Website history and the website as an object of study

Niels Brügger

New Media & Society

New Media & Society

http://nms.sagepub.com

Website history and the website as an object of study

Niels Brügger

New Media Society 2009; 11; 115
DOI: 10.1177/1461444808099574

new media & society

Los Angeles, London, New Delhi, Singapore and Washington DC
Vol11(1&2): 115–132 [DOI: 10.1177/1461444808099574]

ARTICLE

Website history and the website as an object of study

NIELS BRÜGGER
University of Aarhus, Denmark

Abstract

This article puts on the agenda one of the fundamental theoretical questions within the emerging field of website history: how can the object of historical study – the website – be delimited? Its focus is on the 'website' artefact as a medium and a text. After elaborating a definition of the website, as well as discussing how the website is distinct from other possible analytical web objects (the web as a whole, web sphere, webpage and textual web element), the article addresses the challenges of delimiting the archived website. Finally, it outlines some of the key issues in a general discussion of website history.

Key words

medium • text • web • web archiving • web history • web sphere • website

Today, no one would dispute that the internet has been an important part of our communicative infrastructure for some years. Nevertheless, internet history is a relatively blank page, not to mention the subdiscipline of website history, which can be considered an emerging discipline at the intersection between media history and internet history, and which regards the individual website as the unifying entity of historical analysis. Although the number of website studies of all kinds has been growing for several years, genuine historical studies of websites are almost non-existent.[1]

This article puts on the agenda one of the fundamental theoretical challenges within the field of website history: how can the object of historical study – the website – be delimited, both on the web and as an archived website? The aim is to contribute to the development of a theoretical framework enabling the website historian to discuss if and how the bits and pieces of web material that can constitute a website are kept together, so that they form a coherent and delimited object of study.[2]

GENERAL APPROACH

In order to be more specific about the kind of theoretical framework to be discussed, it may be fruitful to introduce a distinction between two types of theories, depending on what they aim at understanding.

On the one hand, we have what could be called theories of objects, which are directed towards how the object of study can be delimited in a systematic and consistent way: for example, theories of what can be understood by 'media', 'communication' and 'the social'. On the other hand, we have theories of explanation, which try to explain and understand the object: for example, the use of semiotics, rhetorical theories, discourse theories, theories of social interaction and the like, which are used with a view to explaining the function and use of media, communication or the social. It can be argued that the former type of theory precedes the latter: we have to know what we study before we can explain it. Thus, since this article sets out to reflect on how to identify the website as an analytical object, it can be regarded as an outline of a theory of the object 'website', while other kinds of theoretical approaches must follow in order to explain the various aspects of websites in more detail. Therefore, the following theoretical framework as well as the reflections on the archived website might be relevant also for website analysis in general, not only for website history. First, website studies need to be more precise about their object of study; second, in practice, most website studies are based on archived websites in order to have a stable object to study and to wcto refer.[3]

Possible analytical objects of media studies

If website history is considered to be a natural part of media history, in continuation of the history of newspapers, film or television, for example, an obvious starting point for the theoretical discussion is the possible delimitations of the object of study within media studies and media history in general.[4]

When faced with a given media artefact, media studies can focus on five different constituents: sender, medium, text, receiver and context, and the mutual relations between these. Although these constituents resemble a communication model, they should be understood as a schematic

representation of the possible analytical objects of media studies and not as a model of communication, seeking to explain the elements and phases in a communicative process (see Brügger, 2002b). For example, with regard to 'television', the media scholar can decide to study the media organization (sender), the technical devices used for production, distribution and reception (medium), the reception of programmes (receiver), specific programmes or the flow of programmes (text), the societal context (macro-context) or the immediate context of use (micro-context). Finally, the relations between the above-mentioned constituents can be studied, for example, the question of how the technical side of the medium affects the producers, the text and its reception.

In particular, the medium itself is important, partly because it binds all the other objects together and partly because its material manner of existence sets up the frameworks for its use in offering a media-specific field of possibilities, thus making possible (or impossible) different ways of acting as sender or receiver or of textuality 'within' the medium.[5] For example, live TV always has been a possibility in making television, and in the early years this potential was used whenever what was aired was not kept on film, since videotapes had not yet been invented. With the advent of tapes the medium changed and live TV began to play a less predominant role, although it remained a potential for many years. This potential was actualized again and used extensively with the invention of satellite television (of course, many economic, organizational and other factors played a role in this development).

Taking these general considerations as a point of departure, it becomes clear that the question of how to delimit the website as an object of study is related primarily to the manner of existence of the 'website' as an artefact, which is why the present focus must be on the medium and the text rather than on the sender, receiver or context. One might argue that leaving the users (sender or receiver) out of consideration is problematic when dealing with the website, since the existence of hyperlinks implies that one has to take the concrete movements of the user into consideration in order to understand the website. This article would argue, first, that this question is not specific to the web, and second, that the signifying elements in newspapers, books, television or on websites can be studied without involving a concrete reader or viewer's way of turning the pages, looking at it or jumping from one window to another. What we study, then, is the text as it presents itself as a field of possibilities to be used (a study that should be an integrated part of any study of the actual use). The user always confronts a text that is already structured in a specific way with specific possibilities, no matter how it is actually used by the user. Thus, the concrete use of the text is considered to be a possibility for an analysis of the medium and the text, rather than a necessity. It could be maintained that when the website historian studies the

text as a field of possibilities, they actually take the viewpoint of the site visitor. However, the website historian's intention is not to visit the website or to re-establish the visitor's experience; rather to analyse the website, at first with a view to delimiting it as an object of study. Therefore, at this stage the website is not approached as an entity experienced by a concrete visitor, but as an object of analysis in its own right. The result of this can be used later as a stepping stone to analysing how the website was actually experienced, often based on other sources than web material.

Medium and text

However, when the website as a medium and a text is in focus, one has to reflect on how it is possible to draw a clear line between the two. This article would argue that the distinction lies in the approach taken by the media scholar, rather than it being inherent in the media artefact. Faced with a newspaper, for example, the focus can be on the paper, the binding and the ink as either physical, material objects, or as an object that through lines, shapes and colours form units endowed with significance (e.g. characters, still images); finally, the focus can be on the relation between material object and signifying units insofar as the paper and ink enhance the use of letters and exclude the use of moving pictures and sound (see Brügger, 2002b). The artefact is the same, but in the first case we consider it a medium, and in the second a text, the difference being whether it is being focused upon as signifying or not.

However, this clear distinction is complicated by digital media, since with the computer a part of the medium is also a text. Today the computer uses an alphabet with two letters (0 and 1) which can be combined on different levels in accordance with certain syntactic rules, thus constituting the grounds for what we see on the screen or hear via the speakers (see Finnemann, 1999). Therefore, an analysis of the computer can focus on the same artefact in three ways. First, the computer-as-medium, that is, the 'box' with the computer, the screen and dots of light and colours on it, the speakers and so forth; second, the computer-as-text, or the dots of light and colours on the screen or soundwaves in the speakers as organized elements of significance; third, the computer-as-medium/text, or the intermediate textual level that the codes of 0/1 and their syntax constitute – a level that is in itself a text, insofar as it is composed of letters and a syntax (see the discussion of materiality and immateriality in Brügger, 2002a).

In this way, digital media are distinctive by embodying a level which can be viewed as both medium and text. However, the text written with the letters 0/1 is not visible and decipherable to the end-user; this is only the case with the text on the screen that is its product. Further, if the distinction between medium and text is based on the fact that in order for something to

be a text it has to be signifying to a reader, then the text written with the letters 0/1 is part of the medium rather than of the text. Still, this by no means implies that this medium/text is unimportant; on the contrary, it is highly important, since as part of the medium it is a decisive part of the conditions for the textual elements.

ELEMENTS OF A WEBSITE ANALYSIS

In light of these general considerations it is argued that, at minimum, the following three elements should be part of a historical analysis of the website:

- its media environment;
- its textual environment; and
- its textuality, that is, the textual elements that constitute the website as a textual phenomenon *strictu sensu*.[6]

Media environment of the website

The media environment in which the website is embedded is the internet, and it must be included in historical analysis of the website since it constitutes the array of possibilities and constraints that frames the other two elements, i.e. the textual environment and the textual elements that constitute the actual website.

Although the internet might seem of no particular interest for a study of the website, its specific manner of existence becomes immediately apparent when undertaking historical studies of five or 10-year-old websites. In 1995, graphics were not in general use, just as streaming of sound and moving images were not possible, and in order to be able to explain this, the internet must be implicated. This article proposes an analysis of the internet as a medium by tracing the historically specific articulations of what can be considered the invariant traits of the internet.

All forms of the internet have the following three characteristics. First, computers must be a component; second, computers are directly connected; and third, when connected, they are able to recognize one another and communicate by means of a shared system of addresses and a shared language: a protocol. Since computers are an essential part of the internet, it 'inherits' the invariant properties characterizing the computer: a mechanical (binary) alphabet consisting of a finite set of letters, each of which is void of semantic content; an algorithmic syntax; and an interface (cf. Finnemann, 1999). None of these invariant traits are necessary in a specific form, but they are necessary as such.

These invariant traits will always be materialized in concrete artefacts that are historically, socially and culturally specific. Concerning the computer (the first of the three traits) we must study the mechanical alphabet (for the computers of today there is not much to say on this point; the alphabet, which

is binary: 0/1); the algorithmic syntax, which varies with the development of software (applications and programs on different layers); and the interface, where some of the following points are of interest: the 'box' with the computer, the screen, the pointing or clicking device, the keyboard, visual in, audio in, the sound source and various storage devices. Concerning direct connectedness, the following are of interest: the devices that connect, the computers that act as 'intermediate stations' by carrying out tasks in relation to the direct connectedness, of which some are specific to the web, and the actual architecture of the connectedness. Finally, concerning the shared system of addresses and shared language, one could focus on internet protocols, protocols for transfer, identification and localization and various shared languages.

Textual environment of the website

The website is not only embedded in a media environment, but also in an environment of signifying elements, or rather two textual environments. The first of these is related to the use of the computer as such, namely, the graphical user interface, with its two key elements which are of great importance for understanding the website: the desktop with clickable and movable icons, folders and the like which serve as a background for the website, and scalable, movable and overlapping windows.

In addition to the desktop and window, which are there no matter what the computer is used for (at least in the case of most personal computers for some years now), the website is embedded in a textual environment that is a necessary and specific condition for the use of the web, namely, the web browser and additional software (however, both can be used for browsing and opening files on the computer). The web browser makes it possible to communicate with web servers, using transfer and location protocols (e.g. http, URL) and to translate what is written in a shared language (e.g. HTML) to letters, images, sounds and positioning in the window. The web browser in itself normally features textual elements that might be of importance in order to analyse a website: for example, the 'Back' button, the 'Go to/History' in the menu bar and the address bar with the URL that all can be elements of navigation *de facto*, even if as such they are not part of a specific website. The variety of additional software allows for the processing of different elements of expression (e.g. QuickTime player, Windows Media players, Adobe Acrobat Reader), and these all constitute textual environments of their own (words, graphics) which are not part of a specific website.

As was the case with the media environment, the textual environment might seem of little interest, but it nevertheless plays a role as a framing facility, which again historical studies will surely make us recognize, since the desktop, browsers and additional software have changed significantly in the course of history.

Textuality of the website

When confronted with a fragmented variety of textual elements that come and go in our browser windows, how can we draw clear demarcation lines that make it possible to delimit the individual website? In order to identify the textual constituents of the web as well as their coherence, this article discusses the following basic concepts: textual element; morphology and syntax; and levels of morphological and syntactical interrelations.

The basic unit of analysis is the textual element, which is a defined coherent textual unit composed of one of the following four formats of expression: written letters (or other characters), still images, moving images and sounds. Textual elements could be a headline and the body text (writing), a photograph (image), a banner advert, a news story from television (moving images), or a piece of music (sound). A textual element can be composed of smaller textual elements (for example, body text can consist of several paragraphs, which can be considered to be individual textual elements).

Analysis of the textual elements can unfold in two dimensions: morphological and syntactical. The morphological analysis focuses on the characteristics of the individual element: how and by what means is a still image constructed? How are the interior structure and coherence of written textual elements established? The syntactical analysis focuses on the rules governing the combinations of elements, as well as on the functions of these combinations, in the sense of the relations between identical types of elements as well as varying combinations (the relation between headlines and body text, images and captions, or written words in images, etc.); and finally, the syntactic dimension can focus on the overall combination of all the types of elements in the complete audiovisual composition.[7]

The interrelations between several textual elements (syntax) as well as the coherence within an individual textual element (morphology) can be studied on three different levels. First, the relations between textual elements (or the limits of an individual textual element) are established through semantic interrelations: for example, by the use of textual cohesion and coherence; the grammatical, lexical construction of a coherent 'world', isomorphy between semantic entities and a global menu structure. Second, the relations between textual elements (or the limits of an individual textual element) are established through formal interrelations, i.e. the forms of expression. For example, the relations between elements of writing are made formally coherent through the formal traits of the words and letters, their positioning (imposed lines of the same length), typography (font, colour, size) and the immediate environment (different forms of separators such as vertical or horizontal lines, boxes, scroll bars); or, for example, the limits of the image element are established by different types of framing. Third, the relations between textual elements (or the limits of an individual textual element) are

established through physically performative interrelations, i.e. the possibility of physically performing an action or a movement, often between elements that are not immediately visible. This could be different forms of continuous movements (vertical or horizontal scroll) or discontinuous 'jumping' (click, hyperlink, mouse-over); thus, the physically performative relationship adds a concrete possibility of action to the other textual elements. However, it is always dependent on the ability to 'graft' onto something which has been expressed already semantically or formally – a word, an image – which is why, for example, a hyperlink never exists without a textual 'substantiation' and must be understood in close relation to this. So moving around in or between textual elements is a semantic, formal as well as a physically performative activity.[8]

With this vocabulary in mind, it can be maintained that from a textual point of view, the website is a coherent textual unit that unfolds in one or more interrelated browser windows, the coherence of which is based on semantic, formal and physically performative interrelations (the same criteria can be used to draw internal demarcation lines in a website, thus delimiting subsites). It could be added that the more widespread these three types of interrelations are, the greater the coherence of the website. Thus, when the website historian aims to characterize the variety of web material in their browser windows, they should try to answer the following three questions: does the web material treat the same subject (semantic cohesion)? Does it look alike (formal cohesion)? Is it possible to come from one browser window to another (physically performative coherence)? If these three questions are answered in the affirmative, we can call the web material a website. However, one of the answers might be a 'Yes, but …', and in this case we might still consider the web material a website if, for example, it consists of a 'core' of coherent webpages with some loosely attached webpages 'around' it, or it is not very coherent but coherent enough to be considered a website. In this sense, coherence is a matter of degree.

THE WEBSITE AND STRATA OF THE WEB

The above theoretical framework can assist the website historian in the task of discussing and maybe even determining to what degree the confused mass of content in web browser windows form a website. It also makes it possible to delimit the website from other possible analytical entities on the web. Within the variety of web material that constitutes 'the web', one can distinguish the following five analytical strata: the web as a whole; the web sphere; the individual website; the individual webpage; and an individual textual web element on a webpage, such as an image.

The concept 'web sphere' was coined by Schneider and Foot and is defined as 'not simply a collection of websites, but as a set of dynamically defined

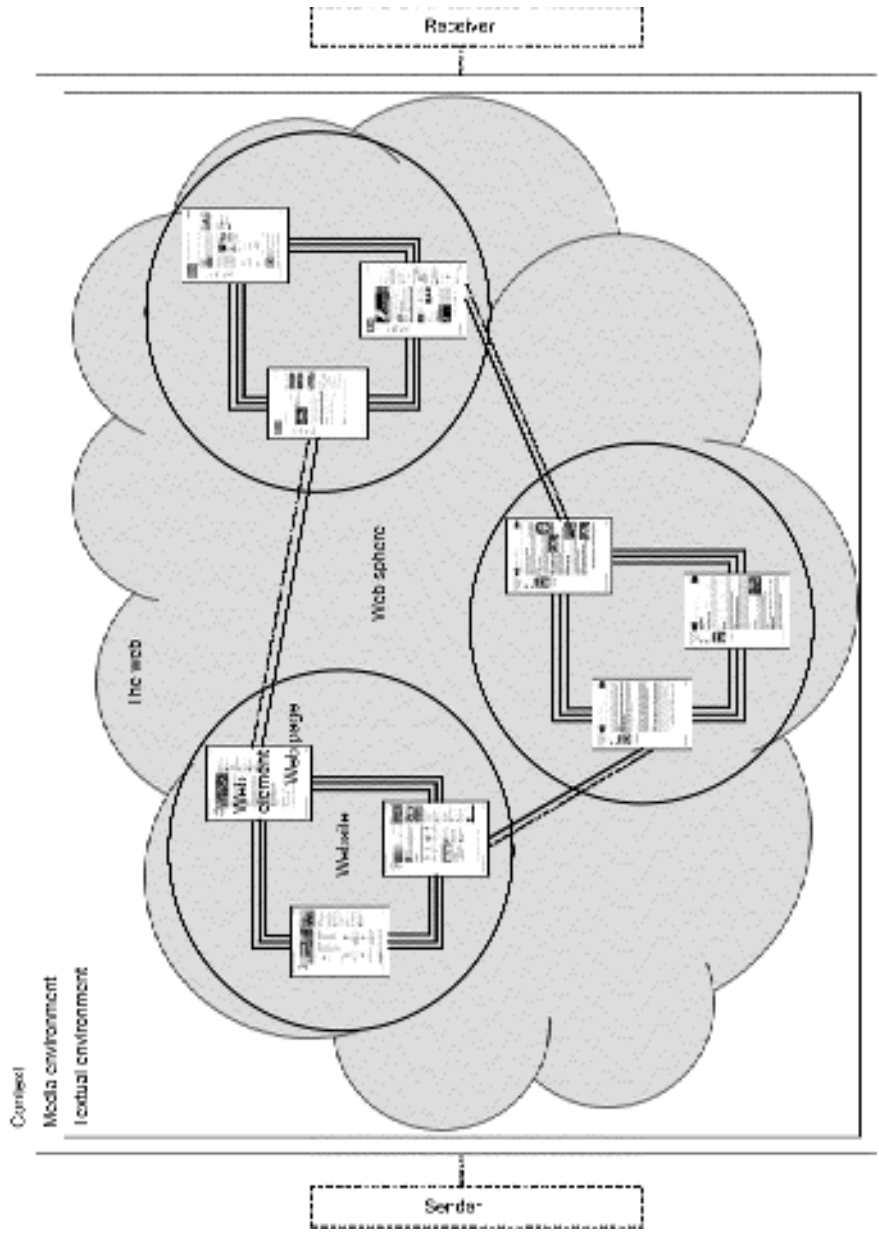
digital resources spanning multiple websites deemed relevant or related to a central event, concept, or theme' (Schneider and Foot, 2006: 20). For example, a web sphere could be web activity in relation to a political event such as the US elections of 2000, 2002 and 2004 (see Schneider and Foot, 2006). As the quotation shows, the web sphere is constructed by the researcher, by selecting websites based on their relevance to the theme in question. On a general level, this is in accordance with the use of the concept in the present article. However, in contrast to Schneider and Foot, the main concern of the present theoretical framework is not the use of websites by actors, but how the researcher can determine that the concrete web material before them is a website, and that it treats the theme in question. The starting point of web sphere analysis – that is, the identification of the web objects of the web sphere (see Schneider and Foot, 2006) – must be based on systematic reflections on how the website as well as the themes are established and supported by textual means such as coherence and creation of meaning (semantics).

Among the five clusters of textuality mentioned above, the website stands apart in the following ways (see Figure 1).

It is clearly distinct from the web as a whole, since the web as such is not supported by any kind of consistent semantic, formal or physically performative interrelations. The website is also clearly distinct from the web element and the webpage: the web element is the minimal textual element on a webpage, while the webpage is clearly delimited by the individual browser window (rarely are webpage and website identical; however, a website consisting of only one webpage is a possibility). Finally, the website is distinct from the web sphere insofar as the website is based on a semantic, formal and physically performative coherence, while the web sphere is characterized only by partial semantic and/or performative interrelations (see the two lines between websites in Figure 1). The 'digital resources' that span 'multiple websites' (Schneider and Foot, 2006) constitute a coherent analytical object because they share a semantic world (the 'central event, concept, or theme' that they treat) and/or because they are linked together; but the relations between them are weaker than those of the website, partly because there are no formal relations and partly because neither the semantic nor the performative relations are necessarily expressed in the same explicit manner on all of the websites that are part of the web sphere.

In summary, the website can be delimited as an analytical object by stating that it is a coherent textual unit which:

- unfolds in one or more interrelated browser windows, the coherence of which is based on semantic, formal and physically performative interrelations;
- is established on the background of, and in an interplay with, a certain internet and textual environment; and



• Figure 1 The elements of a website analysis

- is clearly distinct from the web as a whole, as well as from the web sphere, webpage and web element.

It could be argued that the website can be delimited by other means, such as the web server, the producer of the website or the domain name. However,

the website is not always consistent with the web server and vice versa, since one web server can host several websites, just as one website can span several web servers. The producer of the website can be seen as a unifying editorial entity that exerts overall editorial control and coordination over the webpages, but this editorial role is only useful as a demarcating mechanism insofar as it is expressed on the website through, for example, the use of a logo, footer or disclaimer on each webpage, or by means of a consistent link structure: in other words, insofar as it is expressed by textual means. In fact, the same goes for the domain name: it can constitute a unifying element, but it is a textual web element in line with the other textual elements that create the unity of a website. Although the web server, producer of the website and domain name cannot be considered to be the demarcating mechanism of the website, all three constitute important supplements to the cohesive agent that is created by the semantic, formal and physically performative relations between textual elements.

THE ARCHIVED WEBSITE

Even if the theoretical vocabulary outlined above can help the website historian in the task of delimiting the website as an analytical object in general, this step forward is challenged immediately when the object of historical study is accessible only as an archived website, which is almost always the case. As will be shown later, the archived website differs significantly from the website on the live web as well as from other known archived media types (by 'live web' it is meant 'what is on the web today'). Therefore, the question 'How can the website be delimited as an object of study?' must be supplemented by the question 'How can the archived website be delimited?'[9]

No matter how an archived website has been created and no matter which archive it is found in, the website historian cannot expect it to be an identical copy on the scale of 1:1 of what was actually on the live web at a given time. This is because, on the one hand, the archived website is an actively created and subjective reconstruction, and on the other hand, it is almost always deficient. The combination of these two characteristics has a bearing on the extent to which it is possible to delimit temporally the archived website in a consistent manner.

An actively created and subjective reconstruction

An archived website is always an actively created subjective reconstruction. First, the simple fact that a choice has to be made between different archiving forms and strategies, both in general and in detail, implies that the archived website is based on a subjective decision of either an individual or an institution (see Brügger, 2009b, forthcoming, on different archiving forms and

strategies). For example, it must be decided which form of archiving is to be used, where the archiving should start, how far away from the starting point the archiving is to continue (both in depth and breadth), whether specific file types are to be included or excluded (e.g. images, sounds, flash), whether material is to be collected from other servers and how the material is to be preserved, both here and now and in the long-term perspective. Second, the archived web document is a reconstruction in the sense that it is recreated on the basis of a variety of web elements that stem from the live web and which are reassembled and recombined in the archive.

Thus, the archived website is the result of an active process taking place at the nexus of the 'raw material' which is present on the web and a number of choices with regard to selecting and recombining the bits and pieces at hand. In this sense the archived website does not exist prior to the act of archiving; it is only created in a stable form through the archiving process. Thus, the archived website stands apart from other media types. When archiving newspapers, film, radio and television, the main choices are related to selection of the material, while the archiving process itself almost exclusively consists of taking a copy out of circulation and storing it; no matter who stacks the newspapers or presses the record button on the video recorder, the archived copies are identical to what was once in circulation, just as all copies are identical. In contrast, with websites, choices have to be made in relation to selecting and archiving and we always do more than just remove the web material from circulation; the material is never totally unchanged.

Deficiencies

When compared to what was on the live web, the archived website is almost always deficient. Apart from the deficiency caused by deliberately chosen omissions, two other sources of deficiency can be singled out: those caused by technological problems during the process of archiving, and those related to time. The archived website is very likely to be deficient for technical reasons (software or hardware). For example, words, images or graphics, sounds and moving images can be missing, or some of the possibilities of interaction can be inoperative in the archived web document.[10]

Apart from the technical problems, one of the major reasons for deficiency is related to what could be called the dynamic of updating, i.e. the fact that the web content might have changed during the process of archiving and we do not know if, where and when this happens (see Brügger, 2005; Masanès, 2006; Schneider and Foot, 2004). A brief example can illustrate this:

> During the Olympics in Sydney in 2000, I wanted to save the website of the Danish newspaper *JyllandsPosten*. I began at the first level, the front page, on

> which I could read that the Danish badminton player Camilla Martin would play in the finals a half hour later. My computer took about an hour to save this first level, after which time I wanted to download the second level, 'Olympics 2000'. But on the front page of this section, I could already read the result of the badminton finals (she lost). The website was – as a whole – not the same as when I had started; it had changed in the time it took to archive it and I could now read the result on the front page, where the match was previously only announced. (Brügger, 2005: 22–3).

As this example illustrates, it is obvious that the archived website is deficient, since it is incomplete compared to what was once on the live web: something is lost in the process of archiving due to the asynchrony between updating and archiving. However, it is also deficient in another and less obvious way, in that the archived web document is not only incomplete, but is also 'too complete': something that was not on the live web at the same time, the content of two webpages or website sections, is now combined in the archive and it is difficult to determine what the website was actually like at a given point in time on the basis of these two entities.

In summary, it can be argued that since the actual act of preserving web material almost always changes the material that was on the web in a number of ways, the process of archiving creates a unique version, but not a copy.[11]

Temporal delimitation of the archived website version

As a consequence of these characteristics, a general question presents itself: how can we temporally delimit an archived version of a website? In fact, on the live web, the question of whether a website is 'the same' version or not in temporal terms is not relevant. If we open a webpage and jump to another webpage on the same website by means of a hyperlink, the two webpages are part of the same version simply because they are there at the same time. However, this is not the case with an archived website, since the two webpages might have been archived at different points in time and several archived copies of each of them might even exist, archived at, for example, five-minute intervals. Instead of being strictly related to a distinct point in time, the archived version extends like a time continuum, thus making it difficult to determine not only a temporally coherent version, but also the limits between versions, since in fact one and the same web element can be part of more than one version. The archived websites might appear to be continuous versions with blurred and overlapping borders.

The website historian, then, faces the task of determining whether a variety of webpages are part of one and the same version or not, thus questioning the temporal consistency as well as the temporal limits of the version(s). In this

task they are obliged to choose between two paths. Either a very strict criterion is maintained, that there should be simultaneity between all the archived web elements, or a more pragmatic criterion is accepted, according to which some kind of interval between the web elements is tolerated. In other words, either the version is delimited very precisely, with the result that a lot of versions will consist of only one webpage, or the version is delimited in a more imprecise and inconsistent manner, thereby creating fewer and more 'complete' versions.

To conclude: on the live web the main problem is to delimit the website in spatial terms, while the temporal limits of the website are simply present. However, things work differently when a website archive is entered. Here, not only the spatial but also the temporal limits must be established by the researcher and to a certain extent the archived version is constituted and constructed only *post festum*, after the archiving process. In other words, the subjectivity, active reconstruction and temporal dynamics of updating which characterize the moment of archiving are mirrored in the archive, thus forcing the website historian to delimit the archived website as well as to formulate in detail the methodological deliberations that enable them to do so.

CONCLUSION

The theoretical framework and reflections on the archived website formulated above can be considered a conceptual toolbox that enables the website historian – and the website scholar in general – to discuss to what degree the web material on their screen forms a website in order to create a consistent object of study, as well as to discuss the borders of a website over time – that is, if it had the same boundaries two, five and 10 years ago. However, it can serve also as a starting point for more general reflections on website history. These concluding remarks will outline some of the key issues in this general discussion.

One of the tasks that could be undertaken would be to formulate a historical textual grammar of the website. When the website is defined as above, it would be relevant to ask how the textual elements that constitute the website, along with the semantic, formal and physically performative relations between them, actually have appeared and functioned at various periods in the past in order to identify recurrent patterns and traits. In other words, it is important to examine the ways in which the textual web constituents as well as their combinations actually have been used, so as to formulate the historically changing morphological and syntactical rules of website textuality.

Furthermore, one could point to a cluster of tasks that revolve around a reopening of the different themes which, until now, have been incidental to the focus on specific characteristics of the website. These issues can be

addressed now with a systematic approach to the specificity of the website in mind.

First, it should be discussed whether well-known theories of explanation can be applied unchanged to an object of study such as the website. What does it mean to do rhetorical studies or discourse analysis when, for example, the relations between the fragments of content and expression are performed physically by a hyperlink from one browser window to another, rather than by turning pages or looking at the changing pictures on a television screen? Or when an image is used as a point of departure for a hyperlink, thus making it possible for an image to refer to other textual elements all by itself and in a physically performative sense? Or when written words or pictures suddenly appear with the use of a mouse-over?

Second, both the actual users (sender and receiver) and the context of the website should be reintegrated into historical study, with the historically-specific mediacy and textuality of the website in mind. What role have mediacy and textuality played as fields of potentialities for website users at various points in time in the past? How have the users actually interacted with these possibilities? How has the context of the website – other media, for example – interacted with the website (for example, various cross-media phenomena)?

Third, the website should be resituated in its relation to the other strata of the web, i.e. the web as a whole, web sphere, webpage and web element. These five strata are one another's mutual context, and in many ways they are interdependent, but it could be discussed whether one or more of them constitutes the analytical centre of gravity in historical study of the web. On the basis of the theoretical reflections above, this article would suggest that the website and web sphere be considered the two overriding analytical units on the web, since the other strata are subordinate to them to various degrees, and they are both based on a set of semantic and physically performative relations. The individual web element is difficult to explain without the webpage as its context, just as the webpage is almost always part of a website. The web as such is nothing but a fragmented, heterogeneous and confused mass, if one does not draw some semantic, formal or physically performative lines in it (see Brügger, 2007). However, the website and web sphere are not on equal terms. One may argue that it is possible to analyse an individual website without relating it to a web sphere, but it is less likely that the web sphere can be studied without preliminary identification and study of the websites that constitute the web sphere. In this sense, the website tends to be the encompassing analytical unity on the web and within web history.

As these concluding remarks indicate, the field of website history is still in its infancy and more theoretical research is needed in order to establish the research field. The reflections above are but a small contribution to this task.

Notes

1 Major contributions to internet and web history are Abbate (2000), Gillies and Cailliau (2000), Hauben and Hauben (1997), Henderson (2002), Naughton (2002) and Poole (2005). Some of the few historical studies of websites are Burns (2008), Martin (2004), Schneider and Foot (2006), Tkach-Kawasaki (2003) and Voerman and de Graaf (1998). In addition, currently the author of this article is conducting a research project entitled 'The History of dr.dk, 1996–2006', where the history of the dr.dk website is to be written.

2 In the present context, 'website' is understood as an entity based on the use of protocols and mark-up languages, which originates from the world wide web in the broadest sense rather than other 'clusters' of files on the internet (for example, 'gopher' sites), although they can be analysed probably along the lines outlined in the following.

3 Although one could expect that the fundamental question of what should be understood by 'website' was addressed in the ever-growing number of studies of websites or in general studies of the web, this is not the case. On the contrary, the question is rarely addressed systematically and in depth, if at all (see Burnett and Marshall, 2003; Gauntlett and Horsley, 2004; Kaye and Medoff, 2001; Mitra and Cohen, 1999; Schneider and Foot, 2004 and the brief comments on each of these texts in Brügger, 2007).

4 The following considerations on the analytical object of media studies take as their starting point the reflections in Brügger (2002b).

5 I use the term 'mediacy' for this 'media-ness', the specific being of a medium (see Brügger, 2002b). Mediacy is considered to be a dialectic movement between the potentialities of the medium and the actual use of it. So the medium is both the field of possibilities and the result of its use (see Brügger, 2002b).

6 'Textual' means that the level of analysis is 'higher' than letters or phonemes and sentences. The term 'textual' is understood in a broad sense and is not limited merely to written language, rather it refers to all forms of expression such as written words, still images, moving images and sound.

7 The terms 'morphology' and 'syntax' are used in a broad sense, extending the usual meaning, where focus is on similar elements (written or spoken words) unfolding in one dimension only (a progressive chain of expression). In this context, all kinds of textual elements as well as relations in several directions are comprised in the words 'morphology' and 'syntax'.

8 The word 'performative' is inspired by speech act theories, where the use of a word is the performing of an action. However, in the present context, what is performed is not just a 'speech act' but a physical act: on the web we actually do things with words and images.

9 Web archiving means any form of deliberate and purposive preserving of web material (this definition is elaborated in more detail in Brügger, 2009b, forthcoming; for introductions to and discussions of web archiving, see Brown, 2006, Brügger, 2005, 2008, 2009b, forthcoming; Masanès, 2006).

10 For the test of the quality of archived web documents in various archives, see Brügger (2009a, forthcoming); see also the discussion of quality and completeness in Masanès (2006).

11 This uniqueness of archived versions is confirmed by the test presented in Brügger (2009a, forthcoming), where the appearance of the same website which has been archived on the same date in different archives was examined. Thus, the archived web document must be treated in new ways, compared to other archived media types,

based on a new kind of philology; a small contribution to such a 'website philology' is outlined in Brügger (2008, 2009b, forthcoming).

References

Abbate, J. (2000) *Inventing the Internet*. Cambridge, MA: MIT Press.

Brown, A. (2006) *Archiving Websites: a Practical Guide for Information Management Professionals*. London: Facet Publishing.

Brügger, N. (2002a) 'Does the Materiality of the Internet Matter?', in N. BrÜgger and H. Bødker (eds) *The Internet and Society? Questioning Answers and Answering Questions*, pp. 13–22. Aarhus: Centre for Internet Research.

Brügger, N. (2002b) 'Theoretical Reflections on Media and Media History', in N. Brügger and S. Kolstrup (eds) *Media History: Theories, Methods, Analysis*, pp. 33–66. Aarhus: Aarhus University Press.

Brügger, N. (2005) *Archiving Websites: General Considerations and Strategies*. Aarhus: Centre for Internet Research.

Brügger, N. (2007) 'The Website as Unit of Analysis? Bolter and Manovich Revisited', in A. Fetveit and G.B. Stald (eds) *Digital Aesthetics and Communication*, pp. 75–88. Bristol: Intellect.

Brügger, N. (2008) 'The Archived Website and Website Philology: a New Type of Historical Document?', *Nordicom Review* 29(2): 151–71.

Brügger, N. (forthcoming, 2009a) *Archived Websites between Copies and Versions: Test of Versions in Existing Web Archives*. Aarhus: Centre for Internet Research.

Brügger, N. (forthcoming, 2009b) 'Web Archiving: between Past, Present and Future', in R. Burnett, M. Consalvo and C. Ess (eds) *Blackwell Companion to Internet Studies*. Oxford: Blackwell.

Burnett, R. and P.D. Marshall (2003) *Web Theory: An Introduction*. London: Routledge.

Burns, M. (2008) 'Early ABC Online Interactive Sites. Remembering Public Service Broadcasting: Liberty and Security in Early ABC Online Interactive Sites', *Television and New Media* 9(5): 392–406.

Finnemann, N.O. (1999) 'Modernity Modernised: The Cultural Impact of Computerisation', in P.A. Mayer (ed.) *Computer, Media and Communication*, pp. 141–59. Oxford: Oxford University Press.

Gauntlett, D. and R. Horsley (eds) (2004) *web.Studies*. London: Arnold.

Gillies, J. and R. Cailliau (2000) *How the Web Was Born: The Story of the World Wide Web*. Oxford: Oxford University Press.

Hauben, M. and R. Hauben (1997) *Netizens: On the History and Impact of Usenet and the Internet*. Washington: IEEE Computer Society.

Henderson, H. (2002) *Pioneers of the Internet*. San Diego, CA: Lucent Books.

Kaye, B.K. and N.J. Medoff (2001) *The World Wide Web: A Mass Communication Perspective*. New York: McGraw-Hill.

Martin, F. (2004) 'Net Worth: The Unlikely Rise of ABC Online', in G. Goggin (ed.) *Virtual Nation: The Internet in Australia*, pp. 193–208. Sydney: University of New South Wales Press.

Masanès, J. (ed.) (2006) *Web Archiving*. Berlin: Springer.

Mitra, A. and E. Cohen (1999) 'Analyzing the Web: Directions and Challenges', in S. Jones (ed.) *Doing Internet Research*, pp. 179–202. London: Sage.

Naughton, J. (2002) *A Brief History of the Future: The Origins of the Internet*. London: Phoenix.

Poole, H.W. (ed.) (2005) *The Internet: A Historical Encyclopedia*. Santa Barbara, CA: ABC/Clio.

Schneider, S.M. and K.A. Foot (2004) 'The Web as an Object of Study', *New Media & Society* 6(1): 114–22.

Schneider, S.M. and K.A. Foot (2006) *Web Campaigning*. Cambridge, MA: MIT Press.

Tkach-Kawasaki, L.M. (2003) 'Clicking for Votes: Assessing Japanese Political Campaigns on the Web', in K.C. Ho, R. Kluver and K.C.C. Yang (eds) *Asia.com: Asia Encounters the Internet*, pp. 159–74. London: Routledge Curzon.

Voerman, G. and J.D. de Graaf (1998) 'De websites van de nederlandse politieke partijen, 1994–1998' ['Dutch Political Party Websites 1994–1998'], in *Documentatiecentrum Nederlandse Politieke Partijen Jaarboek 1997* [*Dutch Political Party Central Documents Yearbook 1997*], pp. 238–69. Groningen: Documentatiecentrum Nederlandse Politieke Partijen.

NIELS BRÜGGER is Associate Professor at the Centre for Internet Research, University of Aarhus, Denmark. His primary research interest is website history, and in 2007 he began writing a history of the Danish Broadcasting Corporation's website entitled *The History of dr.dk, 1996–2006*. He has published a number of articles, monographs and edited books, including *Archiving Websites: General Considerations and Strategies* (Centre for Internet Research, 2005) and *Media History: Theories, Methods, Analysis* (ed. with S. Kolstrup, Aarhus University Press, 2002).

Address: Institute of Information and Media Studies, University of Aarhus, Helsingforsgade 14, 8200 Aarhus N, Denmark. [email: nb@imv.au.dk]