

The role of Internet Wayback Machine in a multi-method research project

Elisabetta Locatelli, adjunct professor at Università Cattolica del Sacro Cuore (Italy) and researcher at OssCom (Research Center on Media and Communication of Università Cattolica del Sacro Cuore).

elisabetta.locatelli@unicatt.it

Abstract

If, on the one side, the web offers us a platform where content is searchable and replicable, on the other one, it cannot be forgotten that web content is perishable, unstable and subject to continuous change. This is a challenge for scholarly research about the historical development of web. The research here presented analyzed the historical development of weblogs in Italy investigating their technological, cultural, economic, and institutional dimensions. The approach chosen mixed participant observation, in-depth interviews, and semiotic analysis of blogs and blog posts. Since an important part of the research was about the development of platforms, graphics, layouts, and technology, beside interviews older versions of blogs were retrieved using Internet Wayback Machine. Even if partial versions of the blogs were archived, this part of the research was important to complete data obtained with interviews and blogs' analysis, since individual memory is not always accurate or some blogs were in the meanwhile closed and original posts were not accessible anymore.

Web contents: stable or ephemeral?¹

In this paper we will discuss the use of [Internet Wayback Machine](#) (part of the [Internet Archive](#) project) in a research project into the development of Italian blogosphere.

According to Oxford English Dictionary a blog is “A regularly updated website or web page, typically one run by an individual or small group, that is written in an informal or conversational style”.

Blogs are part of the broader realm of Social Media that are “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content” (Kaplan & Haenlein, 2010, p. 61). Although this is not the place where such a complex and interesting definition can be discussed, it can be highlighted that the term Social Media comprehends, beside other platforms, social networks (like Facebook and Twitter), collaborative platforms (like Wikipedia), content communities (like YouTube), and blogs (Kaplan & Haenlein, 2010; see also Van Dijck, 2013). As Kaplan and Haenlein (2010) pinpoint, blogs “represent the earliest form of Social Media” and “are special types of websites that usually display date-stamped entries in reverse chronological order (OECD, 2007)” (p. 63).

Back in 2004 Schneider and Foot highlighted the ambivalence of web contents, writing that the methodological challenges of the study of the web “stem from the nature of the Web, which is a unique mixture of the ephemeral and the permanent” (p. 115).

About the permanence and stability of web contents, research into “networked publics” (boyd, 2011) assessed that their affordances are

Persistence: Online expressions are automatically recorded and archived.

Replicability: Content made out of bits can be duplicated.

Scalability: The potential visibility of contents in networked publics is great.

Searchability: Content in networked publics can be accessed through search (boyd, 2011, p. 46).

Networked publics are

¹ The research was part of the PhD thesis of the author, discussed on May 22, 2008 at the Università Cattolica del Sacro Cuore (Italy) and later published in Locatelli (2014). The paper is a reworked version of the chapter Blogs as cultural products: A multidimensional approach to their diffusion in Italy (2001–2008) forthcoming in Web 25: Histories from the first 25 Years of the World Wide Web edited by Niels Brügger.

publics that are restructured by networked technologies. As such, they are simultaneously (1) the space constructed through networked technologies and (2) the imagined collective that emerges as a result of the intersection of people, technology, and practice. [...] While networked publics share much in common with other types of public, the ways in which technology structures them introduces distinct affordances that shape how people engage with these environments (p. 2).

Although boyd refers mainly to social networks, her discourse can be extended to Social Media (Kaplan & Haenlein, 2010) too, including thus blogs. For our discourse, it is particularly interesting the focus on the interaction between society (publics) and technology (social networks).

If contents are persistent, replicable, scalable, and searchable, this means that they remain online after they have been published and can be accessed by the user itself, other users, but also by researchers that can also aggregate them under the form of Big Data.

On the other side, web content is perishable, unstable and subject to continuous change.

As Schneider and Foot (2004) pinpoint,

Web content is ephemeral in its transience, as it can be expected to last for only a relatively brief time. [...] Second, Web content is ephemeral in its construction – like television, radio, theater and other “performance media” (Hecht et al., 1993, Stowkowski, 2002). Web content, once presented, needs to be reconstructed or represented in order for others to experience it”. (p. 115).

Although, unlike other performance media, the web

must exist in a permanent form to be transmitted. [...] However, the performance of the Web is somewhat fleeting. Unlike any other permanent media, a website may destroy its predecessor regularly and procedurally each time it is updated by its producer; that is, absent specific arrangements to the contrary, each previous edition of a website may be erased as a new version is produced (Schneider and Foot, p. 115).

Blogs, both in their structure and posts, embody entirely this duality. Blogs, actually, let the users to produce and upload online contents - such as texts, pictures, and videos – autonomously (Blood, 2002; Kaplan & Haenlein 2010).

On the one hand blog contents are persistent. Published contents are archived on the blogging platform, are accessible through search engines and can also be archived in other blogs/websites under the form of links or screenshots, for example. Once removed, a content is no more visible nor searchable on the blog, but it may have been archived otherwise (cache, web archive, screenshot). These contents are indexed by search engines and are available for long time after their production. A problematic side of this availability is related to the right to be forgotten (Guimarães Pereira, Ghezzi, Vesnić-Alujević, 2014) and to the difficulty to remove fake, uncorrect or old contents from the web.

On the other hand, they are ephemeral. Once updated, the new layout replaces the old one that is no more accessible loading the blog and the archives take the new layout too. Edited posts do not keep trace of the changes, unless the blogger desire to keep them, for example under the form of strikethrough text or signaling explicitly that the post was updated.

This duality is particularly evident when the researcher is working on a research project that regards the historical development of blogs across a selected interval.

Building a social history of blogs in Italy

The research here presented worked about the historical development of weblogs in Italy analyzing their technological, cultural, economic, and institutional dimensions (Griswold 1994; Colombo 2003, 2006; Van Dijck 2013).

Following Hine (2000), the internet can be intended both as culture and as a cultural artefact. Considering the internet as a culture means to conceive it as a “place, cyberspace, where culture is formed and reformed” (Hine, 2000, p. 9), whereas, considering it as a cultural artefact means seeing it as a “product of culture”—a technology that is used by different social groups, “with contextually situated goals and priorities. It is also a technology which is shaped by the ways in which it is marketed, taught and used” (Hine, 2000, p. 9).

The same dynamics can be applied to blogs. On the one hand, they have developed a new culture and altered the present one; on the other hand, they were (and are) socially and culturally shaped by the context their part of.

An especially powerful theoretical framework for understanding this duality is to interpret blogs as a cultural product that is “a shared significance embodied in a form” (Griswold, 1986, 1994). The relationship between culture and society is conceived as governed by four dimensions (social world, receiver, cultural object, creator) arranged like a baseball diamond (Griswold, 1994, p. 15).

Colombo (2001, 2003) expands semantically these dimensions into socio-cultural environment, consumption, cultural product, and production. Moreover, defining a medium per se Colombo introduces four additional dimensions —namely the technological, economic, institutional and cultural (Colombo, 2006, p. 30) that in the research here presented were used to analyse the socio-cultural environment and the production of blogs.

In the frame of cultural products like blogs, technology enables the existence of the objects but, at the same time, it is also shaped by user’s habits and practices (Siles, 2012a). The economic dimension refers to the resources needed to access a service or buy a cultural product, as well as to the business and marketing strategies apt to promoting it. The institutional dimension includes all the entities that influence the cultural products: the law, rules, economic subjects, and regulatory bodies. Last but not least, is the cultural dimension. According to Hannerz culture is composed of “overt forms” interpreted by human minds through “externalized meanings” (Hannerz, 1992, p. 4). Blogs, then, are parts of a culture but they also are externalised means that reflect and produce culture.

Methodology

According to the theoretical framework proposed, the methodology should consider the subjective way to live blogging experience, the use of online platforms and blogs' evolution. It is to say, technological, cultural, and social aspects should be analysed together.

Thus, the choice was to build a multi-sited research project (Marcus, 1995) where the main reference to sociology was integrated with other perspectives (Alasuutari, 1999) with a continuous comparison between objectives and techniques, as recommended by Grounded Theory (Glaser, Strauss, 1967; Strauss, Corbin, 1990; Charmaz, 2000).

The qualitative methodology comprised three steps. First, an explorative analysis of the Italian blogosphere and of the early development of blogging in the US was done examining blogs, journal articles, and academic and non-academic publications. Also, some of the early US bloggers were contacted via e-mail. Afterwards, a participant observation of three blogger meetings in Italy was done with short individual interviews. We, then, chose to concentrate on personal blogs. Three phases in the development of personal blogs were identified: early bloggers (1999/2000–2001); the age of Splinder (2003-2006); mainstreaming (2006-2008).

The second methodological step comprised 25 semi-structured interviews (done in 2007) aimed at investigating the individual repertoire of routines and practices involved in maintaining the blog.

The third phase consisted in the semiotic analysis (Cosenza, 2008) of 50 blogs (the interviewees’ plus 25 other blogs sampled following the same criteria) doing an in-depth analysis of blog’s structure and of a selected number of posts.

The sample of interviewees and blogs was selected through theoretical (Mason, 1996; Silverman, 2000) and snowball sampling (Bailey, 1987). The following criteria were considered: blogs’ opening date; city of residence (Milan, Rome and Turin, chosen for the significance of their blogospheres); gender (14 males and 11 females); and the kind of content posted (text, photo, and video).

Using Internet Wayback Machine and mixing methods

An important part of the research was into the development of platforms (such as graphics, layouts, devices) with the goal to understand the very beginning of the blogging activity and its evolution.

This objective was reached one the one side during interviews, eliciting memories and feelings related to the beginning of the blogging activity. On the other side, it was reached by analyzing blogs with a semiotic framework. This analysis regarded the whole blog (structure, graphics, sections, self-description,

links, etc.) and single posts. The first four and the latest four months of blogging activity were analyzed (plus four months in the middle for the first group). Knowing the original aspect of the blog was relevant in order to understand its evolution and also to do a cross-comparison between the blogs of the sample.

The way which proved to be viable to reach this goal was to use the [Internet Wayback Machine](#) as a complementary tool for tracing the historical development of blogs, verifying what emerged during interviews and confronting different stages of the blogs analyzed in order to find similarities and differences. Moreover, some bloggers kept screenshots of the earlier versions of their blogs and shared them with the researcher.

Even if partial versions of the blogs were archived, this part of the research was important to complete data obtained with interviews and blogs' analysis, since individual memory is not always accurate or some blogs were in the meanwhile closed and original posts were not accessible anymore.

Internet Wayback Machine is a free online tool provided by Internet Archive that provides "two complementary measures of website evolution: website age and website updates" (Murphy, Hashim & O'Connor 2008, p. 61). Murphy, Hashim & O'Connor (2008) validated the Internet Wayback Machine as a "viable research tool" (p. 71), highlighting though that the Wayback Machine

has limitations. It archives publicly accessible sites written in simple HTML, but has problems archiving password-protected or dynamic websites (Veronin, 2002). Furthermore, sites can decline inclusion by emailing the Internet Archive or using the Standard for Robot Exclusion (see www.robotstxt.org) to specify files or directories not to crawl (FAQs, 2007). Intellectual property owners concerned about infringements of third party sites can also request removal of such content (FAQs, 2007). Any of these actions stops future indexing, removes site content from the archive, and limits the archives' comprehensiveness. (Murphy, Hashim & O'Connor, 2008, p. 64).

Current Internet Wayback Machine FAQs confirm this limitations that are worth to be considered while using it.

Conclusions

The research evidenced that there was and still is a mutual shaping between technology, culture, society, economy, and institutions as it happened in the US (Siles, 2012a). It also showed that in Italy, despite the early and broad diffusion of blogging, the institutional element has been struggling to find the right way to interpret and regulate the new phenomenon.

Among the several factors that contributed to the shaping of the phenomenon, the research showed that the social and technological context had a strong influence in differentiating blogging 'appropriation' (Silverstone & Hirsch, 1992) among the three highlighted phases. On the contrary, the individual process of adoption (time, space, and privacy management)—were similar across the three phases. This result was obtained confronting materials coming from different sources and analysing different stages of the blogging activity. The continuous comparison of data was inspired both by a hermeneutic approach and by Grounded Theory (Glaser, Strauss, 1967; Strauss, Corbin, 1990; Charmaz, 2000).

Interviews let to understand the subjective meaning of blogging and the individual approach to the phenomenon, meanwhile blogs' analysis let to understand deeply the artefactual nature of blogs under two points of view: on the one side, the individual actions that shaped blogs and, on the other side, how web culture and technological evolution affected blog's appearance and practices of use. For this part, the use of Internet Wayback Machine was useful to obtain data that otherwise could not be reachable.

In conclusion, given the complex theoretical framework of the research that considers technological, cultural/social, institutional, and economic aspects of blogging, a methodological approach that mixed different methodologies proved to be effective to manage the permanent and ephemeral nature of blog contents and to reach the goal to build a history of blogging development in Italy.

Bibliography

Alasuutari, P. (1999). Cultural images of the media. In P. Alasuutari (Ed.), *Rethinking the media audience: The new agenda* (pp. 95–104). London: Sage.

- Bailey, K. (1987). *Methods of social research*. New York, NY: Free Press.
- Blood, R. (2002). *The weblog handbook: Practical advice on creating and maintaining your blog*. Cambridge, MA: Perseus Pub.
- Charmaz, K. (2000). Grounded theory. Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509–535). London: Sage.
- Colombo, F. (2001). L'industria culturale e i suoi margini. In F. Colombo, L. Farinotti, & F. Pasquali (Eds.), *I margini della cultura. Media e innovazione* (pp. 43–68). Milano: Franco Angeli.
- Colombo, F. (2003). *Introduzione allo studio dei media*. Roma: Carocci.
- Colombo, F. (2006). Technological innovation and media complexity. DTT in the light of a new theoretical prospect. In F. Colombo & N. Vittadini (Eds.), *Digitizing TV: Theoretical issues and comparative studies across Europe* (pp. 21–35). Milano: Vita & Pensiero.
- Cosenza, G. (2008). *Semiotica dei nuovi media* (2nd ed.). Roma-Bari: Laterza.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.
- Griswold, W. (1986). *Renaissance rivals: City comedy and revenge tragedy in the London theatre 1576–1980*. Chicago, IL: University of Chicago Press.
- Griswold, W. (1994). *Cultures and societies in a changing world*. Thousand Oaks, CA: Pine Forge Press.
- Guimarães Hannerz, U. (1992). *Cultural complexity: Studies in the social organization of meanings*. New York, NY: Columbia University Press.
- Hannerz, U. (1992). *Cultural complexity: Studies in the social organization of meanings*. New York, NY: Columbia University Press.
- Hine, C. (2000). *Virtual ethnography*. London: Sage Publications.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59–68. doi:10.1016/j.bushor.2009.09.003
- Locatelli, E. (2014). *The blog up: Storia sociale del blog in Italia*. Milan: Franco Angeli.
- Marcus, G. E. (1995). Ethnography in/of the world system: The emergence of multisided ethnography. *Annual Review of Anthropology*, 24, 95–117.
- Mason, J. (1996). *Qualitative researching*. London: Sage.
- Murphy, J., Hashim, N.H., O'Connor, P. (2008). Take Me Back: Validating the Wayback Machine. *Journal of Computer-Mediated Communication*, 13, 60-75.
- boyd, d., (2011). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In Z. Papacharissi (ed.). *A Networked self: Identity, community and culture on social network sites*. New York: Routledge, pp. 39-58.
- Pereira, A., Ghezzi, A., & Vesnić-Alujević, L. (2014). *The ethics of memory in a digital age: Interrogating the right to be forgotten*. Houndmills, Basingstoke: Palgrave Macmillan.
- Schneider, S.M., Foot, K.A. (2004). The web as an object of study. *New Media & Society*, 6(1), 114-112.
- Siles, I. (2012a). The rise of blogging: Articulation as a dynamic of technological stabilization. *New Media & Society*, 14(5), 781–797.
- Silverman, D. (2000). *Doing qualitative research: A practical handbook*. London: Sage.
- Silverstone, R., & Hirsch, E. (1992). *Consuming technologies: Media and information in domestic spaces*. Abingdon: Routledge.
- Strauss, A. L., & Corbin J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Van Dijck, J. (2015). *The culture of connectivity: A critical history of social media*. Johanneshov: MTM.

The Challenges of 25 Years of Data: An Agenda for Web-based Research

Matthew S. Weber

Rutgers University

Despite the numerous challenges that exist, recent years have seen a substantial growth in rigorous and important Web-based research (Brügger & Finnemann, 2013; Milligan, 2016).. For instance, research tracing the development of the Yahoo! GeoCities provides an important snapshot of social organizing in the 1990s (Milligan, 2016). Websites from political campaigns provide critical records for historians and political scientists who wish to examine the discourse and narrative of political events (Davis, 1998; Margolis, Resnick, & Tu, 1997). Moreover, as a source for research, the Web provides for more than historical scholarship. A 2012 study demonstrated that the aggregation of users on Facebook.com could provide social scientists with fertile ground for experiments; Facebook researchers used experimental conditions to test whether or not social influence could impact the likelihood that connected peers would participate in elections (Bond et al., 2012). Arguably, a list of potential research using the Web¹ could parallel the number of Web pages that are currently in operation. And yet, in many regards research utilizing Web data is still limited in a number of ways.

This short article takes a forward-looking perspective to consider key research problems that have challenged researchers in the past, and will likely continue to challenge researchers in the future. The focus herein is on the use Web data as an object of study, although some consideration is given to the use of tools to collect Web data. With that perspective in mind, this short article highlights three specific research challenges that face researchers working with Web data today. First, the scope of the Web poses questions with regards to the size and time dimensions of research. Second, the nature of Web data points to questions regarding the reliability and validity of Web data. Third, the type of data on the Web poses significant questions with regards to the ethics research using Web data.

Questions of Size and Time

Scholars who sets out to conduct research using data from the Web – regardless of field or domain - should address questions regarding size and time dimensions pertaining to their data. Whether these issues are addressed at the outset, or later in data collection and analysis, they will rear their ugly heads inevitably. Questions regarding the data size have implications for all phases of research, as has clearly been demonstrated in prior research. At the most basic level, researchers must address how much data is needed. For instance, 1 million records may sound like a significant amount data, but it is not quite as much if the records are simply lists of uniform resource locators (URLs). On the other hand, 1 million archived Web pages would require a significant amount of sources in order to store, access and analyze. Thus, it is important to ascertain the type of data that is needed, and the fields (e.g. URLs, images, text, video) that are required.

Computationally based historical studies have allowed historians to explore historical records in order to better understand patterns that existed in past generations. Much of this type of analysis utilizes Web data but focuses on small-scale analyses. For instance, researchers have used archived Web data for everything from exploring labor movements in Australia prior to World War I (Mark, 2015) to examining the nature of Web memorials in the wake of the September 11, 2001 terrorist attacks in New York City (Foot, Warnick, & Schneider, 2005) to an examination of Danish public broadcasting (Brügger, 2012). Mark drew upon a small Web archive of digitized songs written by a notable member of the Industrial Workers of the World movement in Australia at the beginning of the 20th Century. It is important to remember that not all Web data originated online; some of the most valuable resources are repositories of material be preserved for future generations. In a second example, Foot, Warnick and Schneider sought to examine the social practice of memorialization as realized

¹ This short article focuses broadly on data drawn from the World Wide Web. Increasingly, this type of data increasingly intersects with Internet data (see O'Reilly 2017 for a relevant discussion of the differences between the two domains).

through digital objects. The authors focused their research on the 9/11 attacks in New York City, but in order to select an appropriate sample they first began with a set of 60 websites, which were then reduced to a subset of 8 based on key variables relevant to the research. In this case, Foot and colleagues wanted websites that were coproduced by a number of individuals, and were sites that focused specifically on memorialization. Thus, although the focus was on native digital content, and there was a relatively large pool of data to draw upon, the authors had specific filters to narrow the scope.

The time dimension is often difficult to disentangle from the size issue; for instance, Brügger's examination was restricted to a single website, but focused on an 11 year period. On the other hand, Foot, Warnick and Schneider looked at eight websites, but focused on a narrow timeframe in 2001. When traversing time, one key variable that should be considered is the frequency of sampling. A more granular sampling window will generate large datasets. If a fine level of granularity is not required, one can capture a wide swath of time with a relatively small dataset (e.g. weekly snapshots vs. hourly snapshots).

Many readers will be more familiar with the Big Data aspect of Web-based research. As the examples from humanities and social science research illustrate, when dealing with Web-based data it is certainly not a requirement to manage millions of data points. On the other hand, large-scale datasets offer a unique opportunity to examine social phenomena at a scale that previously was not possible. Twitter data have been used to understand and predict voting trends in political elections (McKelvey, DiGrazia, & Rojas, 2014), and to predict key topics of discussion in politics (Rill, Reinel, Scheidt, & Zicari, 2014). In these examples, the scale was necessary to enhance the predictive power and validity of the data. From a discourse perspective, political big data has been used to understand how individuals talk about politics, and how different political spheres develop (Colleoni, Rozza, & Arvidsson, 2014). Historical data sources further enabled researchers to examine shifts in political discourse over longer periods of time (Zeng & Greenfield, 2015). In each of these examples, the size of the dataset was notably large, but the decision to utilize data at such a scale was made as a result of the research questions that were asked. These examples underscore the need to let the research question guide the choice of data selection.

Clearly, significant advances in computing technology have provided researchers with improved access to large-scale data sources. Pair these advances with the decreasing cost of high-performance computing, and there is great potential in the future for continued advancements with regards to large-scale research using Web data sources. On the other hand, when thinking about the size and scale of Web-based data, the above discussion underscores the importance of considering the nature of the research questions being asked, as well as the type of analyses being conducted. Moreover, one research question can often be answered in multiple ways; although there is much to be said for the appeal of large datasets, there is often much to be gained from exploring the nuance of smaller datasets. Finally, the aggregate of various scales of research provides a future opportunity for the triangulation of findings to further enhance of understanding of the social world.

Questions of Reliability and Validity

To some degree, any good research study inevitably addresses questions of reliability and validity. When working with Web based data, however, these questions are amplified due to the nature of the data. In some cases, the challenges aren't as complicated. For instance, The New York Times made its entire archive from 1851 to present available and searchable via its Web portal (<http://query.nytimes.com/search/sitesearch/#/>). Similarly, the Digital Preservation unit of the Library of Congress maintains a robust directory of curated digital archive content (<http://www.digitalpreservation.gov/collections/>); the majority of the content includes specific information about the nature of the data and how it was collected.

In the cases above, there are clearly defined parameters that provide information regarding the nature of the aggregated data. In general, however, Web data is plagued by questions of reliability and validity in part

because the data vary so widely. In general, reliability of a study refers to the degree to which that study can be replicated in repeated trials (Wheaton, Muthen, Alwin, & Summers, 1977). On the other hand, validity refers to the degree to which a study measures a selected phenomenon (Golafshani, 2003). These general concepts can similarly be applied to the context of Web data. Reliability, when dealing with Web data, implies that a collection of Web data consistently captures the same data. On the other hand, validity implies that Web data consistently captures data related to the same topic. The underlying concepts are the same, but the ambiguity of the Web domain and the constant flux of both content and technology, creates unique challenges. Thus, reliability and validity questions can emerge in a number of different aspects when working with Web data. In particular, it is often difficult to know how complete Web data are; thus, it is hard to accurately assess the validity of a given dataset. Moreover, the fact that Web data are often in flux means that a study conducted at one point in time may not be reliably replicated in the future.

A number of research strategies have been suggested in order to address issues of degradation and missing data, but overall this is an ongoing challenge facing Web researchers. Some efforts have focused on developing more effective strategies for archiving Web content. One proposal suggests focusing archival efforts on capturing data that changes the most frequently, in order to capture the majority of new content (Spaniol, Denev, Mazeika, Weikum, & Senellart, 2009). Elsewhere, researchers have suggested crawling strategies should prioritize archival efforts based on the size and relative position of websites within their larger ecosystems (Song, Liu, Wen, & Ma, 2004).

A number of recent studies have taken existing sets of archived Web data and examined the content in order to address the pervasiveness of issues related to reliability and validity. For instance, a sample of Web archive data found significant difference in the national origins of archived data (Thelwall & Vaughan, 2004); for instance, the researchers found in their analysis that websites in the United States were overrepresented, which websites in China were underrepresented. Knowing such issues exist is a first step, but any large-scale analysis of such a domain would ultimately have to correct for such biases.

As Milligan (2016) notes, the first step is for researchers simply to be aware of the limitations that exist in the data they are utilizing. Notably, most Web sources are fragments. The best known sources of archived data is arguably the Internet Archive (archive.org), a non-profit organization founded with the mission of archiving Web content and establishing a new library of digital information. The Internet Archive does not, however, contain social media data. Moreover, there are many gaps, especially when one moves beyond .com commercial domains. National libraries around the world have also created country repositories; the British Library contains a notable example, and the Danish government has established a strong Web archiving program. The end result is a mass of data, but also a fragmented set of disconnected collections.

Challenges of reliability and validity are likely to persist moving forward, and accelerate as the Web becomes increasingly mobile and social. One study of Twitter data sources found that there were significant differences in the results of statistical analyses depending on the source of Twitter data (Driscoll & Walker, 2014). One estimate found that approximately 65% of requested archived pages from user searches on the “live web” pointed to Web content that no longer exist on the live Web (AlNoamany, AlSum, Weigle, & Nelson, 2014). From a research perspective, this type of trend means that researchers examining Web data will increasingly need to rely on some form of archived Web database, but researchers will need to be more aware of the limitations of such datasets.

Questions of Ethics

Questions regarding the size, scale, reliability and validity of data are numerous, but they are not insurmountable. In fact, many of the aforementioned research points to a path forward, be it through statistical correction, enhanced specification of data collection techniques or mapping of datasets, among other

approaches, potential solutions exist and present an opportunity for future research. Significant ethical questions lay lurking below the surface of many of the studies this short article has touched on, as well as for many of the data sources.

Web archiving efforts generally work to respect requests for privacy while balancing the need for preservation; for instance, the Internet Archive crawls all available Web content, but does not make content publicly available where there is a clear request to exclude a page from crawls (e.g. robots.txt files). In another example, the national Danish Web archive similarly does not observe robots.txt exclusions; moreover, the Danish effort will also crawl content that is password protected when it is possible to purchase or obtain a password (e.g. to crawl password-protected newspaper archives). In this way, data are not collected if they are protected in a way that is designed to prevent public access (e.g. corporate intranets are not crawled). Twitter data is another ethical grey area; social media data often contains records of social interaction, including personal comments. But Twitter data that is scraped and collected by researchers is often aggregated without permission. Returning to the example of Bond et al.'s (2012) social experiment via Facebook on 61 million votes, what is notable is that participants were never made aware that they were engaging in a scientific research study. The research team manipulated participants by showing an "I Voted" button in some cases, but not in others. In that case, there was no clear potential that research participants could face any harm, but nevertheless most research oversight boards would recommend making participants aware of their role in research.²

Indeed, a central challenge with regards to ethics of Web based research is that many of the existing mechanisms for oversight are working to catch up with technology. In the United States, for instance, Institutional Review Boards vary dramatically from state-to-state and from university-to-university, and have been slow to develop clear guidelines for dealing with Web-based data sources. In the midst of ongoing policymaking about data use, governments are instituting legal regulations to address the issues of big data (for example, legislation stipulating a 'right to be forgotten' now exists in the European Union) (Greenleaf, 2014). In the case of 'right to be forgotten' legislation, current regulation focuses specifically on search results and the right to be excluded from search engines such as Google. Archives are a murkier area of study; it is less clear what rights publics have to be forgotten in archived Web content.

For scholars today, there is an opportunity to engage in research that sets an agenda for discussion of ethics and data privacy. For example, studies such as Lewis's work at Harvard underscore the identifying nature of social media data. Researchers working with Web-based data can contribute to this discussion by helping scholars to better understand *what* data can contribute to identification and *how* researchers can better deidentify existing datasets. Moreover, the policy researchers should have a voice in helping scholars to better understand fair use and privacy rights.

Unanswered Questions and Future Challenges

The balance of this short article points to numerous challenges that surround Web-based research. Although Web-based research presents a number of challenges it is a wide-open landscape for research across disciplines, from the humanities to the social sciences to computer science.

The future of Web-based research also points to the need for new methods of analysis and inquiry. Existing methodological techniques, such as standard linear regressions, are often quite limited in power when used for large data. In addition, these techniques often overestimate the significance of variables. Nevertheless, there are a number of options for research in this space; for example, some have pointed to penalized linear

² See Schroeder (2014) for a more complete accounting of the ethical issues surrounding this study, as well as details regarding the response in both the academic and popular presses.

regression models as a key method for theory building using big data (Hindman, 2015). Analytically speaking, statistical packages such as STATA and R are able to handle large-scale data; in addition, communities of scholars willing to assist with research challenges often support these tools. That said, there is a learning curve that must be overcome by researchers in order to access and engage with data. Sampling issues must be accounted for, as well as temporal biases resulting from when data were collected (Howison & Wiggins). These issues may further limit the predictive capacity of big data (Dhar, 2013), and must be accounted for regardless of discipline.

Looking beyond purely methodological issues, the type of data available to researchers continues to change. In part, librarians and scholars have focused increasing attention on Web archiving as a critical record of societal interaction and recent modern history. The result has been a redoubled effort on improving Web data collection techniques. Moreover, resources and technology for archiving have increased substantially in the past decade. For instance, in 2016 the Archive-IT Web archiving platform had been utilized by more than 350 institutions to archive more than 9 billion documents in more than 2700 archives. A 2013 survey by the National Digital Stewardship Alliance suggests that, “web archiving programs nationally are both maturing and converging on common sets of practices”. The challenges of modern Web research point to 25 years of new opportunities for research and analysis.

References

- AlNoamany, Y., AlSum, A., Weigle, M. C., & Nelson, M. L. (2014). Who and what links to the Internet Archive. *International Journal on Digital Libraries, 14*(3-4), 101-115.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature, 489*(7415), 295-298.
- Brügger, N. (2012). The idea of public service in the early history of DR online. In M. Burns & N. Brügger (Eds.), *Histories of public service broadcasters on the Web* (pp. 91-104). New York, NY: Peter Lang.
- Brügger, N., & Finnemann, N. O. (2013). The Web and Digital Humanities: Theoretical and Methodological Concerns. *Journal of Broadcasting & Electronic Media, 57*(1), 66-80. doi:10.1080/08838151.2012.761699
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication, 64*(2), 317-332. doi:10.1111/jcom.12084
- Davis, R. (1998). *The web of politics: The Internet's impact on the American political system*. Oxford University Press, Inc.
- Dhar, V. (2013). Data science and prediction. *Commun. ACM, 56*(12), 64-73. doi:10.1145/2500499
- Driscoll, K., & Walker, S. (2014). Big data, big questions| working within a black box: Transparency in the collection and production of big twitter data. *International Journal of Communication, 8*, 20.
- Foot, K., Warnick, B., & Schneider, S. M. (2005). Web-Based Memorializing After September 11: Toward a Conceptual Framework. *Journal of Computer-Mediated Communication, 11*(1), 72-96. doi:10.1111/j.1083-6101.2006.tb00304.x
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report, 8*(4), 597-606.
- Greenleaf, G. (2014). Sheherezade and the 101 data privacy laws: Origins, significance and global trajectories. *Journal of Law, Information and Science, 23*(1), 4- 49.
- Hindman, M. (2015). Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences. *The ANNALS of the American Academy of Political and Social Science, 659*(1), 48-62. doi:10.1177/0002716215570279
- Howison, J., & Wiggins, A. en Kevin Crowston. 2011. “Validity Issues in the Use of Social Network Analysis with Digital Trace Data”. *Journal of the Association for Information Systems, 12*(12), 767-797.

- Margolis, M., Resnick, D., & Tu, C.-c. (1997). Campaigning on the Internet Parties and Candidates on the World Wide Web in the 1996 Primary Season. *The Harvard International Journal of Press/Politics*, 2(1), 59-78.
- Mark, G. (2015). Joe Hill Centenary and IWW Songs in Australia. *Labour History*(109), 169-174. doi:10.5263/labourhistory.109.0169
- McKelvey, K., DiGrazia, J., & Rojas, F. (2014). Twitter publics: how online political communities signaled electoral outcomes in the 2010 US house election. *Information Communication & Society*, 17(4), 436-450. doi:10.1080/1369118x.2014.892149
- 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.
- Rill, S., Reinel, D., Scheidt, J., & Zicari, R. V. (2014). PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. *Knowledge-Based Systems*, 69, 24-33. doi:10.1016/j.knosys.2014.05.008
- Song, R., Liu, H., Wen, J.-R., & Ma, W.-Y. (2004). *Learning block importance models for web pages*. Paper presented at the Proceedings of the 13th international conference on World Wide Web.
- Spaniol, M., Denev, D., Mazeika, A., Weikum, G., & Senellart, P. (2009). *Data quality in web archiving*. Paper presented at the Proceedings of the 3rd workshop on Information credibility on the web, Madrid, Spain.
- Thelwall, M., & Vaughan, L. (2004). A fair history of the Web? Examining country balance in the Internet Archive. *Library & information science research*, 26(2), 162-176.
- Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological Methodology* (pp. 84-136). San Francisco: Joseey-Bass.
- Zeng, R., & Greenfield, P. M. (2015). Cultural evolution over the last 40 years in China: Using the Google Ngram Viewer to study implications of social and political change for cultural values. *International Journal of Psychology*, 50(1), 47-55. doi:10.1002/ijop.12125

The Changing Digital Faces of Science Museums

A diachronic analysis of museum websites

Anwasha Chakraborty & Federico Nanni

1. Introduction

Our book chapter is intended to be a starting point to discuss how prominent scientific institutions develop their websites over a period of time to communicate better with their visitors. More specifically, our work presents the formulation of a methodology for using websites as primary sources to trace and examine activities of scientific institutions through the years. The choice to study museums was prompted by the fact that these institutions are perceived by the civil society as authoritative custodians of artifacts, culture and heritage. This is true for science and technology museums as well, which have the additional task of communicating specialised (and often less understood) branches of knowledge.

The three case studies we have chosen include some of the most prominent science museums in Europe: the Science Museum, London; the Deutsches Museum, Munich and the Museo Nazionale della Scienza e della Tecnologia (hereafter, Museo della Scienza), Milan. In our work, we present a periodisation of the phases of significant structural and functional changes of the websites. The chapter chronicles what the websites have done since they were digitally born; what have been the milestones in terms of their development and the trajectory they are pursuing for better communication and relations with the public. It is important to point out that this work does not intend to offer a comparison of the three museums; rather the idea is to use a set of case studies to demonstrate our methodology, which focuses on the analysis of websites as primary sources for studying the recent history of scientific institutions.

2. Importance of museum websites

While there is little attention paid to websites as primary sources in museum studies, they have not been overlooked in discussions on the vast potential of digital domains in improving communication. Morrison (2006) states that a museum these days is not just the building, the collections and the staff, but also the website. Museums can connect with their visitors by providing them with more tools for personalized visits to the institution through their websites (Bowen et al, 1998; Bowen and Filippini-Fantoni, 2004). Marty (2007) argues that websites also boost attendance to the physical museum, and that a museum needs to have an effective and well-designed website, in order to

justify its rationale, attract visitors, showcase its importance in the social life of a specific area and a nation. Day (1997) and Cunliffe et al. (2001) posit the argument that the specific goals and objectives of the museum should be reflected on the website. One such goal is preservation of artifacts and Mason and McCarthy (2008) argue that a strong relationship can be ascertained between museums and computers and other digital tools over the last two decades, created to preserve data and digitize collections for better storage. It is fascinating that one of the most succinct explanations of the potential of websites is to be found in an early document of the Museo della Scienza, back in 1995¹. The document proposed certain characteristics that a successful website must possess, which are still relevant today even with the continuous advancements in the field of web technologies. These features include: dynamism (the ability to update information at a swift pace), hypertextuality (the possibility of creating linkages between articles and webpages both internally and externally), interactivity (considered as the most important ingredient for the success of the website) and the use of direct voice (the need to address the visitors informally).

3. Sources and Methods

In order to conduct our analysis, we employed three types of primary sources. First of all, we used snapshots of websites preserved by the Internet Archive, as these materials hold the potential of presenting the changes in layout, structure and contents of a website through time. Secondly, we conducted a series of interviews with the people who are managing the websites at present. By combining these sources, our goals were (a) to obtain a comprehensive perspective on the evolution of science museum websites and (b) to understand the reasons behind specific changes and decisions. We expanded our analysis further by adopting resources available on the live web. The live web includes the current version of the website and the presence of the museum on social media.

4. Diachronic analysis of the websites' past

4.1 The 'leaflet' museum website

The early use of the website as an information leaflet of the museum corresponds with Schweibenz's (2004) analysis of the typologies of museum websites, in which he

¹<https://web.archive.org/web/20041217025646/http://www.museoscienza.org/INTERNET/prog1995.htm>

describes the first category as the ‘brochure museum’.

The Museo della Scienza has been online since 1995, with an essential informative interface on the activities of the museum, its collections and the opening hours². In 1997, the Deutsches Museum website offered a similar homepage with several pieces of information, both in German and in English. From news regarding the museum, to descriptions of special exhibition to courses and seminar offered by this structure, this website could be identified as the paradigmatic version of Schweibenz’s brochure website discussed earlier. Moreover, a multimedia section (available in 1997 under the rubric “dioramen”) offered interactive demonstrations of specific expositions, videos and other contents. Great attention was paid to describe its archive devoted to the history of science and technology, one of the most important in Europe. Part of the catalogue was presented online and could be directly consulted by users. As pointed out by Annette Lein in the interview, in this early version of the website, “...the aim was to provide information for visitors and to present the exhibitions online with text and photos worldwide. Other main topics were: calendar of events, a list of people working in the museum, a glance at archives, library and research institute, feedback form, contact.”

4.2 The rise and fall of the virtual museum

During the first part of the 2000s, a common trend characterized the efforts that the three museums focused on their digital space. Moving from a conception of the website as a series of pages where pieces of information regarding collections and archives were presented, they started experimenting with virtual reality. The goal was to offer new experiences to the user who would have the opportunity of interact with the collections directly from home.

In 1999, the Museo della Scienza introduced a virtual museum called “Leonardo Virtuale”, created in collaboration with the Polytechnic of Milan. Using this software, users could “walk” and “fly” through the different rooms of the museum and explore its collections. Visitors had the possibility of meeting and chatting with other guests and following a virtual guide. An innovative choice was to adopt Webtalk, a technology developed by the Hypermedia Open Center (HOC) of the Polytechnic of Milan based on VRML (Virtual Reality Markup Language) and Java, which guaranteed a real time interaction between different users.

In 2001 the London Science Museum also introduced the “Exhibitions online”

²<https://web.archive.org/web/20001012004412/http://www.museoscienza.org/INTERNET/sito.html>

section. Similar to a virtual museum, in this case visitors could explore the different collections from their desks, from Babbage's machine to Marie Curie and the history of radioactivity. Moreover, interaction with a series of objects was offered (for example, the possibility of exploring the Apollo 10³ module).

The potential of the web as a place to interact with digital collections gave museum practitioners the chance of experimenting new forms of education. For example, the "Learn and Teach" section of the Science Museum presented several pieces of information for family and school classes and offered a series of activities that could be conducted from home⁴.

Giving users the possibility of interacting with digital objects was a fundamental aspect of the changes in the Deutsches Museum website as well. As early as in 1998 a "Museum Multimedial" was created with a link available on the homepage. During the following years this page offered a large variety of resources⁵: from dioramas to live-cameras, from video-clips to interactive demonstrations.

Even though all three science museums showed ample interest in offering interactive collections online (e.g. in the form of virtual tours), the realisation of the virtual version had to face serious technological impediment, which contributed to the demise of the trend of creating virtual museums. First of all, the majority of internet connections were not sufficiently fast and stable to fully support the usability of these services. Secondly, 3D graphic reconstructions were usually not accurate enough to be considered a real substitute to a visit to the museum. Discussing the diminishing popularity of the concept of the virtual museum, Cavallotti comments on the strategy adopted by Museo della Scienza: "This concept was very popular some years ago. But the museum's website team does not want to make a difference between the virtual and physical museum. The team uses digital tools to talk about the physical museum."

4.3 The outreach-museum website: ensuring greater engagement

³<http://web.archive.org/web/20021002123300/http://www.sciencemuseum.org.uk/on-line/apollo10/intro.asp>

⁴<http://web.archive.org/web/20021001232233/http://www.sciencemuseum.org.uk/education/families/online.asp>

⁵<http://web.archive.org/web/20021019104649/http://www.deutsches-museum.de/mum/start.htm>

While the virtual museum as a trend had started waning by the middle of the 2000s, increased attention was being paid by website managers to present collections in innovative ways (e.g. enhancing digital collections and accurately describing them in dedicated blogs) and to improve communication and dialogue with the users. In museum studies literature, this was the period of increasing number of scholarly works being written about the concepts of participatory designs in museum exhibits, and co-participation of visitors in determining the content inside the museum space. Among them, *The Participatory Museum* (2010) by Nina Simon explains in detail how greater visitor participation can be achieved. She argues that authority over content should be shared between the museum and its visitors, citing O'Reilly's definition of Web 2.0, "an application that gets better the more people use it."⁶ Following the greater focus on more inputs from the public as emphasised by Web 2.0 theories, the websites examined in this chapter show similar structural changes at this stage of the decade to accommodate more voices.

Museums have offered snippets of information regarding their collections and new exhibitions to their users, in particular through newsletters, since the Nineties. However, during the second part of the 2000s there was a greater focus on communicating through blog posts and podcasts to go beyond the traditional authoritative role of these institutions. An excellent example was the Science Museum's "Antenna", a constantly updated resource for science news. Blogs describing the collections were also developed, as the "Stories" from the Science Museum and "Der Blog das Deutsches Museum" shows. The second one, which has been regularly updated since 2009, has been mainly written in German, highlighting their target users.

4.4 The social-media website: managing large networks of communication

Science museums have been constantly trying to improve communications with their visitors since Nineties, through emails and online forums in order to improve interaction with visitors. This could be seen, for example, in the "Let's Talk" section of the Science Museum in 2002⁷. The aim of this section was to answer questions, to create discussion forums about scientific themes and to receive specific feedbacks.

More recently however, these institutions decided to use two of the most frequented

⁶ See <http://www.participatorymuseum.org/chapter3/>

⁷http://web.archive.org/web/20021004110407/http://www.sciencemuseum.org.uk/lets_talk/index.asp#

social networking websites (Facebook and Twitter), which allow many to many communication. By 2011, all three museums have been using Twitter and Facebook for greater engagement with users. On social media, the Museo della Scienza⁸ and the Deutsches Museum are primarily communicating with their national audience, as can be noticed by the language of the posts and tweets. The Science Museum, which was actively using both Facebook and Twitter already in 2009, represents a good example of how to use social media to communicate with the visitors using specific thematic news, hashtags and photos. The digital department has also launched a mobile version of the website, to cater to a very large audience who frequent the web using their phones.

Given that social media is gradually becoming the main dialogic space for museums and visitors, we also need to consider that content generated in these networking spaces will not necessarily be completely available to the museum for specific analyses. In fact, while large amount of user-generated data in these social networking sites can help museums to engage better with the public (e.g. by carrying out large scale visitor surveys), efforts have to be made to obtain them from the parent companies. The history of museum websites in the recent future will not only be contained in the preserved snapshots of their website's pages, but also will be found on the walls and the tweets shared among interested groups.

⁸ Cavallotti mentions that the Museo Scienza has hired a social media personnel on a part time basis from 2015.

References

- Bowen, J. P., Bennett, J. & Johnson, J. (1998). Virtual Visits to Virtual Museums, In D. Bearman and J. Trant (eds.), *MW98: Museums and the Web Conference*, Toronto, Canada.
- Ben-David, A. (2012). The Palestinian diaspora on the Web: Between de-territorialization and re-territorialization, *Social Science Information*, 51 (4), 459-474.
- Ben-David, A. (2015). What Does the Web Remember of its Deleted Past? An Archival Reconstruction of the Former Yugoslav Internet, *Two-day conference at Aarhus University*, Denmark.
- Bowen, J. P., & Filippini-Fantoni, S. (2004). Personalization and the Web from a Museum Perspective, *Museums and the Web Conference*.
- Brügger, N. (2005). Archiving Websites, *General Considerations and State*.
- Brügger, N. (2010). *Web history*, Peter Lang.
- Brügger, N. (2012). Web historiography and Internet Studies: Challenges and perspectives, *New Media & Society*.
- Brügger, N. (2015). A brief history of Facebook as a media text: The development of an empty structure, *First Monday*, 20 (5).
- Cunliffe, D., Kritou, E., & Tudhope, D. (2001). Usability Evaluation for Museum Web Sites, *Museum Management and Curatorship*, 19 (3), 229–252.
- Davies, A. (2007). Virtual museums: Out of this world. *Museums Journal*.
- Day, A. (1997). A model for monitoring web site effectiveness. *Internet Research: Electronic Networking Applications and Policy*, 7(2), 109–115.
- Dougherty, M., Meyer, E. T., Madsen, C. M., Van den Heuvel, C., Thomas, A., & Wyatt, S. (2010). Researcher engagement with web archives: State of the art. *Joint Information Systems Committee Report*.
- Foot, K., Schneider, S. M., Dougherty, M., Xenos, M., & Larsen, E. (2003). Analyzing linking practices: Candidate sites in the 2002 US electoral Web sphere, *Journal of Computer-Mediated Communication*, 8 (4).
- Gomes, D., Miranda, J., & Costa, M. (2011). A survey on web archiving initiatives, *Research and advanced technology for digital libraries*, pp. 408-420.
- Hale, S. A., Yasseri, T., Cowls, J., Meyer, E. T., Schroeder, R., & Margetts, H. (2014). Mapping the UK webspace: fifteen years of british universities on the web, *Proceedings of the 2014 ACM conference on Web science*.
- Henning, M. (2006). *New media - A companion to museum studies*, ed. Sharon MacDonald.
- Howell, B. A. (2006). Proving web history: How to use the Internet Archive, *Journal of*

Internet Law, 9(8), 3-9.

Lepore, J. (2015). The Cobweb: Can the Internet be archived?, *The New Yorker*.

Lyman, P., & Kahle, B. (1998). Archiving digital cultural artifacts. *D-Lib Magazine*, 4 (7).

Marty, P. (2007). Museum Websites and Museum Visitors: Before and After the Museum Visit, *Museum Management and Curatorship*, 22 (4), 337-360.

Mason, D.M., & McCarthy, C. (2008). Museums and the culture of new media: an empirical model of New Zealand museum websites, *Museum Management and Curatorship*, 23 (1), 63-80.

Milligan, I. (2012). Mining the 'Internet Graveyard': Rethinking the Historians' Toolkit, *Journal of the Canadian Historical Association/Revue de la Société historique du Canada*, 23 (2), 21-64.

Milligan, I. (2015). Welcome to the GeoHood: Using the GeoCities Web Archive to Explore Virtual Communities, *Two-day conference at Aarhus University, Denmark*.

Morrison, J. (2006). Cultural websites: Making the network, *Museums Journal*.

Nanni, F. (forth. 2016). Reconstructing a website's lost past - Methodological issues concerning the history of www.unibo.it, *Digital Humanities Quarterly*.

Schweibenz, W. (2004). Virtual museums: The Development of Virtual Museums, *ICOM News Magazine*, 3 (3).

Simon, N. (2010). *The participatory museum*. Museum 2.0.

Srinivasan, R. (2013). Re-thinking the cultural codes of new media: The question concerning ontology, *New Media & Society*, 15 (2), 203-223.

Webster, P. (2015). Will historians of the future be able to study Twitter?, URL: <http://peterwebster.me/2015/03/06/future-historians-and-twitter/>

Wilson, R.J. (2011). Behind the scenes of the museum website, *Museum Management and Curatorship*, 26 (4), 373-389.

Yasko, J. (2007). Museums and web 2.0, *Museum News*, 43 (7).

Zimmer, M. (2015). The Twitter Archive at the Library of Congress: Challenges for information practice and information policy, *First Monday*, 20 (7).