However, this procedure will not be of much help for qualitative semantic studies where the concrete object of study is needed.

4.4 Website, Web Sphere, and the Web as a Whole

When the object of study is neither the web element nor the webpage the problems related to completeness are coupled with another set of challenges revolving around the hyperlinked nature of the web as seen on websites (hyperlinked web pages), within the web sphere (often, but not exclusively hyperlinked websites), and on the web as such (often, but not exclusively hyperlinked web material in general). Since the hyperlink creates a contact between two entities the danger of temporal and spatial inconsistencies between these two entities is imminent.[13]

[13] Rosenzweig maintains that the hyperlink is "the unit of analysis for the computer scientists" and asks: "What is the appropriate unit of analysis for historians?" (Rosenzweig 2003, 760). However, hyperlinks do have a history of their own and can thereby (also) be considered a unit of analysis for historians.

Concerning historical analyses of the website, the web historian may set out to map the changing structure of a website from 1995 to 2005, or he may want to identify the most central sections of the website, based on an analysis of received in-links (made manually or most likely by the use of digital analytical tools). Both studies can be part of more detailed analyses of the website's content, however, content studies usually brings web elements or web pages in focus and not the network of web pages which constitute the website. Since the structure of a website as well as the centrality of specific sections are based on hyperlinks, it constitutes a problem if link source and link target are not archived simultaneously, if both are not archived, and if there exist more versions of some web pages and not of others: which of the versions should be considered link source and link target, respectively? Furthermore, since the study includes the website as it looked between 1995 and 2005, it may be complicated by what could be termed 'the inconsistency of inconsistencies', that is the fact that it is not necessarily the same web pages which are missing – or are duplicated – in each of the years.

When moving from analyses of individual websites to the web sphere these challenges are aggravated, since the web sphere is usually constituted by a much more complex and widespread network of hyperlinks than is an individual website. The temporal and spatial inconsistencies potentially increase, and it is therefore unlikely that a web historian who wants to study the historical changes of the hyperlinked networks on the web in relation to, for instance, parliamentary elections in 2001, 2005, and 2011 – based on material in web archives – will find all nodes of the network archived at the same point in time and in the same depth. But it is very likely that more versions of 'the same' are found. Thus, the subsequent network analysis may be based on either a temporally inconsistent set of link sources and link targets, or on a spatially inconsistent set of web pages from variable depths, or both. And since what is mapped is the developments of the hyperlinked networks in 2001, 2005, and 2011 the inconsistency of inconsistencies mentioned above adds to these complications, but now in a more serious way because of the size of the network. In any case, the entire link structure becomes inconsistent, and a systematic comparison of networks over time may therefore be impeded.

With a view to handling the temporal inconsistency the web historian must choose whether the study should be based on material archived within a very short interval of time, thus minimizing the temporal inconsistency, or if the opposite path should be taken, that is including material archived within a larger time interval, thus increasing the risk of temporal inconsistency. Neither of the solutions are cost free, the first may come at the expense of a very small number of websites to study because less web material has probably been archived, whereas the latter may be threatened by a higher number of different versions because the more that is archived, the more possible versions of 'the same' may be found.

The spatial inconsistency is difficult to handle because it can be very hard to determine if all websites of the web sphere were archived in the same depth and how big each website actually was in the past.

In addition to the possible inconsistencies of the network, the complexity and the size of a web sphere's network of hyperlinks usually implies that the analysis cannot be done manually but must be performed by the use of digital analytical software tools which adds yet another challenge. First, one cannot be sure that the analytical software runs without problems on archived web material, second, it is probably not designed to handle more versions of a web page, which is why the corpus must be carefully prepared for network analysis, and it must probably also be separated from the web archive, for instance by extraction.

The result of these inconsistencies in relation to the web sphere may be a biased network analysis which is not as accurate as it would have been had it been made on the online web. To add insult to injure it can even be hard to determine, if, how, and to what extent the analysis is actually biased, mostly because of the lack of documentation.[14]

Finally, a web historian may want to study the web as a whole, for instance a national web domain such as .de. A national web domain can be considered the background on which all other web activities unfold, and therefore it may be relevant to study how it has developed by focusing on, for instance, the most widespread file types, the number of domain names, the number and the average size of websites, or on the 100 most central websites.[15]

Although the problems related to completeness, inconsistencies, and the use of digital analytical tools remain – and in some cases are aggravated compared to studies of the website and the web sphere – they are in some respects less pressing. First, and paradoxically, the large amount of data implies that the problem of versions may become relatively smaller; for instance, in 2012 the size of the entire Danish domain was 28TB, whereas the size or the possible duplicates archived during the same period with selective and event strategies were 1 and 2 TB, respectively. Second, since the snapshot strategy in most cases archives all websites in the same depth (in the Danish web archive up till 25 levels) the spatial inconsistency is minimized. However, the temporal inconsistency remains since, as mentioned above, it takes several months to take a snapshot of a national domain.

[14] See Brügger 2012a about historical studies of hyperlinked networks related to parliamentary elections.

[15] An analysis of the file types on the Danish domain .dk, based on archived web material in the Danish web archive, showed that the relative relationship between written text and images/video was constant from 2006 to 2011 (75-80% and 20-25%), Brügger 2011c.

5. The Future of Web Historiography

Some have argued that with digital media the humanities and the social sciences are on the threshold of a totally new era, whereas others have claimed that not much is new under the sun (cf. for instance Gold 2012). Maybe a more pertinent characteristic would simply be that computers and online networks are probably here to stay, and that they will continue to be an inevitable condition for large parts of academia, as objects of study, as sources, and as research tools. How this condition is going to affect the heterogeneous array of scholarly disciplines and traditions may be the question, not if.

Historiography is probably also here to stay, digital or not. But the continued spread of digitality within academia forces historiography to pay attention to this, be that in the form of digital history, or by being inspired by neighbouring research areas such as internet studies. For the time being – with the web as a dominant digital artifact in society and in academia – the result of such attention may be a web-minded historiography, or a web historiography in its own right.

If web historiography is to progress as a field of study, and as a supplement to and a continuation of digital history as well as internet studies, a close interplay between the following three focus areas is needed. First, reflections as to theories and methods are vital, general as well as with regard to the use of archived web. Second, more empirical studies are needed to challenge existing web archives, and to be challenged by them. Third, the development of digital analytical tools and research infrastructures is essential, web analytical tools in general as well as tools designed to be used in web archives.

If web historians succeed in making this triad of focus areas interplay in a fruitful way, the future of web historiography is on the right track.

References

Abbate, Janet. 2000. *Inventing the Internet*. Cambridge, Mass.: The MIT Press.

Berry, David M., ed. 2012. *Understanding Digital Humanities*. New York: Palgrave Macmillan.

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

Brügger, Niels. 2009. Website History and the Website as an Object of Study. *New Media & Society* 11 (1-2): 115-32.

Brügger, Niels, ed. 2010a. *Web History*. New York: Peter Lang.

Brügger, Niels. 2010b. Web history, an emerging Field of Study. In *Web History*, ed. Niels Brügger, 1-25. New York: Peter Lang.

Brügger, Niels. 2010c. Website history: An analytical Grid. In *Web History*, ed. Niels Brügger, 29-59. New York: Peter Lang.

Brügger, Niels. 2011a. Web Archiving – between Past, Present, and Future. In *The Handbook of Internet Studies*, ed. Mia Consalvo and Charles Ess, 24-42. Oxford: Wiley-Blackwell.

Brügger, Niels. 2011b. 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*, ed. David W. Park, Nicholas W. Jankowski and Steve Jones, 283–98. New York: Peter Lang.

Brügger, Niels. 2011c. WWW 20 år – og stadig mest tekst. *Magasin – fra Det Kongelige Bibliotek* 24 (4): 46-50.

Brügger, Niels. 2012a, forthcoming. Historical Network Analysis of the Web. *Social Science Computer Review*.

Brügger, Niels. 2012b, forthcoming. Web Historiography and Internet Studies: Challenges and Perspectives. *New Media & Society*.

Brügger, Niels, and Maureen Burns, eds. 2012. *Histories of Public Service Broadcasters on the Web*. New York: Peter Lang.

Cohen, Daniel J., Michael Frisch, Patrick Gallagher, Steven Mintz, Kirsten Sword, Amy Murrell Taylor, William G. Thomas III, and William J. Turkel. 2008. The Promise of Digital History. *The Journal of American History* 95 (2): 452-91.

Cohen, Daniel J., and Roy Rosenzweig. 2006. *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia: University of Pennsylvania Press.

Consalvo, Mia, and Charles Ess, eds. 2011. *The Handbook of Internet Studies*. Oxford: Wiley-Blackwell.

De Maeyer, Juliette. 2012, forthcoming. Towards a hyperlinked Society: a critical Review of Link Studies. *New Media & Society*.

Dutton, William H., and Paul W., eds. 2010. *World Wide Research: Reshaping the Sciences and Humanities*. Cambridge, Mass.: MIT Press.

Ess, Charles, and Mia Consalvo. 2011. Introduction: What is 'Internet Studies'? In *The Handbook of Internet Studies*, ed. Mia Consalvo and Charles Ess, 1-8. Oxford: Wiley-Blackwell.

European Science Foundation (ESF). 2011. *Research Infrastructures in the Digital Humanities*. Strasbourg: European Science Foundation (ESF).

European Strategy Forum on Research Infrastructures (ESFRI). 2006. *European Roadmap for Research Infrastructures. Report 2006*. Luxembourg: Office for Official Publications of the European Communities.

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

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

Hauben, Michael, and Hauben, Ronda. 1997. *Netizens: On the History and Impact of Usenet and the Internet*. Washington: IEEE Computer society.

Henderson, Harry. 2002. *Pioneers of the Internet*. San Diego: Lucent Books.

Hunsinger, Jeremy, Lisbeth Klastrup, and Matthew Allen, eds. 2010. *International Handbook of Internet Research*. Dordrecht: Springer.

The Information Society. 2005. Special issue on ICT Research and Disciplinary Boundaries: Is 'Internet Research' a Virtual Field, a Proto-Discipline, or Something Else? 21 (4).

Jacobson, Susan. 2012. Transcoding the News: An Investigation into Multimedia Journalism published on nytimes.com 2000-2008. *New Media & Society* 14 (5): 867-85.

Jankowski, Nicholas W., ed. 2009. *e-Research. Transformation in scholarly Practice*. New York, London: Routledge.

Kimpton, Michele, and Jeff Ubois. 2006. Year-by-year: From an Archive of the Internet to an Archive on the Internet. In *Web Archiving*, ed. Julien Masanès, 201-12. Berlin: Springer.

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

Naughton, John. 2002. *A brief History of the Future: The Origins of the Internet*. London: Phoenix.

Poole, Hilary W., ed. 2005. *The Internet: A historical Encyclopedia*. Santa Barbara: ABC/Clio.

Rosenzweig, Roy. 2003. Scarcity or Abundance? Preserving the Past in a digital Era. *The American Historical Review* 108 (3): 735–62.

Rosenzweig, Roy. 2004. How will the Net's History be written? Historians and the Internet. In *Academy & the Internet*, ed. Helen Nissenbaum and Monroe E. Price, 1-34. New York: Peter Lang.

Schneider, Steven M., Kirsten A. Foot, and Paul Wouters. 2009. Web Archiving as e-Research. In *e-Research*, ed. Nicholas W. Jankowski, 205–21. New York: Routledge.

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

Svensson, Patrik. 2011. The Digital Humanities as a Humanities Project. *Arts and Humanities in Higher Education* 11 (1-2): 42-60.

Thaller, Manfred, ed. 2012. Controversies around the Digital Humanities. *Historical Social Research* 37 (3).

Thomas, William G. III. 2004. Computing and the historical Imagination. In *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens and John Unsworth, 56-68. Oxford: Blackwell.

Weber, Matthew S. 2012. Newspapers and the long-Term Implications of Hyperlinking. *Journal of Computer-Mediated Communication*, 17 (2): 187-201.

Wellmann, Barry. 2011. Studying the Internet through the Ages. In *The Handbook of Internet Studies*, ed. Mia Consalvo and Charles Ess, 17-23. Oxford: Wiley-Blackwell.