



UvA-DARE (Digital Academic Repository)

Doing Web history with the Internet Archive

Screencast documentaries

Rogers, R.

DOI

[10.1080/24701475.2017.1307542](https://doi.org/10.1080/24701475.2017.1307542)

Publication date

2017

Document Version

Final published version

Published in

Internet Histories: Digital Technology, Culture and Society

License

CC BY-NC-ND

[Link to publication](#)

Citation for published version (APA):

Rogers, R. (2017). Doing Web history with the Internet Archive: Screencast documentaries. *Internet Histories: Digital Technology, Culture and Society*, 1(1-2), 160-172. <https://doi.org/10.1080/24701475.2017.1307542>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Doing Web history with the Internet Archive: screencast documentaries

Richard Rogers

Universiteit van Amsterdam, Amsterdam, The Netherlands

ABSTRACT

Among the conceptual and methodological opportunities afforded by the Internet Archive, and more specifically, the WayBack Machine, is the capacity to capture and “play back” the history a web page, most notably a website’s homepage. These playbacks could be construed as “website histories”, distinctive at least in principle from other uses put to the Internet Archive such as “digital history” and “Internet history”. In the following, common use cases for web archives are put forward in a discussion of digital source criticism. Thereafter, I situate website history within traditions in web historiography. The particular approach to website history introduced here is called “screencast documentaries”. Building upon Jon Udell’s pioneering screencapturing work retelling the edit history of a Wikipedia page, I discuss overarching strategies for narrating screencast documentaries of websites, namely histories of the Web as seen through the changes to a single page, media histories as negotiations between new and old media as well as digital histories made from scrutinising changes to the list of priorities at a tone-setting institution such as whitehouse.gov.

ARTICLE HISTORY

Received 30 December 2016
Accepted 13 March 2017

KEYWORDS

Website history; Web history; digital history; WayBack Machine; Internet Archive; screencast documentary

Introduction: digital history, Web history and website history

The Internet Archive and the Web archives of national libraries are thought as sources for “digital history”, which refers to history-writing with Web-based materials (Brügger, 2012; Cohen & Rosenzweig, 2006; Internet Archive, 2017; Rosenzweig, 2003). The creation and maintenance of Web archives often are justified for digital history purposes, considering the wealth of online materials not only compared to other media but also because they encompass them. The argument for the specificity of Web archives thus lies in the growth of “born-digital” materials, in contrast to digitised ones of media archives. It also rests especially upon their use by future historians, when they come to write the history of particular periods, such as the 1990s. The value of the archived Web is thus often thought to lie in its special contents that are otherwise unavailable elsewhere and in its future use by historians, as Milligan notes:

Imagine a history of the late 1990s or early 2000s that draws primarily on print newspapers, ignoring the [Internet] technology that fundamentally affected how people share, interact, and leave historical traces behind (2016).

Internet (and Web) history, on the other hand, may be distinguished from digital history, as it concerns employing the Web to tell its own story, in the tradition of medium or media history. Whilst there are exceptions, Web archives are not as often justified as sources for specific Internet and Web histories (Ben-David, 2016; Goggin & McLelland, 2017; Stevenson, 2016). Moreover, Internet histories may be written largely without them (Abbatte, 2000; Ryan, 2011). Indeed, be it for digital, Internet or Web history, actual historian use of Web archives remains limited (Brock, 2005; Dougherty et al., 2010; Hockx-Yu, 2014; Weltevrede & Helmond, 2012).

How to reconsider the value of Web archives to Web history? The point of departure here is to reintroduce a more specific type of Web history – website history – and put forward an approach to its study and at the same time a productive use of Web archives (Brügger, 2008). That is, the screencast documentary is both an approach to studying website histories and also a means to provide researcher use for Web archives, which itself is understudied (Dougherty et al., 2010). It takes advantage of the organisation of the Internet Archive, and especially the interface and query machine built on top of it to access its contents.

Whilst it recently has added a keyword search, for over a decade now, the WayBack Machine has had as its primary (and default) input field a single URL. Using digital methods, or tool-based methods to extract and analyse Web data and objects for social and cultural research, the screencast documentary approach put forward here captures the outputs of the WayBack Machine (list of archived pages with dates), screenshots the unique ones, and arrays them in chronological order so as to play back the history of the website in the style of time-lapse photography (Rogers, 2013).

In the following, narrations or particular goals for telling the history of a website are put forward. They offer means to study the history of the Web (as seen through a single website like Google Web Search), the history of the Web as media (such as how a newspaper has grappled with the new medium) as well as the history of a particular institution (such as the US White House or marriage, as seen through a leading wedding website). Arguably, the first is a form of Web and medium history, the second media history, and the third digital history, however much each also blends the approaches and blurs the distinctions.

It should be pointed out that the WayBack Machine of the Internet Archive is itself a Web-historical object. In a sense, it also tells the story of the Web, or at least a particular period of it, through the manner in which it primarily grants access to websites. By the default means by which it is queried and also how archived web pages are interlinked, the WayBack Machine of the Internet Archive has organised a surfer's Web circa 1990s rather than a searcher's Web of the 2000s or a scroller's of the 2010s (with a smartphone).

Here, it is argued that the WayBack Machine also lends itself to a particular historiography that is embedded in the screencast documentary approach, namely a single-site or site-biographical method of recounting history. Having developed that argument in brief, the piece concludes with how to put to use the WayBack Machine of the Internet Archive to tell single-site histories as screencast documentaries.

The WayBack Machine: surf the Web as it was, or use the Internet Archive as source

The WayBack Machine of the Internet Archive, with its original slogan “surf the Web as it was”, was conceived and presented in part as a solution to the 404 problem, the response



Figure 1. Alexa toolbar, with WayBack icon to access the Internet Archive, circa 2004. Author screenshot.

code signifying that the file or web page is not found. With the Alexa toolbar installed in a browser (in the 1990s, see [Figure 1](#)), the Web browser user confronted by a 404 error message would receive a flashing WayBack icon on the toolbar that indicates that the missing page is in the Internet Archive. (If the button did not flash, there was no archived version, and the page had been lost.) In return for Alexa's solution to the 404 problem as well as the content at the Internet Archive, the user would aid in populating the archive. That is, when downloading the toolbar, permission would be given to have his/her browsing activity logged, and web pages or sites that a user visited would be sent to Alexa. If a site was not yet in the archive, a crawler would visit it, and thus grew the Internet Archive. Later, high-traffic and other significant sites would be earmarked for regular archiving.

The WayBack Machine's architecture, designed and launched in the mid-1990s, aimed to furnish an ideal surfer's experience, frictionless and without dead ends. Once onto a website in the archive, clicking links takes the surfer to the page closest in time, and if unavailable to the page on the live Web. The surfer jumps through time as if in an atemporal hyperspace, one of the earliest Web metaphors or structuring devices for a document universe without directories or search engines. The WayBack Machine thus sacrifices temporal matching for smooth navigation, and as such embeds a period in Web history, in an experience that could be described as more living museum of a surfer's space than historian's meticulous archive.

Apart from the "way it was" experience, the WayBack Machine is also suggestive of particular research practices and ultimately historiographical approaches. With respect to the research practices, there are largely two afforded by the interface. At archive.org, the `http://` prompt invites the inputting of a single page URL so as to summon its history. At the outset, in other words, one is invited to submit a URL and pursue its history through two outputs, one of which shows minute changes to the contents of the pages in the archive (additions and deletions), and another that invites the exploration of a fuller arc, where one can click backward and forward arrows through larger chunks of the page's history.

In the original results page, asterisks next to date stamps indicate changed content on the web page. One may thus peruse a web page's history to spot the crucial, detailed change (or "diff" in computational language). As a research output, one perhaps would wish to put two or more pages side-to-side, highlighting the specific, telling diff, such as an infringement of one's intellectual property, which is a common use case of the WayBack Machine in the legal arena, discussed in more detail below.

[I]n *Telewizja Polska USA, Inc. v. Echostar Satellite Corp.*, the plaintiff alleged that the defendant was using the plaintiffs trademark name in violation of its intellectual property rights. In response, the defendant introduced the printout of the defendant's archived webpage dated before the plaintiff received the trademark of its brand (Gazaryan, 2013, p. 221).

The form of output navigation for exploring the fuller arc of history is the timeline (see [Figure 2](#)). Instead of pouring over the detailed changes, with the timeline, one makes



Figure 2. WayBack Machine banner that accompanies archived web page loaded in a browser. Example is Myspace.com, indicating the date it changed from a social networking to a music-oriented social entertainment site. Source: <http://web.archive.org/web/20101116021305/http://www.myspace.com/> (accessed 28 December 2016).

a sweep through the interface and content of a web page over the years with an eye towards the broader themes, such as the introduction and subsequent locking down or removal of comment spaces and other interactive features on websites that once made new media new.

The interface to the Internet Archive thus creates at once a surfer's experience from a particular period in Web history whilst also affording modes of historical work that privilege focusing on the minute as well as the sweeping change to a single page.

Digital source criticism

Seen from the perspective of digital history (history-writing with Web materials), the WayBack Machine also could be said to invite the user to seek a specific source, and scrutinise it for its veracity because it is a Web source. Here, with the WayBack Machine, one brings the Web, and its pages, into the evidentiary arena of source criticism. There are at least three sets of questions, or aspirations for the “digitally reborn” sources online now that they appear as web.archive.org URLs rather than in their original name space state (Brügger, 2012). Once captured and put back online, the archived web pages face tests, from a series of scholarly discourses, before they may employed as proper sources. In law do they count as duplicated sources, in the social sciences as valid and in history as sufficient substitutes for missing materials? From the start, one of the more popular use cases for the Internet Archive, apart from the 404-not-found error whilst surfing, has been as evidence (Howell, 2006). One could go back in time to a website for evidentiary purposes, checking for trademark and intellectual property infringements, as was the case with its first-time deployment in US courts in 2003 when print-outs from the WayBack Machine were introduced as exhibits (Eltgroth, 2009). Here, the questions concern the extent to which one can treat the archived page as a duplicate of the original one no longer online, or in a lesser test, at least warrant through testimony that it represents accurately the material the site owner put online. In the event, the archived website need not be a duplicate in code and data to be admissible; rather it only need to be an accurate representation.¹

Apart from its authenticity in legal arenas, a web page faces scrutiny as a source for scholarly referencing purposes, in order to anchor an account of events, for example. In the very first place, the challenge put to the Web as source may rest upon its overall (historical) reputation problem, as a medium of pirates, pornographers, conspiracy theorists and self-publishers (Dean, 1998). As the fake news scandals surrounding the US presidential campaigns of 2016 pointed to anew, it is a space with and without professional editors, and has been subject to the question of its quality, even as the Web further domesticated, in its nearly thirty years of use (Thelwall, Vaughan, & Björneborn, 2005).

More to the point is the question of whether URLs should be referenced in the first place as sources, and if offline, whether a WayBack URL could stand in. Apart from the reputation problem, it is often argued that the Web's ephemerality, or perhaps its uneven maintenance, disproves its worthiness as source. Referenced URLs break, as links rot (Klein et al., 2014; Veronin, 2002). In this context, the WayBack Machine may be viewed as a set of well-tethered (rather than broken) source links. The Internet Archive thus becomes an early attempt at providing permanence to ephemeral Web sources, in a lineage of such attempts from the tradition of hypertext (permalinks in blogs and edit history retention in wikis) to that of library science (DOI numbers). Once accepted as not only references but reference-able, Web sources that break and are reborn in the Internet Archive face further tests. Are the archived ones "valid"? Such a determination relies, among other things, on whether the date stamps of archived web pages, including new archived versions, match the dates of the web pages when online, an issue studied by a series of authors (Dougherty & Meyer, 2014; Dougherty et al., 2010; Murphy, Hashim, & O'Connor, 2007). In the event, the Internet Archive has met validation challenges concerning web page (and thus content) age.

For referencing, a WayBack URL rather supplements than replaces an original URL. In a recent edition to the MLA style guide, even (original) broken URLs should be referenced, with access date, for the reader may be able to "evaluate the credibility of the site that published the source, or locate the source under a new URL" (Gibson, 2016, n.p.). In all, the MLA recommends adding the WayBack URL to the reference after the broken URL, rather than pruning the citation through the use of the archived URL only (Internet Archive, 2016).

For historians, a further test concerns whether a reborn website in the archive was ever online as such in the first place (Brügger, 2012). Websites reconstituted by the archiving appear to be damning critiques of their value as historical sources (Russell & Kane, 2008). Especially newspapers, as proverbial first drafts of history, are susceptible to hodge-podge archival reconstructions, where certain plugged-in content is saved at another time than the front page of the newspaper, and when one recombines it in the archive, the "digitally reborn source" becomes a novel artefact of its archiving process. Even given the missing original, the question steps beyond whether the incomplete, archived source is acceptable, in the spirit of save what one can. When writing digital history, or using the Web as historical source, being a scholar of the history of the Web (and dynamic websites), together with the history of its archiving (and the treatment of dynamic websites), becomes crucial.

Web historiographies in brief

As discussed above, the architecture of the WayBack Machine of the Internet Archive invites website or web page histories, given that one fetches the history of a URL through the interface, and peruses it looking for minor changes with the aid of the asterisks in the classic interface, or with a broad sweep, forward clicking month by month, examining the larger thematic changes to the life and times of the site.

Before introducing examples of website histories, in the style of Jon Udell's pioneering screencast documentary of the edit history of a Wikipedia page, it is instructive to mention that the biographical (in which a website history would fall) is among at least four

dominant traditions of Web archive collection and usage. The second tradition is of a special collection, where typically elections, disasters and changes of power or transitions are archived, such as US elections and the installation of a new pope (Schneider & Foot, 2004). Here, the approach to Web historiography is event-based. In the archiving, there is also an attention cycle to consider, both the run-up to an election and transition as well as its aftermath. Archiving agility (especially for a disaster) is called for.

A third type of Web historiographical approach is embodied in the efforts by national libraries to demarcate and save “national” Webs, beginning with the preservation of the official public record and continuing often with a carefully considered definition of a website of relevance to national heritage (Jacobsen, 2008; Rogers, Weltevrede, Borra, & Niederer, 2013). For example, the Danish, pioneers in Web archiving, define a relevant national website as having at least one of four properties: in the top-level country .dk domain, written in Danish, about a Danish subject matter (e.g. the author Hans Christian Andersen) or material of relevance to the Danish or Denmark, the last type of which expands the material to such an extent that it becomes a matter of editorial selection, bringing the librarians back into Web content curation (after the demise of the online directories).

The fourth one, an autobiographical, is the most recent, and concerns Web properties that are essentially no longer considered websites, at least as we have known them to be as accessible without a password and residing for the most part on an open Web. Whether they are social media platforms or smartphone apps, they are difficult to collect and preserve, and improbable to make accessible at any scale, owing to the fact that they are personal, behind user logins, or have other novel social and technical constraints. Here, the approaches to storing differ in that just the data are captured (e.g. by individuals-requested data dumps from Facebook) or by videorecording a user interacting with her mobile phone. The collection becomes the video together eventually with the smartphone itself. More recently, at Rhizome, the digital arts collective, the “webenact” technique, put online as webrecorder.io, has been developed to capture or record a social media user’s pages so as to re-enact them or play them back. The work was developed on the heels of the critically acclaimed performance piece of the user of Instagram, Amalia Ulman (Rhizome, 2014).²

Website histories and the screencast documentary

From the standpoint of Web historiography, a website history or single-site biography may be understood as the unfolding of the history of the website, and with it a variety of stories may be told. First, the history of a website could be seen to encapsulate the larger story of the history of the Web. In one example discussed in detail below, the history of the changes to the front page of Google.com (in particular, the tabs) may be read as the history of the demise of the human editors of the Web, and the rise of the back-end, of the algorithm, taking over from the librarians. From the history of a website, second, one also may tell the story of the history of media, such as how a newspaper, a radio station, or a television channel grappled with the Web, over time (Bødker & Brügger, 2017). Has the old media form, so-called, embraced new media features, only to settle back into a digitised version of its original self? How have newspapers domesticated the blog, or

tamped down the comment space where readers can talk back to the institution referred to historically as gatekeepers?

In a screencast documentary of the history of *nytimes.com*, the newspaper has experimented repeatedly with new media forms, beginning as a separate entity from the print version, without any reference to the print version or to subscriptions (Hermens, 2011). It was directed at a Web-only audience with such features as “cybertimes” and forums. Often these special new media forms would be jettisoned, though some have remained such as a curated comment space as well as novel newspaper navigation through “most emailed”, “most viewed” and “recommended for you”.

The third strategy is telling the history of an idea, individual, organisation, institution or other entity to which a website has been dedicated. Examining the evolution of the contents of the “issues” tab at *whitehouse.gov* shows at a glance how the priorities of the US presidential administration have changed, sometimes abruptly; after the 9/11 attacks on the World Trade Center and the Pentagon in 2001, most all issues on *whitehouse.gov* included the word “security”, only gradually to broaden their scope in the years to come (Rogers, 2013). In another case, examining the history of *theknot.com* over a ten-year period, researchers found how a simple advice and registry site became a complex wedding planner, multiplying expenses and product placements, concluding that nowadays for weddings “no expense should be spared” (Livio, Mataly, & Schuh, 2012). Thus, one view on the evolution of the institution of marriage, and its commercialisation, may be reconstructed through a single-site history.

Techniques for making screencast documentaries of the history of a web page

There are practical aspects to creating a screencast documentary of the history of a web page. At the Digital Methods Initiative at the University of Amsterdam, colleagues and I have created tools and techniques to compile the archived versions of a web page, so as to assemble them chronologically as a movie. There are four steps: make a list of the archived pages, capture or download them, load them in a moviemaker and record a voiceover. In the first step, to make a list of the archived pages, use the tool, the Internet Archive WayBack Machine Link Ripper.³ One enters the URL to be captured from the WayBack Machine (e.g. <http://www.google.com>), and the tool creates a list of links of its archived pages, removing duplicates by default, and providing options concerning the capture interval (e.g. daily or monthly). To study minute changes to the web page over time, one chooses daily snapshots, and for a fuller arc of history, monthly. In the second step, the WayBack Machine URL list (a text file) is subsequently inputted into a screenshot generator (such as the browser extension, Grab Them All).⁴ Screenshots are made of each archived web page. The pages need to load in the browser for the screenshot to be made, so it is advisable to fine-tune the amount of time between screenshots so as to make sure the pages have loaded before the screenshots are taken. The third step is to load the screenshots into an image viewer such as iPhoto, and make a project in movie-making software such as iMovie (or Windows Movie Maker). Finally, the voiceover is recorded, and the movie is ready for playback.

For the voiceover, consideration should be made of the narrative strategy. In the “Heavy Metal Umlaut”, Jon Udell establishes the literary and social value of the screencast documentary, previously known for software instructions of use (2005). In the screencast,

Udell deploys a simple narrative strategy that could be employed as a starting point. He opens with an overview of his subject matter, the revision history of the Wikipedia article on the heavy metal umlaut. Through a “quick flight” of the changelog (speeding up the chronological loading of the pages), he shows the growth and occasional vandalism of the article, speaking with awe about Wikipedians’ vigilance (see Figure 3). Subsequently, he introduces four themes, and treats them one by one. The spinal tap theme concerns the typographical as well as factual question of the n-umlaut (or heavy metal umlaut). In the vandalism piece, he is impressed by the dedication shown by the Wikipedians, cleaning the graffiti and reverting other offensive edits only minutes after they have been made. He spends time talking about the organisation of the article, and how the table of contents matures over time. (The focus on the changes to the table of contents led colleagues and I to build a tool, the Wiki TOC scraper, that captures a Wikipedia article’s table of contents, and, with the use of the slider, shows its changes over time.) Finally, Udell mentions issues of cultural sensitivity, and in particular how the look of the font and the n-umlaut is no longer associated with Nazism (as it was initially), but rather is described as Germanic. Without summarising the four themes, Udell concludes the screencast documentary by returning to the first edit and jumping to the last, making mention of the achievement of a “loose federation of volunteers”, in this new type of content creation,

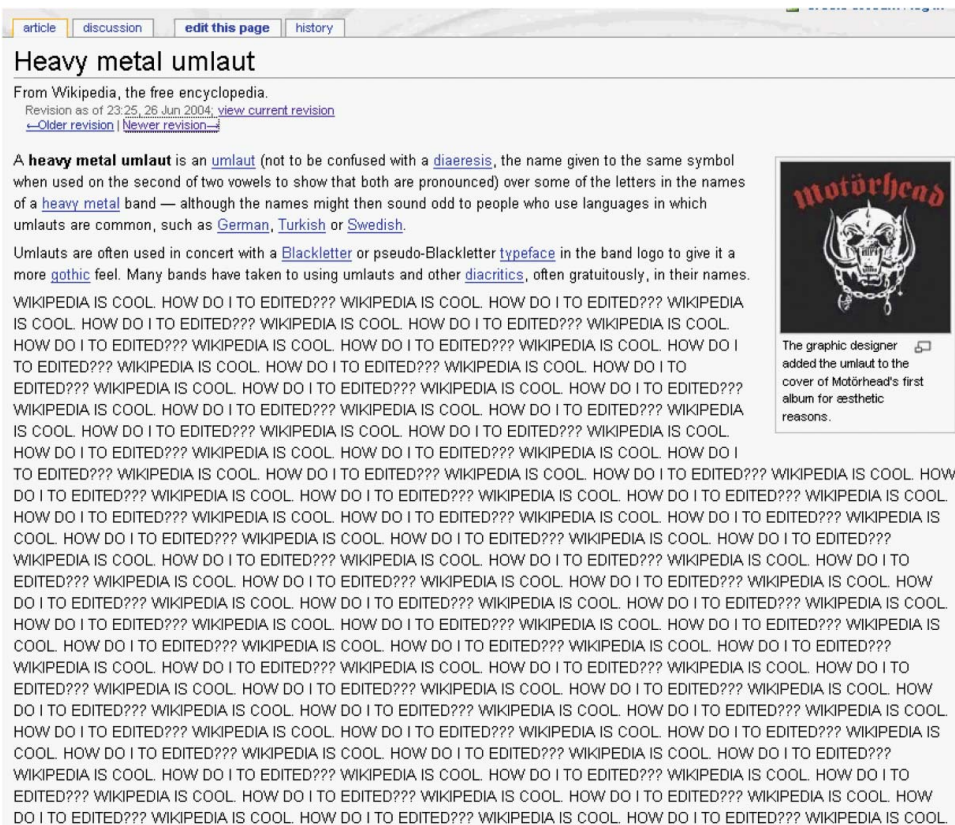


Figure 3. Screenshot from Jon Udell’s “Heavy Metal Umlaut”, screencast documentary (2005), discussion of graffiti.

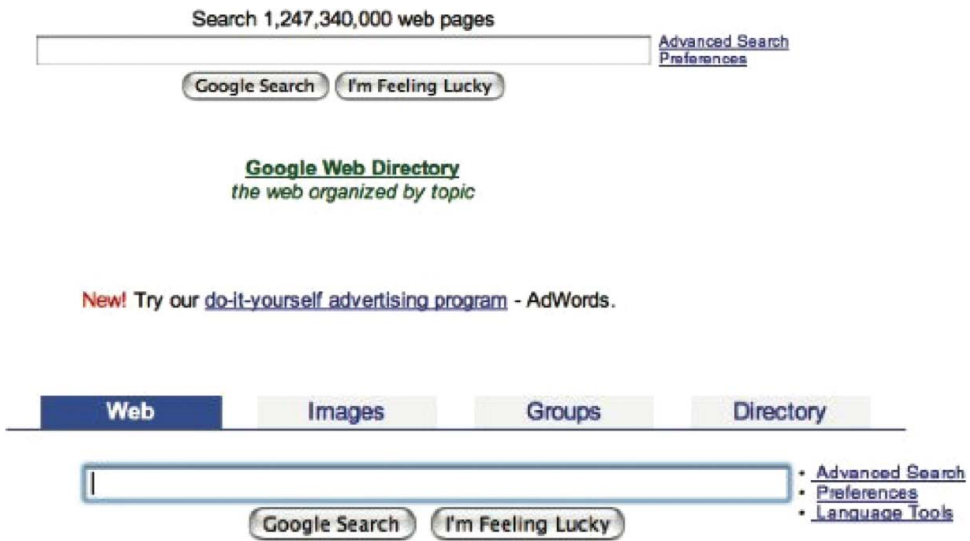


Figure 4. Google's Directory on Google's front page in 2000 (top), and receiving tab status in October of 2001 (bottom). Excerpt from Digital Methods Initiative and Kim de Groot, "The Demise of the Directory: Web librarian work removed in Google", Information Graphic, 2008, http://www.govcom.org/publications/drafts/GCO_directoryfall.pdf.

otherwise known as the wisdom of the crowd. In the edit revision history of a single Wikipedia article, it's as if Web history was made. The screencast thus captures the birth of user-generated content.

"Google and the Politics of Tabs" is the first single-site history made that follows in Udell's footsteps. It is the history of Google seen through its interface from 1998 until late 2007, and through it tells a larger story about the history of the Web (see Figure 4) (Rogers & Govcom.org, 2007). It makes use of all the available, updated Google front-pages in the Internet Archive, captured and played back, in the style of time-lapse photography. Google and the Politics of Tabs chronicles the subtle changes to the Google front-page real estate, showing the services that have risen to the interface, achieving tab status, and the others that have been relegated to the "more" and "even more" buttons. As its main theme, it tells the story of the demise of the directory (particularly, dmoz.org's), and how the back-end algorithm has taken over the organisation of Web information at the expense of the human editors and the librarians.

Conclusions: the value of capturing website histories

Website histories tell stories of the Web, media and cultural or political history. In terms of the stories to be told in the voiceover narrative of a web page history, one could be of loss; something of value has been taken, or replaced. In Google and the Politics of Tabs, which details a decade's worth of subtle changes to the Google.com's interface, ultimately, the algorithm has taken over from the librarian on the Web. Another is about transformation, or even continuity. Despite massive change around it, the object or subject has remained remarkably the same (or nearly so). Despite transformation, it has returned to its original form. As discussed above, the enthusiastic embrace of new media

or its stubborn resistance is made the subject of the screencast by scrutinising how a newspaper, radio station, or television channel website has evolved. Has the old media form, so-called, radically embraced cyberspace and new media features, only to settle back (largely) into a digitised form of its original self, as in the case of *nytimes.com*? How have newspapers domesticated the blog, or tamed the comment space where readers once could talk back to the institution referred to historically as gatekeepers? Here, the story concerns incorporating new media into established practices. In each case, one is considering the overall narrative of change, concentrating on a limited number of story-lines, and leaving out the rest. The third strategy is to allow the history of an idea, individual, organisation, institution or other entity to unfold in the changes to a website. The wedding as institution could be simple, or it can be industrialised, as a website, *theknot.com*, and the Web is further monetised with the rise of e-commerce. One can thus build in the recipe of a great novel. Capture the life and times through the changes occurring to an institution – on its leading website.

On 20 January 2017, with the incoming Presidential administration, *whitehouse.gov* changed dramatically. A story in the *New York Times* opened: “Within moments of the inauguration of President Trump, the official White House website on Friday deleted nearly all mentions of climate change [...] part of the full digital turnover of *whitehouse.gov*, including taking down and archiving all the Obama administration’s personal and policy pages” (Davenport, 2017). Capturing “transitions” such as the Papal in 2005 by the Library of Congress is an event-based Web historiography, pioneered in the websphere technique that curates a collection of thematically related and interlinked sites over a period of some months. One also may capture such transitions through website histories, where changed front-pages are made into screenshots (or otherwise captured) and played back as a screencast documentary or even as an animated gif. Here, the display of content removal tells the story of changes in political (and policy) priorities. One may also focus on additional sections or pages on the website, such as the changes under the “issues” tab, where after 20 January 2017, *whitehouse.gov* had such issues as “America First Energy Plan” and “America First Foreign Policy”, which are distinctive in (sloganeering) style and substance to those on 19 January 2017 prior to the administration turnover (see [Table 1](#)).

Table 1. Top issues at *whitehouse.gov*. Source: WayBack Machine of the Internet Archive ([archive.org](#)).

<i>19 January 2017</i>
Civil rights
Climate change
Economy
Education
Foreign policy
Health care
Iran deal
Immigration action
<i>20 January 2017</i>
America first energy plan
America first foreign policy
Bringing back jobs and growth
Making our military strong again
Standing up for our law enforcement community
Trade deals that work for all Americans

The Internet Archive (and Web archives generally) are commonly thought of as sources for “digital history”, however much actual historian use of Web archives appears to be limited (and is understudied). With such use, digital source criticism becomes a focal point with concerns about how in the archiving a “digitally reborn” source may be reconstituted in a form that never existed in the first place. Here is a particular case where digital history may draw from Web history, and its study of different forms of ephemerality (Chun, 2013). Indeed, Web archives have not necessarily been justified for the purposes of telling Internet and Web history, however much active use may be made of them by researchers in that field. Above, I reintroduced the notion of “website history” and put forward a particular approach to it (screencast documentary) that allows one to pursue a variety of histories: Web, media as well as digital history.

The screencast documentary approach derives from digital methods, or the use of tool-based methods for Web data extraction and analysis. The research affordances of the WayBack Machine are the point of departure, for it provides a list of stored pages (and an indication of which ones have new content, also known as the “diffs”) that can be captured, and played back in the style of time-lapse photography. The website history, it is argued, could be seen in the Web historiological tradition of website biography, which is distinctive from event, national or autobiographical styles of collection and curation. Once captured, the website history may be narrated; in the examples given, the stories revolved around loss, continuity and transformation. They concern how the history of a single website may encapsulate the history of the Web, how so-called old media perpetuates itself in the new media, and how the transformation of an institution may be captured.

Notes

1. Until the burden was lessened in the late 1960s, when photographs were introduced to the courts of law, the fact finder was often asked to produce the negative together with the photograph so it could be authenticated and there was assurance it had not been doctored (Eltgroth, 2009).
2. See also docnow.io, which is a tool and a community developed around supporting the ethical collection, use, and preservation of social media content.
3. <https://tools.digitalmethods.net/beta/internetArchiveWaybackMachineLinkRipper/>.
4. There is also an option to download the html of the archived pages.

Acknowledgements

The author would like to acknowledge the work of Erik Borra, Anne Helmond, Michael Stevenson and Esther Weltevrede on the Digital Methods unit dedicated to ‘The Website as Archived Object,’ from which this article draws.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by the European Union’s Horizon 2020 research and innovation programme [grant number 732942].

Notes on contributor

Richard Rogers is a professor of New Media & Digital Culture, Media Studies, University of Amsterdam.

References

- Abbate, J. (2000). *Inventing the internet*. Cambridge, MA: MIT Press.
- Ben-David, A. (2016). What does the Web Remember of its Deleted Past? An archival reconstruction of the former Yugoslav top level domain. *New Media & Society*, 18(7), 1103–1119.
- Bødker, H., & Brügger, N. (2017). The shifting temporalities of online news: The Guardian's website from 1996 to 2015. *Journalism*. doi:10.1177/14648849166891.
- Brock, A. (2005). 'A belief in humanity is a belief in colored men': Using culture to span the digital divide. *Journal of Computer-Mediated Communication*, 11(1), article 17.
- Brügger, N. (2008). The archived website and website philology: A new type of historical document? *Nordicom Review*, 29(2), 151–171.
- 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.
- Chun, W. (2013). *Programmed visions: Software and memory*. Cambridge, MA: MIT Press.
- Cohen, D. J., & Rosenzweig, R. (2006). *Digital history: A guide to gathering, preserving, and presenting the past on the web*. Philadelphia: University of Pennsylvania Press.
- Davenport, C. (2017, January 20). With Trump in charge, climate change references purged from website. *New York Times*.
- Dean, J. (1998). *Aliens in America: Conspiracy cultures from outerspace to cyberspace*. Ithaca, NY: Cornell University Press.
- Dougherty, M., & Meyer, E. T. (2014). Community, tools, and practices in web archiving: The state-of-the-art in relation to social science and humanities research needs. *Journal of the Association for Information Science and Technology*, 65(11), 2195–2209.
- Dougherty, M., Meyer, E. T., Madsen, C., van den Heuvel, C., Thomas, A., & Wyatt, S. (2010). *Researcher engagement with web archives: State of the art*. London: JISC.
- Eltgroth, D. (2009). Best evidence and the Wayback Machine: Toward a workable authentication standard for archived Internet evidence. *Fordham Law Review*, 78(1), 181.
- Gazaryan, K. (2013). Authenticity of archived websites: The need to lower the evidentiary hurdle is imminent. *Rutgers Computer and Technology Law Journal*, 39(2), 216–245.
- Gibson, A. (2016, November 2). URLs: Some practical advice, MLA style center. New York, NY: Modern Language Association. Retrieved from <https://style.mla.org/2016/11/02/urls-some-practical-advice/>
- Goggin, G., & McLelland, M. (2017). Introduction: Global Coordinates of Internet histories. In G. Goggin & M. McLelland (Eds.), *The Routledge companion to global internet histories*. New York, NY: Routledge.
- Hermens, E. (2011). *The New York Times - a web historiography, screencast documentary*. Amsterdam: University of Amsterdam. Retrieved from <https://vimeo.com/32319207>
- Hockx-Yu, H. (2014). Access and scholarly use of web archives. *Alexandria*, 25(1), 113–127.
- Howell, B. A. (2006). Proving web history: How to use the Internet Archive. *Journal of Internet Law*, 9(8), 3–9.
- Internet Archive. (2016). FAQ, San Francisco. Internet Archive. Retrieved from <https://archive.org/about/faqs.php#265>
- Internet Archive. (2017). Internet Archive. Retrieved from <http://www.archive.org>
- Jacobsen, G. (2008). Web archiving: Issues and problems in collection building and access. *LIBER Quarterly*, 18(3–4), 366–376.
- Klein, M., van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., & Zhou, K. (2014). Scholarly context not found: One in five articles suffers from reference rot. *PLoS ONE*, 9(12), e115253.

- Livio, M., Mataly, J., & Schuh, M. (2012). *TheKnot.com - a website historiography, screencast documentary*. Amsterdam: University of Amsterdam. Retrieved from <https://www.youtube.com/watch?v=5cxVXJthETA>
- Milligan, I. (2016). Lost in the infinite archive: The promise and pitfalls of web archives. *International Journal of Humanities and Arts Computing*, 10(1), 78–94.
- Murphy, J., Hashim, N. H., & O'Connor, P. (2007). Take me back: Validating the WayBack Machine. *Journal of Computer-Mediated Communication*, 13(1).
- Rhizome. (2014). *Amalia Ulman: Excellences & perfections*. New York, NY: Rhizome. Retrieved from <http://webenact.rhizome.org/excellences-and-perfections>
- Rogers, R. (2013). *Digital methods*. Cambridge, MA: MIT Press.
- Rogers, R. & Govcom.org. (2008). *Google and the politics of tabs, screencast documentary*. Amsterdam: Govcom.org. Retrieved from <https://movies.digitalmethods.net/google.html>
- Rogers, R., Weltevrede, E., Borra, E., & Niederer, S. (2013). National Web studies: The case of Iran online. In J. Hartley, J. Burgess, & A. Bruns (Eds.), *A companion to new media dynamics*. Oxford: Wiley-Blackwell.
- Rosenzweig, R. (2003). Scarcity or abundance? Preserving the past in a digital era. *The American Historical Review*, 108(3), 735–762.
- Russell, E., & Kane, J. (2008). The missing link: Assessing the reliability of Internet citations in history journals. *Technology and Culture*, 49(2), 420–429.
- Ryan, J. (2011). *A history of the internet and the digital future*. London: Reaktion.
- Schneider, S., & Foot, K. (2004). The Web as an object of study. *New Media & Society*, 6(1), 114–122.
- Stevenson, M. (2016). Rethinking the participatory web: A history of HotWired's 'new publishing paradigm,' 1994–1997. *New Media & Society*, 18(7), 1331–1346.
- Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. *Annual Review of Information Science and Technology*, 39(1), 81–135.
- Udell, J. (2005). Heavy metal Umlaut, screencast documentary. Retrieved from <http://jonudell.net/udell/2005-01-22-heavy-metal-umlaut-the-movie.html>
- Veronin, M. A. (2002). Where are they now? A case study of health-related web site attrition. *Journal of Medical Internet Research*, 4(2), e10.
- WayBack Machine. (2017). WayBack Machine of the Internet Archive. Retrieved from <http://www.archive.org/web>.
- Weltevrede, E., & Helmond, A. (2012). Where do bloggers blog? Platform transitions within the historical Dutch blogosphere. *First Monday*, 17(2).