

# WEB ARCHIVABILITY

A Pathway Towards Dialogue

Written by Itza A. Carbajal | December 2016

Dodging the Memory Hole Scholarship Recipient

## Introduction

The days of rushing to a newsstand to read about the latest and greatest no longer exists. Breaking News now comes to us through constantly changing websites, databases, or mobile applications all viewable on those bright computer screens ready for immediate consumption. Since the introduction of the internet, the ways humans consume the news continues to transform everyday. Readers can be recognized as the social media user reading online articles on the news feed, the online subscriber tuning in daily to national newspaper headlines, to the researcher gathering data from news related databases and the teacher following educational blogs. Online news informs people about their past, surroundings, and possible futures. If online news now represents a unique testimony of humanity's existence then there must exist a mandate to preserve it. The question then becomes not why, but how can people approach such an integral duty and who bears the responsibility of such a task.

Preserving humanity's digital memories will not be a simple task. Web archiving continues to evolve just as online technology evolves, despite recent advances. For the practice of web archiving to gain a foothold, educational components and constant dialogue calling for preservation friendly practices and standards in web content development must be emphasized. This research paper proposes the incorporation of the concept of web archivability in discussions about preserving online news. The following information will touch upon issues surrounding the concept of web archivability including the notions behind web archivability, a brief history of online news, the differences between web archive stakeholders and a brief discussion on current web archivability implementations.

## Terminology and Scope

For purposes of clarity, this paper will define a few terms of interests for the reader. Web archivability will be defined as a metric measuring the ability of web content to be preserved as a means of evaluating the degree of how likely a web based resource can be preserved. Web based structures while not a formal definition will be used as a general term that includes websites, online databases, blogs, and others in particular those built using HTML, CSS, Javascript, PHP, and Python and dependent on using a web browser for viewing. Other digital news structures such as mobile applications will not be directly addressed as they have yet to be measured in formal research papers using web archivability approaches.

## What is web archivability

When building a website, web developers and designers take into account multiple factors including accessibility, performance, search engine optimization, standards compliance, and usability.<sup>1</sup> But what about preservation considerations such as web archivability, the

---

<sup>1</sup>Vangelis Banos, Yunhyong Kim, Seamus Ross, and Yannis Manolopoulos. 2013. "CLEAR: A Credible Method to Evaluate Website Archivability." Lisbon, Portugal. [http://purl.pt/24107/1/iPres2013\\_PDF/CLEAR%20a%20credible%20method%20to%20evaluate%20website%20archivability.pdf](http://purl.pt/24107/1/iPres2013_PDF/CLEAR%20a%20credible%20method%20to%20evaluate%20website%20archivability.pdf).

measurement of how easily a website's components including structure, presentation, content, and functionality can be saved for future access and use? Web archivability, as a method of measurement of how complete or accurate the archival copy will mirror the original, could provide the gateway to collaboration between the creators of the content and the people in charge of preserving it.

The concept of web archivability can be traced to a 2013 study by researchers Vangelis Banos, Yunhyong Kim, Seamus Ross, and Yannis Manolopoulos titled "CLEAR: a credible method to evaluate website archivability." Web archivability as a metric unit can be the defining factor on whether a web based structure can be archived and to what extent the captured copy will represent an accurate and complete representation. Just as many web archiving programs focus on evaluating the quality after the web capture, the practice of web archivability can be seen as a component of a pre-screening process. Web archivability can also be used to gauge the difficulty levels of web capture as it relates to web structure components such as HTML, CSS, images, robots.txt, valid hyperlinks, Javascript or video.

Web archivability also addresses the impact of the ability of capturing tools, the "reliability of the mechanism used to archive, [and] the frequency at which the mechanism is run" on the final preservation copy.<sup>2</sup> For purposes of this paper, this aspect will be briefly mentioned, but must be considered when facilitating conversation between stakeholders. Web archiving tools affect the extent of what can be expected from the final web archive item. For example Heritrix3 struggles with sites using robots.txt, using by default older file-naming conventions if conducting recurring crawls and often interprets URLs with hash symbols as being the same page if the only difference occurs after the symbol. Web based structure creators contribute most directly to these problems and can most easily provide solutions instead of relying on the web archiving team to grapple with alternative capture methods.

### **What is Online News**

A few decades ago, the idea of preserving the news could be easily categorized as belonging to one of the three main technologies: television, radio, or print. The early origins of online news can be traced to the early 1990s, as news agencies such as the TIME already began to embrace the electronic option.<sup>3</sup> This interest in utilizing the internet to reach mass audiences also coincided with the emergence of preservation work of other digital media such as the foundation of the Internet Archive in 1996.<sup>4</sup> By 1995, despite a number of news agencies providing electronic sources, the "news websites were typically little more than repositories of

---

<sup>2</sup> Mat Kelly, Justin F. Brunelle, Michele C. Weigle, and Michael L. Nelson. 2013. "On the Change in Archivability of Websites Over Time." *Research and Advanced Technology for Digital Libraries*, Lecture Notes in Computer Science, 8092 (July): 36.

<sup>3</sup> Allan Stuart. 2006. *Online News: Journalism and the Internet*. Maidenhead, England; New York: Open University Press: 15

<sup>4</sup> M. Toyoda and M. Kitsuregawa. 2012. "The History of Web Archiving." *Proceedings of the IEEE 100* (Special Centennial Issue): doi:10.1109/JPROC.2012.2189920: 1441.

reports previously published elsewhere...”<sup>5</sup> On April 19th 1995 as an indirect result of the Oklahoma City domestic terrorism bombing, the role of online news would drastically change forever.

During the Oklahoma City tragedy, people flocked to online sources in an attempt to gather information at various times of the day with an almost unprecedented immediacy. This in turn caused some traditional news sources to question their age-old practices, as they realized “the sudden liberation of newspapers from the time constraints associated with print...”<sup>6</sup> As online breaking news stories regarding the Oklahoma City bombing caught the world’s attention, the newspaper print industry recognized the value in producing immediate news. It became “increasingly obvious that they would not be able to compete with their electronic rival...”<sup>7</sup>

Fast forward to the present, years after the early days of online news. Today people utilize online news in a variety of ways: subscription databases, web news aggregators, web-only news sites, blogs, online newspaper databases, newspaper websites, news media websites, grassroots journalism sites, WikiLeaks, and social network sites, to name a few. As the diversity of online news expands, the concept of what constitutes news and what does not becomes progressively more complex. At a recent gathering known as Dodging the Memory Hole: Saving Online News, some attendees even hinted that news to them consisted of the actual data or content, not just the source or presentation of that data. If this is the case, then the entirety of the web could be considered news, resulting in an even bigger need for preservation efforts in order to ensure access for future use. This, in turn, complicates the issue of web archiving to an even larger and more complex nature.

### **Why Web Archiving**

Before web archiving can even begin, stakeholders from various areas must be convinced of the importance of web archiving. In thinking about archives, one concept currently emphasized by the National Society of American Archivists comes to mind: enduring value. Online news could easily be considered as having enduring value, although stakeholders will differ as to the length of how long materials should be preserved. For many of these stakeholders, importance must be framed according to their own expectations. Memory or education-based institutions maintain a strong history of placing attention and resources towards preservation efforts. These institutions oftentimes sense an innate intangible worth in materials for purposes of higher learning and for the safeguarding of society’s historical memory. In a 2012 survey conducted by the National Digital Stewardship Alliance, a surge in web archiving

---

<sup>5</sup> Ibid 16.

<sup>6</sup> Ibid 20.

<sup>7</sup> Ibid 21.

activities could be visibly seen mostly in not for profit institutions.<sup>8</sup> Web archives serve as testimony to the historical role these institutions play; “it is the public face of the institution and its community.”<sup>9</sup>

In the for-profit field, enduring value may not be as discernable. For some news agencies, the news created for online consumption ceases to matter the moment readers stop reading. This issue brings into question the need for active and clear discussions on returns on investment for those not in the business of preservation. Unfortunately majority of news agencies in the past have paid little to no attention to preservation efforts mostly due to a perceived notion of having little monetary value or little historical value to their business needs. Others may have interest, but little resources to pursue these sorts of initiatives. This can be seen through the example of early television broadcast and their “scant regard to preservation.”<sup>10</sup> Other factors also heavily contributed to the lack of preservation of early television content including high costs of equipment, the fluid nature of the material, and concerns of obstructing the rights or workers or creators. Alongside these obstacles, there also existed a strong “lack of appreciation of the cultural significance of television”<sup>11</sup> While television along with radio has slowly gained the recognition of value it deserves, web archives appear to be in a similar situation as the early productions of television and radio.

As the number of web archiving initiatives grows, the significance of web archiving gains traction. More institutions view web archives as “innovative systems that acquire, store and preserve information published on the web” for future use cases by various communities as well as future generations.<sup>12</sup> The biggest issue then falls on the divide between those already doing the work and those not. If the bulk of the web archiving work falls on a small group producing the least amount of online news, then the need for larger and more extensive resources also increases. Sadly, most not-for-profit institutions do not possess the ability to increase their funding to match the ever increasing demands placed on them. Online news will also potentially grow as a result of the news industry restructuring itself leading to a greater need for more initiatives and resources for archiving.<sup>13</sup> The challenge of web archiving cannot be tackled by

---

<sup>8</sup> NDSA Content Working Group. *Web Archiving Survey Report*. Report. National Digital Stewardship Alliance. June 19, 2012. Accessed November 11, 2016. [http://www.digitalpreservation.gov/documents/ndsa\\_web\\_archiving\\_survey\\_report\\_2012.pdf](http://www.digitalpreservation.gov/documents/ndsa_web_archiving_survey_report_2012.pdf)

<sup>9</sup> Duckworth A. Antracoli, S., Silva, J., & Yarmey, K. (2014). Capture all the URLs: First steps in web archiving. *Pennsylvania Libraries*, 2(2), doi:<http://dx.doi.org/10.5195/palrap.2014.67>: 156

<sup>10</sup> Chris Perry and Simon Coward. 2010. "Swiped or Wiped? Kaleidoscope's Part in the Recovery of Lost Television Programmes." *Critical Studies In Television* 5, no. 2: *Humanities Full Text (H.W. Wilson)*, EBSCOhost (accessed December 8, 2016): 49

<sup>11</sup> Ibid

<sup>12</sup> Gomes, Daniel, João Miranda, and Miguel Costa. 2011. "A Survey on Web Archiving Initiatives." In *Research and Advanced Technology for Digital Libraries: International Conference on Theory and Practice of Digital Libraries, TPD L 2011, Berlin, Germany, September 26-28, 2011. Proceedings*, edited by Stefan Gradmann, Francesca Borri, Carlo Meghini, and Heiko Schuldt. Berlin, Heidelberg: Springer Berlin Heidelberg. [http://dx.doi.org/10.1007/978-3-642-24469-8\\_41](http://dx.doi.org/10.1007/978-3-642-24469-8_41): 408

<sup>13</sup> Amy Mitchell and Jesse Holcomb. 2016. "State of the News Media 2016." *Pew Research Center's Journalism Project*. June 15. <http://www.journalism.org/2016/06/15/state-of-the-news-media-2016/>.

one entity and for this reason, the responsibility of web archiving should be more evenly distributed.

### **Web Archiving Stakeholders**

When thinking about why people may want to archive online news, a brief description of stakeholders should be highlighted. While the final users of archives will vary drastically from repository to repository, the term stakeholders here applies to the user groups at the point of creation, the moment of preservation, as well as the points of use after preservation has commenced.

The first group focuses on the people involved in the creation of the online news source including news agency staff, board members, funders, content creators such as writers or photographers, web developers for the online structures, as well as countless other groups that contribute to the initial publication. This group in particular must be approached from the beginning, as the allocation of resources and implementation of web design plans critically define the level of success of any web archiving initiative. In particular, web design professionals in this group must be clearly instructed to consider the implications of website development on the likelihood their design can be archived. At times, members of this group may also exist in the second group such as the case with the New York Times archives and their effective reuse of archival materials for future publications.<sup>14</sup>

The second group includes the people in charge of preservation efforts, which would also vary according to institution. In the realm of web archiving, this stakeholder group could encompass archivists, digital project coordinators, data curators, student technicians, metadata specialists, and web capture tool developers. This group primarily focuses on the act of web archiving, but often has little say in modifying aspects of the web content after creation. In order to alleviate the need for more resources including time and labor, this group should make an active effort to communicate their expectations and needs with the first stakeholder group.

The final group is the most easily recognized as it encompasses users of archival materials including researchers, student, journalists, educators, policy analysts, etc. Historically this group has been viewed as a passive consumer rather than an active participant, but that approach can be detrimental if the news agency requires clear indications of profitable use as a case for the investment in web archiving efforts. More than the two previous groups, this group helps define the extent of valuable and usefulness of a web archive. In relation to web archivability, the needs of this group should strongly influence the ways completeness and accuracy are defined. As Brunelle et al. point out, users in this group can define the “relative

---

<sup>14</sup> Hare, Kristen. "How The New York Times Is Using Unpublished Images from the Archives to Tell Stories It Missed the First Time." Poynter. February 02, 2016. Accessed December 1, 2016. <http://www.poynter.org/2016/how-the-new-york-times-is-using-unpublished-images-from-the-archives-to-tell-stories-it-missed-the-first-time/394336/>.

value of embedded resources and assign a damage rating to archived pages as a way to evaluate archival success.”<sup>15</sup>

### **Web Archivability in Practice**

Web archiving evaluation tools and methods look at different factors such as accessibility, ingest quality, metadata & standards compliance alongside the technical structures to assess the web archive readiness of some of the nation’s top news providers. The following two approaches by no means exhaust the possibilities of how online news archiving can approach web archivability. Rather, these two instances provide contrasting strategies – one being the act of advocating for proper practices within the creating community –the other being the use of an almost out-of-the-box external evaluation application.

Web archivability will almost always rely on the people making the decisions on what ends up on the website and how. From the use of different scripting languages to third party platforms, the decisions gone into the development of a web source will drastically influence the the archivability of a website. For this reason, many memory institutions that also exist alongside the creating institution establish protocols used towards the creation of web sources. At Stanford, the university libraries system created a division geared towards the preservation of online content, mainly those created by and for the institution’s community.<sup>16</sup> This division, alongside the activities of capturing online content, also focuses on providing guidance to different university divisions, especially the web developers. For the Stanford web archiving team, web builders represent facilitators of web archiving who can most easily follow archivability recommendations presented by the web archiving division. Other agencies such as the Library of Congress also follow a similar practice of promoting certain practices in order to facilitate the ingestion of web archives. On their website they address issues of what formats to use, what standards to implement, as well as areas to be wary of.<sup>17</sup>

Maintaining clear lines of communication between divisions within an institution plays a crucial role in the success of a web archive project, but at other times a lack of time and resources requires quicker solutions. As mentioned previously, the development of the term web archivability stems from a study on how web archiving can occur effectively. In this study the researchers as part of the Credible Live Evaluation of Archive Readiness (CLEAR) method developed ArchiveReady.com, a free online application component of the Credible Live Evaluation used as an external measurement tool for assessing a web based structures

---

<sup>15</sup> Brunelle, Justin F., Mat Kelly, Hany SalahEldeen, Michele C. Weigle, and Michael L. Nelson. 2015. “Not All Mementos Are Created Equal: Measuring the Impact of Missing Resources.” *International Journal on Digital Libraries* 16 (3–4). doi:10.1007/s00799-015-0150-6: 283

<sup>16</sup> Stanford University Libraries. “Guidance on Building Archivable Websites.” 2016. Accessed October 3. <http://library.stanford.edu/blogs/digital-library-blog/2014/09/guidance-building-archivable-websites>.

<sup>17</sup> Library of Congress. 2016. “Recommended Formats Statement for Websites.” Web page. *Preservation*. Accessed October 8. <https://www.loc.gov/preservation/resources/rfs/websites.html>.

archivability.<sup>18</sup> ArchiveReady.com uses the archivability facets: Accessibility, Cohesion, Metadata & Standards Compliance included in CLEAR and now CLEAR+, the evolved version of the original. These quantitative metrics used in “diagnosing whether [the web based structure] has the potentiality to be archived with completeness and accuracy”<sup>19</sup> can radically impact a web archiving program’s decision to capture web based news. When users input a target URL into ArchiveReady.com, the web-based application analyzes the components of the web structure against the aforementioned archivability facets. ArchiveReady.com then imitates a web crawler by initiating multiple HTTP connections to retrieve and “analyse only the URL submitted by the user, it does not evaluate the whole website recursively.”<sup>20</sup> These connections are then evaluated and the measurements of archivability are calculated using the CLEAR method. Afterwards users see an overall rating, individual ratings for the five archivability facets, as well as different website attributes including possible errors, positive components, and additional information. One of the most attractive features of a service such as ArchiveReady.com rests on the seemingly easy implementation of verification for web based structures. Instead of assigning a staff person to manually check each potential web archive for errors, this service bundles up a five distinct areas that contribute to website completeness and verify them all simultaneously.

Other approaches include providing sitemaps, maintaining updated hyperlinks, and incorporating proper and complete metadata.<sup>21</sup> Web archivability standards can be customized to address abilities and limitations of the agreed upon web capture tool. With the ever increasing need for web archives, these tools can only increase in numbers and abilities.

## Conclusion

Using web archivability approaches barely scratches the surface on the problem of web archiving. Web technologies continue to change at an alarming rate. If memory institutions or long term preservation entities wish to keep up, there must be a shift in strategy. More attention should be focused on opening lines of communication among different stakeholders. This would then be coupled with developing clear standards on what can be archived and what steps need to be taken in order to complete the archiving process. Ultimately a successful web archiving initiative will not only include a strategy towards web archivability, but also an emphasis on commitment from contributing stakeholders. For smaller news providers or governmental agencies in conflict zones, web archivability may not be easily implemented either as result of low resources or resistance to accessibility. Web archives depend heavily on consent and cooperation, and without either, efforts will fall flat. In order for future generations to have the

---

<sup>18</sup> “ArchiveReady.com: Website Archivability Testing Tool.” 2016. Accessed October 3. <http://archiveready.com/>.

<sup>19</sup> Banos

<sup>20</sup> Banos

<sup>21</sup> Helen Hockx-Yu. 2012. “How to Make Websites More Archivable?” *UK British Library Web Archive Blog*. September 5. <http://blogs.bl.uk/webarchive/2012/09/how-to-make-websites-more-archivable.html>.



ability to witness their past as documented through online news, partnerships and dialogue focused on making online content archivable must start today.

## References

Antracoli, A., Duckworth, S., Silva, J., & Yarmey, K. (2014). Capture all the URLs: First steps in web archiving. *Pennsylvania Libraries*, 2(2), 155-170.

doi:<http://dx.doi.org/10.5195/palrap.2014.67>

“ArchiveReady.com: Website Archivability Testing Tool.” 2016. Accessed October 3.

<http://archiveready.com/>.

Bailey, Jefferson. "The NDSA Web Archiving Survey Post." *The Signal*. June 5, 2012.

Accessed November 24, 2016. <http://blogs.loc.gov/thesignal/2012/07/the-ndsa-web-archiving-survey/>.

Banos, Vangelis, Yunhyong Kim, Seamus Ross, and Yannis Manolopoulos. 2013. “CLEAR: A Credible Method to Evaluate Website Archivability.” Lisbon, Portugal.

[http://purl.pt/24107/1/iPres2013\\_PDF/CLEAR%20a%20credible%20method%20to%20evaluate%20website%20archivability.pdf](http://purl.pt/24107/1/iPres2013_PDF/CLEAR%20a%20credible%20method%20to%20evaluate%20website%20archivability.pdf).

Brunelle, Justin F., Mat Kelly, Michele C. Weigle, and Michael L. Nelson. 2015. “The Impact of JavaScript on Archivability.” *International Journal on Digital Libraries* 17 (2): 95–117. doi:10.1007/s00799-015-0140-8.

Center for Research Libraries. *Preserving News in the Digital Environment: Mapping The Newspaper Industry In Transition*. Report. April 27, 2011. Accessed November 30, 2016.

Davis, Robin Camille. 2016. “Die Hard: The Impossible, Absolutely Essential Task of Saving the Web for Scholars.” *Faculty Publications and Research*, May.

[http://academicworks.cuny.edu/jj\\_pubs/76](http://academicworks.cuny.edu/jj_pubs/76).

Dwyer, Jim. "Harvesting Government History, One Web Page at a Time." N.Y/Region. December 1, 2016. Accessed December 2, 2016.

[http://www.nytimes.com/2016/12/01/nyregion/harvesting-government-history-one-web-page-at-a-time.html?\\_r=1](http://www.nytimes.com/2016/12/01/nyregion/harvesting-government-history-one-web-page-at-a-time.html?_r=1).

Hare, Kristen. "How The New York Times Is Using Unpublished Images from the Archives to Tell Stories It Missed the First Time." Poynter. February 02, 2016. Accessed December 1, 2016. <http://www.poynter.org/2016/how-the-new-york-times-is-using-unpublished-images-from-the-archives-to-tell-stories-it-missed-the-first-time/394336/>.

Hockx-Yu, Helen. 2012. "How to Make Websites More Archivable?" *UK British Library Web Archive Blog*. September 5. <http://blogs.bl.uk/webarchive/2012/09/how-to-make-websites-more-archivable.html>.

Gomes, Daniel, and Miguel Costa. 2014. "The Importance of Web Archives for Humanities." *Journal Of Humanities & Arts Computing: A Journal Of Digital Humanities* 8, no. 1: 106-123. *Academic Search Alumni Edition*, EBSCOhost (accessed December 8, 2016).

Gomes, Daniel, João Miranda, and Miguel Costa. 2011. "A Survey on Web Archiving Initiatives." In *Research and Advanced Technology for Digital Libraries: International Conference on Theory and Practice of Digital Libraries, TPDL 2011, Berlin, Germany, September 26-28, 2011. Proceedings*, edited by Stefan Gradmann, Francesca Borri, Carlo Meghini, and Heiko Schuldt, 408–20. Berlin, Heidelberg: Springer Berlin Heidelberg. [http://dx.doi.org/10.1007/978-3-642-24469-8\\_41](http://dx.doi.org/10.1007/978-3-642-24469-8_41).

Kelly, Mat, Justin F. Brunelle, Michele C. Weigle, and Michael L. Nelson. 2013. "On the Change in Archivability of Websites Over Time." *Research and Advanced Technology for Digital Libraries*, Lecture Notes in Computer Science, 8092 (July): 35–47.

Kelly, Mat, Michael L. Nelson, and Michele C. Weigle. n.d. "The Archival Acid Test: Evaluating Archive Performance on Advanced HTML and Javascript."

Library of Congress. 2016. "Recommended Formats Statement for Websites." Web page. *Preservation*. Accessed October 8. <https://www.loc.gov/preservation/resources/rfs/websites.html>.

Mitchell, Amy, and Jesse Holcomb. 2016. "State of the News Media 2016." *Pew Research Center's Journalism Project*. June 15. <http://www.journalism.org/2016/06/15/state-of-the-news-media-2016/>.

NDSA Content Working Group. *Web Archiving Survey Report*. Report. National Digital Stewardship Alliance. June 19, 2012. Accessed November 11, 2016. [http://www.digitalpreservation.gov/documents/ndsawebarchiving\\_survey\\_report\\_2012.pdf](http://www.digitalpreservation.gov/documents/ndsawebarchiving_survey_report_2012.pdf)

Padilla, Thomas. "How Easy Is It to Search the Complete Archives of the New York Times Online?" History News Network. February 18, 2007. Accessed November 28, 2016. <http://historynewsnetwork.org/article/34785>.

Perry, Chris, and Simon Coward. 2010. "Swiped or Wiped? Kaleidoscope's Part in the Recovery of Lost Television Programmes." *Critical Studies In Television* 5, no. 2: 48-59. *Humanities Full Text (H.W. Wilson)*, EBSCOhost (accessed December 8, 2016).

Rollason-Cass, Sylvie, and Scott Reed. 2016. "Living Movements, Living Archives: Selecting and Archiving Web Content During Times of Social Unrest." Accessed October 9. <http://www-tandfonline-com.ezproxy.lib.utexas.edu/doi/pdf/10.1080/13614576.2015.1114839?needAccess=true>.

Shein, Esther. 2016. "Preserving the Internet." *Communications of the ACM* 59 (1). <http://cacm.acm.org/magazines/2016/1/195738-preserving-the-internet/fulltext>.

Stanford University Libraries. 2016. "Archivability." *Web Archiving*. Accessed October 3. <https://library.stanford.edu/projects/web-archiving/archivability>.

Stanford University Libraries. "Guidance on Building Archivable Websites." 2016. Accessed October 3. <http://library.stanford.edu/blogs/digital-library-blog/2014/09/guidance-building-archivable-websites>.

Stuart, Allan. 2006. *Online News: Journalism and the Internet*. Maidenhead, England; New York: Open University Press.

Toyoda, M., and M. Kitsuregawa. 2012. "The History of Web Archiving." *Proceedings of the IEEE* 100 (Special Centennial Issue): 1441–43. doi:10.1109/JPROC.2012.2189920.

Wortham, Jenna. 2016. "How an Archive of the Internet Could Change History." *The New York Times*, June 21. <http://www.nytimes.com/2016/06/26/magazine/how-an-archive-of-the-internet-could-change-history.html>.