Emily Drachman

A Review of *History in the Age of Abundance?: How the Web is Transforming Historical Research* by Ian Milligan

Whether we like it or not, the 1990’s is history. To study that history without consulting the internet would be to create an entirely incomplete record. There’s a wealth of information to be found on the internet from an incredible variety of sources, ranging from news publications to public reactions on social media. The nature of the internet is that it’s constantly updating, and there’s no finished product. It’s an ephemeral place where things are constantly appearing and disappearing. So how then do we use this ever-evolving source while it’s in flux? To study recent history, it will be invaluable to be able to consult webpages as they existed at the times we wish to study. Ian Milligan tackles these topics head on in his book *History in the Age of Abundance?: How the Web is Transforming Historical Research*. In our era of big data, it is essential to think about the web as a source in historical research and how to handle this amount of information in research. Milligan asserts that while research methods long used by historians remain important, the unique nature of web content requires new methodology to be able to extract useful information. Milligan approaches the subject of web archiving as a historian, giving him the perspective of an end user of archives, while also having a strong understanding of the role of archivists and the technical knowledge needed to understand digital born material.

Due to the nature of the subject, much of the investigation of research is relatively new, existing in the internet history and information disciplines. Milligan is very in tune to the research being done in these realms. One particular scholar who has been instrumental in the study of web archives is Neils Brügger, his is a name which comes up again and again when doing research on the subject. Milligan himself places his book in conversation with Brügger in his introduction, “The implications of web archives from an archival and internet studies standpoint have been exhaustively studied by Niels Brügger, who has probably done the most to bring web archive researchers into a cohesive scholarly community.”[[1]](#footnote-1) Brügger’s work is cited by Milligan and both are interested in the digital humanities and web archives but Milligan’s work has solid foundations in his own background with historical research and the changing methodologies of research where web archives are involved. With this work, Milligan expands on the existing literature to discuss the impacts that the use of web archives will have on historical research.

Chapter one, “Exploding the Library,” introduces a somewhat technical history of the internet. Milligan makes it clear from the beginning that one of the new problems with research in web archives is dealing with more information than one person could possibly read, but here he introduces another barrier to entry, the technical understanding required to know what you are looking at and how to interact with archived webpages. His explanations are certainly accessible, especially for those who have a basic understanding of being on the internet as a user, but it’s a lot of information all at once and specific terminology and acronyms can make some of this section feel denser. While you could technically print out a webpage and store it in an archival box, in doing so, you would lose a significant part of the document. To truly understand the content of a webpage, you have to be able to see the links and separate image files. As Milligan says, webpages are “made possible through the interplay of many different files.”[[2]](#footnote-2) A crucial part of understanding a web source is to be able to investigate how it is situated as a group of elements which make up linked pages.

Due to the amount of information to be interpreted from a web archive, Milligan argues that the traditional method of close reading for historical research may not be as effective with web sources. It’s here that he introduces the method of “distant reading” which involves looking at metadata trends in groups of webpages in order to glean information from increasingly large collections. Milligan suggests that it is going to be close reading and distant reading in combination which will be the future of historical study with web archive material. Milligan’s discussion of methodology is rooted in comparison and contrast with traditional archives. He does very well at drawing comparisons between traditional archival practice and web archiving to understand their similarities and differences. This is a tool he employs through his book to much success, especially for those who have an understanding of archival practice, even to a small degree.

Chapter two, “Web Archives and Their Collectors,” is essentially a history of web archiving. Milligan details how web crawlers work to collect webpages and what can get in their way, such as link rot and corporate neglect of user content. He also uses this chapter to explain how web archives are not immune to biases in what and who is being archived. The web can be a democratizing space but certainly still falls victim to accessibilty. Online content creation is more open than traditionally publishing but biases are still present. How web archivists determine what to collect also plays a large role in collections. Because there is so much web content, it can’t all be collected and thus we, or algorithms we produce, must make choices about what is preserved. Once again Milligan draws parallel to traditional archival practice, where due to space, resources, or acquisition guidelines, archivists have to make similar decisions.

Chapter three, “Accessing the Records of Our Lives,” details the practical experience of researching in a web archive. As Milligan explains, web crawlers are not able to capture every webpage at the same time, or even all of the elements within one page, at the same time. For some archived webpages this “temporal shift” can create some strange phenomena. With the complexity of this problem, Milligan does well giving a simple understandable example before scaling it up to convey the research problems these temporal shifts could lead to. He begins with an example of a weather page where the temperatures and forecasts listed on the homepage suggest one kind of weather, while the radar map, captured by the crawler at a different time, shows something entirely different due to being archived by the crawler at a different time. From there, Milligan scales up his example to news sites, where a captured homepage can show differing information than in captured articles as the site may be updated as information changes.

Milligan points to the constantly updating nature of the web as a unique challenge to archiving this kind of material. Even various elements that make up a webpage may be in formats which have become unsupported such as Flash. Flash was a way to have more dynamic websites including animations, now however these elements are no longer viewable, making archiving difficult. Milligan states, “To see what these sites looked like, we will need to resort to magazines and other print sources with screenshots. The surprising durability of print again rears its head.”[[3]](#footnote-3) Milligan points out the value in print material where a digital archive at this point is just not equipped to handle an out of use format. It would probably be easy to write about web archiving and look down on print material, its own problems of storage and degradation, but through this book Milligan makes clear again and again that web archiving is no better and cannot take the place of traditional archives.

Part of Milligan’s argument is that the methods for historical research in web archives are different from that in traditional archives. It is in his third chapter where we get a closer look into what “distant reading” entails. Web archives come with a variety of metadata which can be investigated algorithmically. Investigating web archives in this way can demand different kinds of research questions. Milligan points out examples involving determining relationships between sites through frequency of linking between them to determine relationships. He also details using topic modelling to identify keywords, and what words appear grouped together frequently. This leads to an ease of certain kinds of research questions which can be answered more easily by metadata analysis than close reading. Investigating metadata answers different questions than historians may be used to. In the conclusion of chapter three, Milligan states “A fear in all of our work with web archives, as seen in this chapter, is that the focus on tools and what is technically possible might begin driving research, rather than the research questions driving tools development.”[[4]](#footnote-4) Milligan brings up this fear only in the conclusion of this chapter and it feels a bit dropped on its own. He goes on in the same conclusion to assert, “Of course, historians’ questions have always been framed in part by the availability of sources and the infrastructure available to interpret them. So, turning to a different kind of record does not necessarily represent a dramatic transformation. Yet we still need to consciously avoid letting technology dictate the historical research agenda, especially insidiously.”[[5]](#footnote-5) Milligan’s main argument is about the ways in which historical research will change. Here he very briefly touches on the impacts of those changes on the kinds of research that will be done and the kinds of questions that will get answered, but I think his argument would benefit from a more thorough discussion of this. I would also question that if we must change our research questions because of the scale of the information at our disposal, is the information we’re looking for really accessible?

In the fourth chapter, “Unexpected Needles in Big Haystacks,” Milligan reiterates the importance that researchers understand the technology behind web archives. To look for materials without understanding the algorithms that collect and then return results in a search of them can make it difficult to find things. With such an abundance of material web archives can be difficult to navigate, and unlike traditional archives, there aren’t always institutions or archivists to turn to for guidance on navigation.

Milligan References his case study of GeoCities many times through the first four chapters of his book, and chapter five, “Welcome to GeoCities, Population Seven Million” is where he presents his investigation in full. First Milligan describes GeoCities as any other community or civilization to be studied, “The GeoCities web archive allows us to explore a virtual ruin, and all of the technical and ethical quandaries that such a place presents for historical research.”[[6]](#footnote-6) The ethical concerns in web archiving may appear more apparent than in a physical archive because content created by average everyday people tends to be more accessible online. It’s through the lens of GeoCities, a vast network of personal webpages founded in 1994 and ultimately closed down in 2009, that Milligan discusses the ethics of web archiving material from everyday people and investigating the question of how much privacy we can reasonably expect online. Milligan asserts that often there is not time to debate the ethical dilemma of collection when it comes to web archives. In the case of GeoCities, there was very little time between the announcement of the shutting down of the site and when the site would go offline. Another important point Milligan makes is that web archives often don’t have institutional guidelines to direct them through these tough decisions. An interesting distinction Milligan makes is the difference between what is right legally versus ethically. While there’s nothing legally wrong with using information posted online and citing it in research, there may be an ethical grey area with using content which is clearly personal or had only been meant for close friends or family such as online diaries or conversations amongst a community on message boards. Milligan uses the example of a conversation in a public space such as a café as an example, while there is nothing legally wrong with listening into the conversation of the people at the table next to you, ethically most everyone understands that there is an expected level of privacy there.[[7]](#footnote-7) In response to these ethical dilemmas, Milligan suggests that researchers must be mindful of the content they are using, and to operate off of what the “expectation of privacy” would have been for any given source.[[8]](#footnote-8) Milligan seems to make the argument that compared to a traditional archive, stewarded by archivists who are processing a collection, because of the nature of web archives, in scope and in collection process, the responsibility of ethics falls more on the researcher than those creating the collection.

Chapter six, “The (Practical) Historian in the Age of Big Data, is the final chapter before the conclusion of the book and addresses the state of web archiving and some of Milligan’s thoughts on where we go from here in the field. Through much of the book, Milligan made the choice to avoid talking about specific current technologies, as due to the nature of the subject, things are updated and changed frequently, and the information would become outdated. Here however, he takes the opportunity to expand on the current technology in a more specific way. Web archives don’t have the same kind of infrastructure that traditional archives do, and thus are missing some of the ease of access that come from archivists, librarians, and established practices. Milligan points out that the archiving and collecting of web archives has been made relatively simple thanks to many open-source tools, but the accessing of web archives can still be quite a process. As somewhat of an explanation of his own book, Milligan writes, “During this rapid development, the right place to explain all of this is not in a book, but within blogs, social media, and rapidly evolving conference papers. Even technical journal articles can become dated too quickly. However, the basics of knowing what is inside a collection, text analysis, and network diagrams strike me as core elements that will lie at the heart of web archival research for years to come.”[[9]](#footnote-9) Given his assessment that a book is not necessarily the best format for this subject, in this chapter he offers references to further reading, mostly in the form of links and online resources.

Overall, Milligan gives a very thorough understanding of web archiving and the methodologies for using web archives in historical research. He asserts that web archiving demands some of the same but also many new considerations as compared to traditional archives, both in collection and research. Milligan is very successful in rooting web archival practices as stemming from traditional archival work but with its own new challenges due to the nature of web content. He does well illustrating his points with examples and case studies. Not only does he explain the phenomena he is pointing out, but also point to research questions which could be answered by researching in web archives. While his focus is on changing methodologies, he does mention the potential impact of technology and methodology influencing research questions rather than the other way around. A further discussion of these impacts is perhaps out of the scope of this book, but I think a fuller understanding of the implications of how web archives will affect the kind of historical research we will do is important. Milligan opens the door for this conversation but doesn’t provide much more than some foundational questions. His work is very accessible and a good place to get a comprehensive understanding of web archiving, especially with a basic technological understanding that comes from being online or an understanding of traditional archival practices.

Reference

Milligan, Ian. *History in the Age of Abundance? : How the Web Is Transforming Historical Research*. Montreal: McGill-Queen’s University Press, 2019. <https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2086633&authtype=sso&custid=s8440772&site=eds-live&scope=site>.

1. Ian Milligan, History in the Age of Abundance? : How the Web Is Transforming Historical Research (Montreal: McGill-Queen’s University Press, 2019), 25. [↑](#footnote-ref-1)
2. Milligan, History in the Age of Abundance?, 45. [↑](#footnote-ref-2)
3. Ibid, 118. [↑](#footnote-ref-3)
4. Ibid, 142. [↑](#footnote-ref-4)
5. Ibid, 142. [↑](#footnote-ref-5)
6. Ibid, 171. [↑](#footnote-ref-6)
7. Ibid, 203. [↑](#footnote-ref-7)
8. Ibid, 210. [↑](#footnote-ref-8)
9. Ibid, 232. [↑](#footnote-ref-9)