Appraisal Talk in Web Archives

ED SUMMERS

ABSTRACT The Web is a vast and constantly changing information landscape that by its very nature seems to resist the idea of the archive. But for the last 20 years, archivists and technologists have worked together to build systems for doing just that. While technical infrastructures for performing web archiving have been well studied, surprisingly little is known about the interactions between archivists and these infrastructures. How do archivists decide what to archive from the Web? How do the tools for archiving the Web shape these decisions? This study analyzes a series of ethnographic interviews with web archivists to understand how their decisions about what to archive function as part of a community of practice. It uses critical discourse analysis to examine how the participants' use of language enacts their appraisal decision-making processes. Findings suggest that the politics and positionality of the archive are reflected in the ways that archivists talk about their network of personal and organizational relationships. Appraisal decisions are expressive of the structural relationships of an archives as well as of the archivists' identities, which form during mentoring relationships. Self-reflection acts as a key method for seeing the ways that interviewers and interviewees work together to construct the figured worlds of the web archive. These factors have implications for the ways archivists communicate with each other and interact with the communities that they document. The results help ground the encounter between archival practice and the architecture of the Web.

RÉSUMÉ   Le Web est un paysage informationnel vaste et en changement constant qui, par sa nature même, semble s'opposer à l'idée de l'archive. Pourtant, depuis les vingt dernières années, les archivistes et technologues ont travaillé de concert afin de bâtir des systèmes qui feraient exactement ça. Bien que les infrastructures technologiques pour archiver le Web ont été abondamment étudiées, on en sait étonnamment peu à propos des interactions entre les archivistes et ces infrastructures. Comment les archivistes décident de ce qui sera archivé du Web? Comment les outils d'archivage du Web modèlent leurs décisions? La présente étude analyse une série d'entretiens ethnographique avec des archivistes du Web afin de comprendre comment leurs décisions concernant ce qui doit être archivé s'articulent en fonction d'une communauté de pratique. Elle utilise l'analyse critique du discours pour examiner comment l'utilisation du langage par les participants joue un rôle dans leurs processus de prise de décision d'évaluation. Les résultats suggèrent que les politiques et le positionnement des archives sont reflétés dans la manière dont les archivistes parlent de leurs réseaux de relations personnelles et organisationnelles. Les décisions d'évaluation sont l'expression des relations structurelles d'une archive et des identités de l'archiviste, qui sont forgées au cours des relations de mentorat. L'introspection agit comme méthode essentielle pour voir la façon dont les intervieweurs et les interviewés travaillent de concert pour construire les mondes façonnés des archives du Web. Ces facteurs ont des répercussions sur les façons dont les archivistes communiquent entre eux et interagissent avec les communautés qu'ils documentent. Ces résultats aident à ancrer la rencontre entre la pratique archivistique et l'architecture du Web.

Introduction

Archival appraisal is generally understood to be the work that archivists do to identify materials that have sufficient enduring value to justify their being cared for in an archives.[1] It is typical for archives to have different notions of value, and these are often expressed in the collection development policies that archivists use in the selection work they do while processing collections. Appraisal decisions continue to be made as collections are cared for and as the demands of new records impinge on the archive's ability to store them.[2] While the values ascribed to individual archives differ, the activity of appraisal is central to the work of all archivists. The cumulative effect of these appraisal decisions shapes the historical record and, by extension, our knowledge about the past and our social memory.[3]

This value-driven process of appraisal has many facets, which sometimes can seem to suffuse all of the archivist's work. For example, the values that drive appraisal also find expression in the ways archives are arranged and described, which in turn determine how they are accessed.[4] To describe this moment at the inception of an archive with more specificity, Eric Ketelaar coined the term *archivalization*:

> It is *archivalization*, a neologism which I invented, *meaning the conscious or unconscious choice (determined by social and cultural factors) to consider something worth archiving.* Archivalization precedes archiving. The searchlight of archivalization has to sweep the world for something to light up in the archival sense, before we proceed to register, to record, to inscribe it, in short before we archive it.[5]

---

1   Richard Pearce-Moses, "Appraisal," in *A Glossary of Archival and Records Terminology* (Chicago: Society of American Archivists, 2005), 22–23.

2   Leonard Rapport, "No Grandfather Clause: Reappraising Accessioned Records," *American Archivist* 44, no. 2 (1981): 143–50.

3   Trond Jacobsen, Ricardo L. Punzalan, and Margaret L. Hedstrom, "Invoking 'Collective Memory': Mapping the Emergence of a Concept in Archival Science," *Archival Science* 13, no. 2–3 (2013): 217–51.

4   Elizabeth Yakel, "Archival Representation," *Archival Science* 3, no. 1 (2003): 1–25.

5   Eric Ketelaar, "Tacit Narratives: The Meanings of Archives," *Archival Science* 1, no. 2 (2001): 131–41.

The searchlight of archivalization that Ketelaar imagines for us is the process of deciding what to remember, no matter what material form the archive takes. In this article, I use *archivalization* and *appraisal* somewhat interchangeably, but I use *archivalization* to refer specifically to the initial moments in which a decision is made about what to preserve and what not to preserve. Ketelaar goes on to remind us that "technology changes the archivable."[6] The technologies of record production that we create inevitably shape both *what* and *how* records get archived.[7]

In this article, I will explore how these expressions of archivalization, the specific moments of appraisal, are being performed in web archives in order to gain insight into how the infrastructure of the Web is shaping our attempts to preserve it. I will argue that, while there is a rich research literature concerned with the development and assessment of web archiving technologies, the archival community is only just beginning to investigate how archivists and the tools for web archiving co-produce each other as part of a *socio-technical* system. In order to investigate these moments of archivalization, I will use critical discourse analysis to closely examine how web archivists talk about their appraisal decisions. The article will close with some observations on how the global address space of the Web and the immediacy of its underlying protocols have occasioned a profound shift in the nature of appraisal, particularly with regard to the trust relationship between the documenter and the documented.

Background

As the World Wide Web has become a prominent, if not the predominant, form of global communications and publishing over the last 30 years, we have seen web archiving emerge as an increasingly important activity. The Web is an immense and constantly changing information landscape that fundamentally resists the idea of archiving it all.[8] The Web is also a site for constant breakdowns

6   Ibid.

7   Joan M. Schwartz, "'We Make Our Tools and Our Tools Make Us': Lessons from Photographs for the Practice, Politics, and Poetics of Diplomatics," *Archivaria* 40 (Fall 1995): 40–74.

8   Julien Masanès, "Web Archiving Methods and Approaches: A Comparative Study," *Library Trends* 54, no. 1 (2006): 72–90.

in the form of broken links, failed business models, unsustainable infrastructure, obsolescence, and general neglect. Web archiving projects work through various measures to stem this tide of loss – to save what is deemed worth saving before it is gone. While selecting content, archivists also necessarily allow other archival records to drop out of the line of sight, to degrade and be forgotten.[9] This process of forgetting operates by necessity because archivists are both unable and unwilling to store all records. For example, it is generally understood that institutional archives store between one and five percent of their organizations' records.[10] Electronic records and the Web present similar challenges as our ability to generate data is far outstripping our ability to store it.[11] In all its mundane and messy details, the process of web archiving shapes our knowledge of the recent past in the present.[12]

Not surprisingly, the emergence of the Web and the production of web archives has required the development of new tools, protocols, standards, collaborative networks, and expertise.[13] And so today, the practice of appraisal can no longer be done without the assistance of specialized technologies and automated agents that retrieve selected content for the archives, discover new related content, and provide the archivist with a sense of the dimensions of the entities we call web pages, websites, and domains.[14] While the archival research community has undertaken significant collaborative efforts to survey the tools, techniques, and goals of web archival practices and has developed research methods for studying them, it still has significant work to do in order to understand the day-to-day practices of archivists engaged in appraising web content for archives.

---

9 Verne Harris, "Antonyms of Our Remembering," *Archival Science* 14, no. 3–4 (2014): 215–29.

10 Terry Cook, "'We Are What We Keep; We Keep What We Are': Archival Appraisal Past, Present and Future," *Journal of the Society of Archivists* 32, no. 2 (2011): 173–89.

11 John Gantz and David Reinsel, "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," *IDC iView*, no. 2012 (2012): 1–16.

12 Niels Brügger, "When the Present Web Is Later the Past: Web Historiography, Digital History, and Internet Studies," *Historical Social Research/Historische Sozialforschung*, 2012, 102–17.

13 Masanès, "Web Archiving Methods and Approaches."

14 Ed Summers and Ricardo Punzalan, "Bots, Seeds and People: Web Archives as Infrastructure," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (New York: Association for Computing Machinery, 2017), 821–34, doi: 10.1145/2998181.2998345.

Appraisal Practices

Over the past 100 years, the archival community has developed various sets of appraisal practices that help guide the selection process. These practices have emerged partly as a result of the contingencies of their time: the technologies of writing, printing, photography, and digitization and their respective abilities to generate content at increasingly high rates.[15] The practices are often but not always identified as the practical results of appraisal theories. Indeed, over the last 30 years, appraisal has been guided by a growing critical awareness that the archive both reflects and creatively shapes the society of which it is a part.[16]

Over this same period, archives have also begun to create collection development policies that communicate what types of materials an institution is interested in collecting, for example, materials related by topics, ages, media types, and types of records creators. In a survey of 100 manuscript repositories in the United States, Cynthia Sauer found that many archives still lacked collection development policies, even while evidence suggested that collection development policies were effective tools guiding decisions about donations to archives.[17] While prototypes for these policies have existed since the 1980s,[18] as Jennifer Marshall's content analysis of collection development policies shows, these policies are still geared largely toward potential donors.[19] Collection development policies describe the *what* and the *why* of archival collections but fall short of describing *how* individual collections are to be built.

However, collection development policies are only one example of the types of tools that archivists have deployed for appraising records. Another tool that aids appraisal is the record retention schedule, which codifies when records should be either destroyed or transferred to an archives and provides a timetable for

---

15   Richard J. Cox, "Archivists and Collecting," in *Encyclopedia of Library and Information Sciences*, ed. Marcia Bates and Mary Niles Maack, 3rd ed. (Boca Raton, FL: CRC Press, 2010), 208–20.

16   Cook, "'We Are What We Keep; We Keep What We Are'"; Terry Eastwood, "Towards a Social Theory of Appraisal," in *The Archival Imagination: Essays in Honour of Hugh A. Taylor*, ed. Barbara Lazenby Craig and Hugh A. Taylor (Ottawa: Association of Canadian Archivists, 1992), 71–89.

17   Cynthia Sauer, "Doing the Best We Can? The Use of Collection Development Policies and Cooperative Collecting Activities at Manuscript Repositories," *American Archivist* 64, no. 2 (2001): 308–49.

18   Faye Phillips, "Developing Collecting Policies for Manuscript Collections," *American Archivist* 47, no. 1 (1984): 30–42.

19   Jennifer Marshall, "Toward Common Content: An Analysis of Online College and University Collecting Policies," *American Archivist* 65, no. 2 (2002): 231–56.

managing both the records to be retained and those to be discarded.[20] Sampling techniques may be used to systematically select records from a larger set that cannot be accessioned in full.[21] Macro-appraisal and functional analysis provide frameworks for assessing the value of records based on where and why they were created, their function, and their social impacts – rather than solely on their informational value.[22] Documentation strategy is a process-oriented methodology that allows records creators, archivists, and archives users to collaboratively determine what archives to collect based on the topics, types, and geographic locations of records.[23] Indeed, archival appraisal is a complex topic within archival studies – one that cannot be adequately summarized here.[24] Instead, the remainder of this article investigates the intersection of archival appraisal and practices of web archiving, which will be discussed next.

Web Archives

Starting as early as 1996, the Internet Archive's automated agents started crawling from link to link on the Web, archiving what they could along the way.[25] Not long afterwards, organizations belonging to the International Internet Preservation Consortium (IIPC) began building their own collections of web content. These collections can be either of country domains, like the top-level *.uk* domain collected by the British Library,[26] or of specific websites that have

20   Barbara Nye, "Records Retention Schedules" in Bates and Maack, *Encyclopedia of Library and Information Sciences*, 4465–69.

21   Evelyn Kolish, "Sampling Methodology and its Application: An Illustration of the Tension Between Theory and Practice?," *Archivaria* 38 (Fall 1994): 61–73; Terry Cook, "'Many Are Called But Few Are Chosen': Appraisal Guidelines for Sampling and Selecting Case Files," *Archivaria* 32 (Summer 1991): 25–50.

22   Terry Cook, "Macro-Appraisal and Functional Analysis," *Journal of the Society of Archivists* 25, no. 1 (2004): 5–18.

23   Helen W. Samuels, *Varsity Letters: Documenting Modern Colleges and Universities* (London: Scarecrow Press, 1998); Tom Nesmith, "Documenting Appraisal as a Societal-Archival Process: Theory, Practice, and Ethics in the Wake of Helen Willa Samuels," in *Controlling the Past: Documenting Society and Institutions: Essays in Honor of Helen Willa Samuels*, ed. Terry Cook (Chicago: Society of American Archivists, 2011), 31–50.

24   For a recent overview of the historical development and evolution of appraisal thinking in archives, see Fiorella Foscarini, "Archival Appraisal in Four Paradigms," in *Currents of Archival Thinking*, 2nd ed., ed. Heather MacNeil and Terry Eastwood (Santa Barbara, CA: Libraries Unlimited, 2017), 107–34.

25   Gordon Mohr, Michael Stack, Igor Ranitovic, Dan Avery, and Michele Kimpton, "An Introduction to Heritrix: An Open Source Archival Quality Web Crawler," in *4th International Web Archiving Workshop*, 2004, accessed January 18, 2020, https://webarchive.jira.com/wiki/download/attachments/5441/Mohr-et-al-2004.pdf.

26   Daniel Gomes, Sérgio Freitas, and Mário J. Silva, "Design and Selection Criteria for a National Web Archive," in *Research and Advanced Technology for Digital Libraries: Proceedings of 10th European Conference, ECDL*

been selected according to a collection development policy.[27] In the last 5 to 10 years, service providers such as Archive-It, Hanzo Archives, Perma, Archive-Social, and MirrorWeb have emerged that allow archives in academic institutions, government, and business to create their own web archives for historical purposes as well as for compliance with e-discovery and other regulations. Web archiving research is now routinely presented at conferences such as the International Conference on Digital Preservation (iPRES), the ACM/IEEE Joint Conference on Digital Libraries (JCDL), the ACM Conference on Web Science (WebSci), and events of the Research Infrastructure for the Study of Archived Web Materials (RESAW) and the Coalition for Networked Information (CNI). While an exhaustive overview of this literature is outside the scope of this article,[28] a brief discussion of research relevant to the study of appraisal in web archives is warranted.

Previous research has measured the number and size of web archives[29] as well as the contributions of automated web archiving agents.[30] There has also been some analysis of how collection development policies at major national

---

2006, ed. Julio Gonzalo, Costantino Thanos, M. Felisa Verdejo, and Rafael C. Carrasco (Berlin: Springer-Verlag, 2006), 196–207; Margaret E. Phillips, "What Should We Preserve? The Question for Heritage Libraries in a Digital World," *Library Trends* 54, no. 1 (2006): 57–71.

27   Steven M. Schneider, Kirsten Foot, Michele Kimpton, and Gina Jones, "Building Thematic Web Collections: Challenges and Experiences from the September 11 Web Archive and the Election 2002 Web Archive," in *Proceedings of the 3rd Workshop on Web Archives, in Conjunction with the 7th European Conference on Research and Advanced Technologies for Digital Libraries*, ECDL 2003, ed. Julien Masanès, Andreas Rauber, and Gregory Cobena (n.p.: n.p., 2003), 77–94, accessed January 18, 2020, http://bibnum.bnf.fr/ecdl/2003 /proceedings.php?f=schneider.

28   While the topic of web archives is spread across the literature on digital libraries, digital preservation, digital curation, and archives, a good overview of the history, theory, and practice of web archiving can be found in Julien Masanès, ed., *Web Archiving* (Berlin: Springer, 2008). For more recent coverage of the various types of web archiving efforts, with a specific focus on the use of web archives from a research perspective, see Niels Brügger and Ralph Schroeder eds., *The Web as History: Using Web Archives to Understand the Past and the Present* (London: UCL Press, 2017).

29   Daniel Gomes, João Miranda, and Miguel Costa, "A Survey on Web Archiving Initiatives," in *Research and Advanced Technology for Digital Libraries*, ed. Stefan Gradmann, Francesca Borri, Carlo Meghini, and Heiko Schuldt (Berlin: Springer, 2011), 408–20.

30   Kalev Leetaru, "How Much of the Internet Does the Wayback Machine Really Archive?" *Forbes*, November 16, 2015, accessed June 3, 2019, http://www.forbes.com/sites/kalevleetaru/2015/11/16/how-much-of-the-internet -does-the-wayback-machine-really-archive/; Kalev Leetaru, "Why It's So Important to Understand What's in Our Web Archives," *Forbes*, November 25, 2015, accessed June 3, 2019, http://www.forbes.com/sites /kalevleetaru/2015/11/25/why-its-so-important-to-understand-whats-in-our-web-archives/.

archives enact appraisal[31] and some general overviews of archival processes in web archives.[32] Indeed, there is no shortage of research material about the need for web archiving or the technical approaches for achieving it.

Surveys of the web archiving community have been conducted by Jennifer Marill, Andy Boyko, and Michael Ashenfelder, for the IIPC,[33] and more recently by the National Digital Stewardship Alliance (NDSA).[34] These surveys provide useful longitudinal descriptive statistics about web archiving activity, particularly with regard to the tools used for web archiving. The 2011 and 2013 NDSA surveys provide a high-level overview of how organizations manage, fund, and train web archiving teams. The 2013 survey contains the first questions about the types of content being collected, for example, social media, databases, videos, audio content, blogs, and art. How these categories were determined is not entirely clear, but this approach marks a significant shift away from emphasizing the technologies being used and toward understanding the appraisal dimensions of web archiving. Questions about inter-institutional, collaborative web archiving in both the 2011 and 2013 surveys indicate a growing interest and engagement with co-operative approaches. However, apart from a short series of interviews with web archivists involved in the selection of content, published by the Library of Congress,[35] we have surprisingly few in-depth studies of how this

31   Jinfang Niu, "Appraisal and Custody of Electronic Records: Findings from Four National Archives," *Archival Issues* 34, no. 2 (2012): 117–30.

32   Jinfang Niu, "An Overview of Web Archiving," *D-Lib Magazine* 18, no. 3–4 (2012), accessed January 18, 2020, http://www.dlib.org/dlib/march12/niu/03niu1.html.

33   Jennifer Marill, Andrew Boyko, Michael Ashenfelder, and Gina Jones, "Web Harvesting Survey" (n.p.: International Internet Preservation Consortium, 2004), accessed January 18, 2020, https://web.archive.org/web/20150413181742/http://www.netpreserve.org/sites/default/files/resources/WebArchivingSurvey.pdf.

34   National Digital Stewardship Alliance, *Web Archiving Survey Report* (n.p.: National Digital Stewardship Alliance, 2012), accessed January 18, 2020, http://www.digitalpreservation.gov/documents/ndsa_web_archiving_survey_report_2012.pdf; Jefferson Bailey, Abigail Grotke, Kristine Hanna, Cathy Hartman, Edward McCain, Christie Moffatt, and Nicholas Taylor, *Web Archiving in the United States: A 2013 Survey* (n.p.: National Digital Stewardship Alliance, 2014), accessed January 18, 2020, http://www.digitalpreservation.gov//documents/NDSA_USWebArchivingSurvey_2013.pdf; Jefferson Bailey, Abigail Grotke, Edward McCain, Christie Moffatt, and Nicholas Taylor, *Results of a Survey of Organizations Preserving Web Content* (n.p.: National Digital Stewardship Alliance, 2017), accessed January 18, 2020, http://ndsa.org/documents/WebArchivingintheUnitedStates_A2016Survey.pdf; Matthew Farrell, Edward McCain, Maria Praetzellis, Grace Thomas, and Paige Walker, "Web Archiving in the United States: A 2017 Survey" (n.p.: National Digital Stewardship Alliance, 2018), accessed January 18, 2020, https://osf.io/m8wzr/.

35   Kimberly D. Anderson, "Appraisal Learning Networks: How University Archivists Learn to Appraise Through Social Interaction" (PhD dissertation, University of California, Los Angeles, 2011); Abbie Grotke, "Ask the

work is enacted and achieved by archivists. Some notable examples of such work are discussed in the next section.

Appraisal in Web Archives

Recent work by Emily Maemura et al.[36] explores the provenance of web archives, and in particular, the types of information that can be provided in archival description to ensure that researchers can understand the various conditions that gave rise to an archive. In their case study, Maemura et al. analyzed the creation over time of three web collections and developed a conceptual framework to guide the documentation of provenance for web archives. This framework identifies three essential elements of provenance documentation for web archives: the scope of collecting, the processes of collecting, and the context within which these processes take place. The authors' key insight is that web archiving is a socio-technical phenomenon that happens in time, amidst shifting social and historical conditions. Understanding the socio-technical conditions that give rise to a web archive requires a socio-technical perspective – that is, a perspective that examines the ways that human agency and technologies co-produce each other but avoids privileging either the social or technical aspects as determining the other – and is essential for researchers who attempt to interpret the records that the archive contains.

The element from Maemura et al.'s framework that is most relevant to our analysis of appraisal is the scope of collecting – which helps especially in gaining insight into the specific moments of archivalization. Maemura et al. itemize several useful technical factors (i.e., timing, access restrictions, and tool configurations) that appear to influence the archivalization of web content. However, while provenance documentation is primarily concerned with the chain of

Recommending Officer: The Civil War Sesquicentennial Web Archive," *The Signal* (blog), Library of Congress, August 15, 2011, accessed January 18, 2020, http://blogs.loc.gov/digitalpreservation/2011/08/ask-the -recommending-officer-the-civil-war-sesquicentennial-web-archive/; Abbie Grotke, "Ask the Recommending Officer: Indian General Elections 2009 Web Archive," *The Signal* (blog), Library of Congress, January 31, 2012, accessed January 18, 2020, http://blogs.loc.gov/digitalpreservation/2012/01/ask-the-recommending-officer -indian-general-elections-2009-web-archive/; Michael Neubert, "Five Questions for Will Elsbury, Project Leader for the Election 2014 Web Archive," *The Signal* (blog), Library of Congress, October 26, 2014, accessed January 18, 2020, http://blogs.loc.gov/digitalpreservation/2014/10/five-questions-for-will-elsbury-project -leader-for-the-election-2014-web-archive/.

36   Emily Maemura, Nicholas Worby, Ian Milligan, and Christoph Becker, "If These Crawls Could Talk: Studying and Documenting Web Archives Provenance," *Journal of the Association for Information Science and Technology* 69, no. 10 (2018): 1223–33.

custody of a particular set of records and with determining where the records came from and establishing their authenticity, appraisal decisions shape not only what is present in the archives but also what is missing.[37] While there have been significant changes that would encourage archival description to speak directly to the conditions surrounding both appraisal and the moments of archivalization,[38] these conditions have not typically been central concerns of provenance, which tracks the particular material and genealogical histories of a set of records. Maemura et al. conclude that more empirical work is necessary to understanding the actual work practices of archivists, and this is particularly true with regard to understanding the articulation of appraisal decisions.

In terms of the work practices of web archivists, Jessica Ogden, Susan Halford, and Leslie Carr[39] conducted an ethnographic study of labour practices at the largest public web archives, the Internet Archive. Ethnography provided both method-ological and theoretical instruments for integrating observation, interviews, and documentary analysis into an investigation of the day-to-day activities of web archivists at work at the Internet Archive. Specifically, the researchers employed the methodology of trace ethnography to examine how human and non-human agents worked together as part of a socio-technical system.[40] This approach allowed them to explore the materiality of web archiving by discussing both the relationship between this practice and the production of artifacts (both digital and analog) and the role of the environmental factors (the policies, activities, infra-structure, and communities) that actively inform practice and production.

Through its decades of web archiving work, the Internet Archive has developed a heterogeneous set of work practices that help guide its web archiving efforts. These practices include manual inspection and maintenance of web crawlers; specialized analysis of collected content (to discover candidates for archiving);

---

37 Rodney G.S. Carter, "Of Things Said and Unsaid: Power, Archival Silences, and Power in Silence," *Archivaria* 61 (Spring 2006): 215–33.

38 Tom Hyry, Diane Kaplan, and Christine Weideman, "'Though This Be Madness, Yet There Is Method In't': Assessing the Value of Faculty Papers and Defining a Collecting Policy," *American Archivist* 65, no. 1 (2002): 56–69; Jennifer Douglas, "Toward More Honest Description," *American Archivist* 79, no. 1 (2016): 26–55.

39 Jessica Ogden, Susan Halford, and Leslie Carr, "Observing Web Archives: The Case for an Ethnographic Study of Web Archiving," in *Proceedings of WebSci '17* (Troy, NY: Association of Computing Machinery, 2017), accessed January 18, 2020, https://eprints.soton.ac.uk/410123/.

40 R. Stuart Geiger and David Ribes, "Trace Ethnography: Following Coordination through Documentary Practices," in *44th Hawaii International Conference on System Sciences* (n.p.: IEEE, 2011), 1–10, accessed January 18, 2020, http://www.stuartgeiger.com/trace-ethnography-hicss-geiger-ribes.pdf.

and the use of custom tools (to prioritize and/or block particular web domains) and activity streams from platforms such as WordPress, Wikipedia, and Twitter (to transform the "labour of the crowds"[41] into appraisal metrics). While these practices provide significant evidence of the ways appraisal is being enacted in web archives, the size, history, and collecting scope of the Internet Archive when compared to most other web archiving organizations make it an anomaly. Consequently, there are still significant questions about how archivists in less specialized workplaces enact web archivalization.

In a thematic analysis of ethnographic interviews with web archivists, Summers and Punzalan explored the various dimensions of decision-making that impact how web content is appraised.[42] This study revealed six areas that impacted the work of appraisal in web archives: crawl modalities, information structures, resources, people, tools, and breakdown. Echoing Maemura et al. and Ogden et al., Summers and Punzalan also highlighted the importance of attending to the actual work practices of archivists as they interacted with each other and with non-human actors such as software tools, systems, and infrastructures.[43] While these thematic findings, based on an analysis of ethnographic field notes, were useful, questions remained about the specific nature of the work practices themselves, particularly with regard to emerging communities of practice for web archivists.

Given this background, it is evident that while there are well established appraisal practices in archives, the archival field is still in the process of adjusting these practices to the new medium of the Web. This situation is complicated by the fact that, even though the core standards and protocols of the Web have remained fairly constant,[44] their deployment and articulation have evolved rapidly in the context of rising social media and platform logics.[45] As a

---

41 Ogden et al., "Observing Web Archives," 8.

42 Summers and Punzalan, "Bots, Seeds and People."

43 A socio-technical approach to web archiving is also present in Anat Ben-David and Adam Amram, "The Internet Archive and the Socio-Technical Construction of Historical Facts," *Internet Histories* 2, no. 1–2 (2018): 179–201. Ben-David and Amram look specifically at the ways algorithmic practices impact the contents of archives and the types of epistemological claims that are made using them.

44 Ian Jacobs and Norm Walsh, eds., "Architecture of the World Wide Web, Volume One" (n.p.: World Wide Web Consortium, 2004), accessed January 18, 2020, http://www.w3.org/TR/webarch/.

45 Tarleton Gillespie, "The Politics of 'Platforms,'" *New Media & Society* 12, no. 3 (2010): 347–64; José van Dijck, *The Culture of Connectivity: A Critical History of Social Media* (Oxford: Oxford University Press, 2013).

result, the field of web archiving is situated within a much larger socio-technical system that has been undergoing rapid change unlike any we have seen since the invention of print. The practice of archival appraisal is now jointly enacted by archivists, designers, software developers, and the automated agents that are embodied in the infrastructures that co-produce the archive.[46]

Critical Archival Studies

The archival research community's growing interest in the political role of the archive also plays a significant role in discussions about appraisal and archivalization. Archives have traditionally been built and sustained through an investment of resources that are normally reserved for established institutions that can represent powerful interests.[47] Awareness of this has brought sustained attention to the role of archival appraisal as both a site for social memory[48] and a vehicle for social justice.[49] The seemingly mundane details of documenting selection decisions take on a distinctly political dimension that should be addressed when discussing how web content is selected for archives. Recent work on critical archival studies establishes a theoretical stance for such an exploration, which started in the 1990s as part of the postmodern turn in archival studies.[50] As Caswell, Punzalan, and Sangwand explain, "Critical archival studies broadens the field's scope beyond an inward, practice-centered orientation and builds a critical stance regarding the role of archives in the production of knowledge and different types of narratives, as well as

---

46   Summers and Punzalan, "Bots, Seeds and People." For a broader overview explaining that archival appraisal is not solely the concern of archivists but always involves the work of records creators, and records users in a much wider records ecosystem, see Sue McKemmish, Frank Upward, and Barbara Reed, "Records Continuum Model," in Bates and Maack, *Encyclopedia of Library and Information Sciences*, 4447–59.

47   F. Gerald Ham, "The Archival Edge," *American Archivist* 38, no. 1 (1975): 5–13; Howard Zinn, "Secrecy, Archives, and the Public Interest," *Midwestern Archivist* 2, no. 2 (1977): 14–26; Cook, "'We Are What We Keep; We Keep What We Are.'"

48   Brien Brothman, "The Past That Archives Keep: Memory, History, and the Preservation of Archival Records," *Archivaria* 51 (Spring 2001): 48–80.

49   Randall C. Jimerson, *Archives Power: Memory, Accountability, and Social Justice* (Chicago: Society of American Archivists, 2009); David A. Wallace, "Locating Agency: Interdisciplinary Perspectives on Professional Ethics and Archival Morality," *Journal of Information Ethics* 19, no. 1 (2010): 172–89; Ricardo L. Punzalan and Michelle Caswell, "Critical Directions for Archival Approaches to Social Justice," *Library Quarterly* 86, no. 1 (2016): 25–42.

50   See Terry Cook, "Archival Science and Postmodernism: New Formulations for Old Concepts," *Archival Science* 1, no. 1 (2001).

identity construction."[51] By applying a critical archival studies lens, my study was able to reveal some of the motives and architectural tensions behind the observed practices.

For the archival community to define a set of roles, policies, or technologies for appraising web content, it is important to understand how appraisal is currently enacted in web archives. Accordingly, I formulated a number of questions to guide my research: What can this enactment of appraisal practices tell us about the role of web archives today in relation to the archives we have and those we do not have? Is there an emerging community of practice[52] around the appraisal of web content? If so, what are its characteristics and dynamics? Answering these questions can help to inform the design of appraisal tools (policies and technologies) and can also provide guidance to practitioners as they collectively grapple with the challenges that these new forms of media present for archives and the organizations of which they are a part.

On the one hand, critical archival studies reminds us that we must not lose sight of the forest for the trees when examining web archiving practice up close, as this article attempts to do. Without a theoretical grounding like critical archival studies, it would be easy to slip into simply recounting individual practices without drawing them together to form a coherent analysis of appraisal in web archives. On the other hand, critical archival studies connects my research question concerning appraisal practice in web archiving work with the methodology chosen for my study, to which I turn next.

Methodology

One way of investigating the phenomenon of archivalization is to qualitatively analyze how archivists talk about their appraisal work: to look at the words they use, the conventions they have established, the context they share, the ways they

---

51   Michelle Caswell, Ricardo Punzalan, and T-Kay Sangwand, "Critical Archival Studies: An Introduction," *Journal of Critical Library and Information Studies* 1, no. 2 (2017).

52   See Jean Lave and Etienne Wenger, *Situated Learning: Legitimate Peripheral Participation* (Cambridge, UK: Cambridge University Press, 1991) for the original development of the concept of community of practice (CoP); for the later development of this concept, see Etienne Wenger, *Communities of Practice: Learning, Meaning, and Identity* (Cambridge, UK: Cambridge University Press, 1998). As discussed in the conclusion of this article, CoP's attention to practice as a site for learning provides a useful theoretical perspective for this article's analysis.

learn from each other in communities of practice, and the political work that these communicative practices perform.[53]

To address these research objectives, I undertook a critical discourse analysis (CDA) of a set of ethnographic interviews of individuals involved in selecting web content for archives. I employed CDA because it offers a theoretical framework, grounded in critical theory, for analyzing the way in which participants' use of language reflects identity formation, figured worlds, and social relations, while also addressing the larger socio-cultural context in which practice takes place. CDA helps to examine how language use connects with issues of ideology and power, which are particularly relevant when considering archival appraisal as an inherently political act, in line with a critical archival studies' perspective. Sociolinguist James Paul Gee, a practitioner and theorist of CDA, noted that "there are solid linguistic, even grammatical grounds, on which to argue that all language-in-interaction is inherently political."[54] Indeed, CDA is a theoretical approach to language use rather than a method as such, and those who practice CDA bring a variety of discourse analysis methods to bear in their analyses.[55]

Gee's research centres on the fields of education, literacy, and media studies. This focus makes his work particularly relevant for analyzing the ways archivists talk about web archives. He elucidates seven *building tasks* that language performs to reflect and produce social relations. These building tasks involve

- significance: how language is used to foreground and background certain things;

- activities: how language is used to enact particular activities;

- identity: how language is used to position specific identities and make them recognizable;

---

[53] Wenger, *Communities of Practice*.

[54] James Paul Gee, "Discourse Analysis: What Makes It Critical?" in *An Introduction to Critical Discourse Analysis in Education*, ed. Rebecca Rogers (London: Routledge, 2004), 49–80.

[55] Ruth Wodak and Michael Meyer, eds., *Methods of Critical Discourse Analysis* (London: Sage, 2001).

- relationships: how language is used to construct relationships between people and things;

- politics: how notions of value and norms are established in the use of language;

- connections: how language is used to connect and disconnect ideas, activities, and objects; and

- sign systems and knowledge: how language positions particular sign systems, or ways of knowing and believing.[56]

Gee also provides a set of methodological tools that support the analysis of linguistic performances, or building tasks – in other words, tools that are used to dissect the ways that language produces social relations. For Gee, words do actual work in the world: "Whenever we speak or write, we always and simultaneously build one of seven things or seven areas of 'reality.'"[57] While I draw on several of Gee's building tasks and tools in my analysis, as I immersed myself in my transcription data, I became particularly focused on the building tasks related to *relationships*, *identity*, and *politics*.

While language is important, it is not the only means by which archivists build community in their work. CDA also allows the researcher to examine language use in relation to non-linguistic elements such as technology, infrastructure, and setting. Although this study focuses specifically on linguistic discourse, software utilities, infrastructures, and the geographic dispersion of Internet communication provide important dimensions for understanding the work of appraisal in web archives. Some of these factors emerge below in the discussion of the results.

---

56 James Paul Gee, *How to Do Discourse Analysis: A Toolkit*, 2nd ed. (London: Routledge, 2014), 95–98.

57 Ibid., 94.

Data Collection

I specifically chose ethnographic interviews as a data collection method in order to provide a glimpse into participants' views of their own work.[58] My own participation as a technologist in the web archiving community facilitated recruitment because I was familiar with several of the interview subjects and had an understanding of their work as archivists. It was therefore essential to make self-reflection a key component of the analysis; field notes gathered during the interview process provided a means for doing this reflective work.[59]

An initial set of participants was selected using purposeful sampling from a list of attendees at a web archives conference I attended in Ann Arbor, Michigan, on November 12–13, 2015. Snowball sampling was employed to widen the number of informants beyond this initial set. Stratified sampling was used to also include technologists, volunteers, and researchers who were involved in web archiving practices. As this conference skewed toward North American participants, my research findings reflect the practices and views of that community of practitioners. In order to compensate for the large number of users of Archive-It, currently the leading provider of web archiving services in the United States, I also recruited customers of other web archiving service providers, such as Hanzo and ArchiveSocial, and members of the Archive Team community.

The study recruited 39 individuals (21 female, 18 male), 27 (13 female, 14 male) of whom agreed to be interviewed. Due to technical problems, 4 interviews were not recorded, which resulted in the study collecting audio recordings and transcripts of interviews with 23 individuals (12 female, 11 male).

Each interview lasted approximately an hour and was allowed to develop organically as a conversation. Informants were encouraged to describe their work in web archives and how they had come to it. After this general introduction and discussion, the conversation developed through follow-on questions about the informants' work and history. The goal of the interviews was to invite participants to describe specific situations where they needed to decide whether or not to archive web content.

---

58   Lucy Suchman, "Making Work Visible," *Communications of the ACM* 38, no. 9 (1995): 56–64.

59   Robert M. Emerson, Rachel I. Fretz, and Linda L. Shaw, *Writing Ethnographic Fieldnotes*, 2nd ed. (Chicago: University of Chicago Press, 2011).

Interviews were conducted via Skype, and video and audio recordings were then transcribed using the Scribie transcription service. Each participant provided informed consent via email. Participants were located across the United States, so in-person interviews were impractical. While some participants in the study wished to be named, others did not, so pseudonyms are used in place of all personal and organizational names in the excerpts below.

A set of codes derived from an inductive thematic analysis of the field notes for these interviews was used to analyze the transcripts of the interviews themselves. This set of codes was chosen specifically to highlight the correspondence between the findings from the field notes and the language used by participants in the interviews.[60]

Coding the interview transcripts in this way allowed for deep immersion in the interview data, which yielded targeted insights into specific interview segments that spoke particularly well to issues related to communities of practice and appraisal. The coded transcripts were used to identify key participants and specific segments to further analyze using Gee's seven building tasks. Returning to the audio recordings of these interviews, performing closer transcription, and attending closely to the ways that specific words were chosen and enunciated provided the core data used in the discussion of findings below.[61] All the transcriptions included in this article use the following notation:

TABLE 1   *Notations used in interview transcripts*

//	final intonation contour, like a period in writing
/	non-final intonation contour, like a comma
[segment]	overlapping talk
=	latching: two utterances that follow one another without any perceptible pause
WORD	an emphasized or stressed word

60   The results of this phase of the data collection and analysis are described in Summers and Punzalan, "Bots, Seeds and People."

61   While there is no standard set of transcription notations, the critical discourse analysis (CDA) methodology focuses attention not only on what was said, but also on how it was said.

In addition to Gee's seven building tasks, I used several of his specific tools for discourse analysis; these are mentioned in the context of the findings they helped illuminate.[62]

Findings

The findings draw from interviews with three subjects, who I refer to pseudonymously as Jim, Jack, and Carly. These excerpts, from the 30 hours of interviews coded for analysis, have been selected specifically because of the way these informants talked about their appraisal work and the manner in which they involved other people in their decision-making processes.

Gee's seven building tasks, which allow language users to shape social realities, were chosen to provide a framework for insights into how and why appraisal in web archives is being performed – not to quantify or otherwise make generalized claims about the practice as a whole. As I performed close readings of the transcripts, Gee's building task related to relationships appeared particularly useful for examining the participants' use of language. Gee explains this building task by associating it with his relationships building tool, which is used to analyze relationships found in language: "For any communication, ask how words and various grammatical devices are being used to build and sustain or change relationships of various sorts among the speaker, other people, social groups, cultures and/or institutions."[63]

Focusing on relationships present in the archivists' use of language in turn exposed two more of Gee's building tasks: those related to identities and politics. Regarding identities, Gee advises researchers to "ask what socially recognizable identity or identities the speaker is trying to enact or to get others to recognize."[64] And regarding politics, Gee is primarily concerned with how language performs the distribution of "social goods" or enacts day-to-day politics rather than with formal systems of government. The discussion of the findings below is organized around three themes that emerged when using the questions Gee

---

[62]   The entirety of Gee's *How to Do Discourse Analysis* details a collection of methodological tools that help in the analysis of language using critical discourse analysis. Some of these tools are linked directly to his idea of seven building tasks, but others are linked to ideas about saying, doing and designing, situated meaning, social meaning, and intertextuality.

[63]   Gee, *How to Do Discourse Analysis: A Toolkit*, 121.

[64]   Ibid., 116.

suggests in relation to these three building tasks: the themes of hierarchies, mentorship, and structures.

Since discourse analysis is not used a great deal in archival studies, a brief note on the excerpts is warranted before diving into the specific examples. The excerpts included below are not typical quotations because they aim to capture the stresses, pauses, and rhythms of speech using the aforementioned notation. The lines are sparse, numbered, and spatially organized to reflect the individual utterances that form the recognizable speech units that will be analyzed. The length of the excerpts varies depending on the amount of context that is useful for the discussion. The excerpts used here are not intended to be representative of archivists in general but are used to provide insights into particular factors that are at play in the appraisal of web content.

Hierarchies

Jim works as an archivist in a non-profit organization that does a large amount of web archiving as well as digitization. In the following excerpt, Jim describes a situation where a large amount of data was being archived from a video streaming provider that was going out of business. I was asking Jim to recall a time when he had needed to make a decision about whether or not to archive a particular website or document. The excerpt provides a particularly salient snapshot of the type of discussion that goes on in his organization when archivists decide how to archive a large amount of content from the Web:

Line #	Speaker	Utterance
1	Jim	The petty disk /
2		that I have at my disposal /
3		without having to go to higher ups /
4		is about 10 terabytes //
5	Ed	mm-hmm //
6	Jim	So if I find a job and the job is like an eight-terabyte job /
7		I don't need to bring it up with the Archives //
8		Um, if I discover that it's gonna be 30 40 50 petabytes /

9		I go to my superior /
10		Ariana Reese /
11		and Ariana initially will go /
12		"Why? /
13		Convince me this is a good idea." //
14	Ed	mm-hmm
15	Jim	And I'll say, "Well, it's cause of this this." She'll go, "Okay that sounds like a good idea. Go ahead." /
16		you know, like as a stopgap //
17		But then it will be like /
18		Well it's extremely controversial /
19		it's the stuff=and then she'll be like /
20		"Okay well /
21		if it could possibly blow back on the Archives /
22		or if we could potentially be facing some kind of issue with it /
23		let's go have a chat with Greg" /
24		and now it's me and Greg and Ariana saying /
25		"Are we gonna do this?" //
26	Ed	mm-hmm
27	Jim	Now when I mention these /
28		I mention them like it's some sort of whatever /
29		you know we're talking /
30		I'd have to say that that's me and Greg and Ariana going /
31		whether this job /
32		has happened like /
33		six times in the last three years /
34	Ed	mmm
35	Jim	maybe? /
36		Ooom [and Ariana]
37	Ed	[and did] /
38		was Real TV /
39		was this one of the examples? /

40		the Real TV one or?
41	Jim	Well Greg helped me get in contact with the employees /
42		Greg was already on the ground with it.
43	Ed	Oh okay //
44	Jim	and Greg /
45		KNEW /
46		that it was going to be a lot of data /
47		and was like /
48		"Okay so [be a little more] /
49	Ed	[ahhhh]
50	Jim	careful with this"

As I asked Jim to recall a specific occasion when he had needed to make an appraisal decision, he recalled a situation that emphasized relationships with other employees in his organization. Note the use in line 3 of "higher ups," which invokes the concept of an organizational hierarchy or chain of command that involved his manager, Ariana, and Greg, the director of the archives. These relationships are foregrounded and frame the decision that is being made. Invoking the organizational hierarchy in this manner lends weight and formality to the appraisal decision, while also working to lead us away from a discussion of the appraisal criteria. The moment of archivalization is surfaced and then effaced.

Just as the hierarchy is emphasized, the details of the actual decision-making process are elided with rapid speech and the use of "this this" in line 15 to refer to the actual appraisal criteria. We do not actually know what Jim says to Ariana to persuade her that the video content is worth saving: the "this" references are stressed through repetition, but they both lack a referent. Jim indicates that these conversations are infrequent and that the initial decision to archive this content came directly from Greg. The decision to archive this content started at the top, came down, and then went back to the top of this appraisal ladder again. The circularity and vagueness of these hierarchical relationships suggests that they could be operating as a rhetorical device to formalize what is otherwise a much less structured and more organic process.

One additional relationship that is identified only near the end of this segment is the connection between Greg and Jim, inside the archives, and an

employee outside at the video streaming organization. Archives often refer to these connections as donor relationships, because they broker communications between individuals or organizations that are donating materials and the receiving archives. Greg is described as helping Jim contact an employee and as already being "on the ground" with the process, which casts the archives in the role of actively seeking content and not simply receiving content. This is a highly significant donor relationship, which we will return to below.

Mentorship

Jack is an archivist at a large university in the United States, which he joined a few years ago after leaving a previous job as an archivist at another university. In this segment, Jack describes how he came to work on a web archive that documents the activities of the fracking industry.

Line #	Speaker	Utterance
1	Jack	I came to University A actually wanting to /
2		to drive some more um /
3		I guess professional engagement around the legacy of fracking in this state //
4		I came from the University B where /
5		we had a lot of really intense collections around environmentalism /
6		and energy development in the state /
7		and it was a sort of an area of programming for the archives there //
8		And one of my close colleagues there had done a project /
9		basically sort of like anticipating the next energy boom in the state /
10		which coincidentally was fracking //
11		which also coincidentally was um something that 30 or 40 years ago /
12		a company wanted to um /
13		explore by /

14		detonating a series of underground nuclear explosions to
15		stimulate the gas um /
16		to the surface [chuckle] //
17	Ed	Whaaat?
18	Jack	And we had some collections on on that /
19		sort of like the protest effort that um /
20		that killed that endeavour //
21		So she had done just a lot of really interesting stuff=
22		oral histories exploring the boom and bust and /
23		so I always kind of had in the back of my mind that coming /
24		to University A /
25		I had other reasons for coming /
26		but that in coming here
27		this was an issue I wanted to explore //
28		But I didn't really have an opportunity to push it /
29		until I saw / some news announcement somewhere /
30		I don't know where /
31		but University C announcing that they were /
32		going to start this project /
33		to document the fracking anti-fracking activism in the state //
34		And I immediately took it to our associate dean /
35		who at the time was Mark Dalton //

Again, we see that relationships figure prominently in this description of how decisions are made about what to archive from the Web. In lines 34–35, a hierarchical relationship between Jack and Mark is positioned as one of the key moments of appraisal, as we saw in the previous example. We also see several organizational relationships traced between Jack's current university (A), his

previous university (B), and an institutional collaborator (C). However, unlike in the previous example, these are not hierarchical relationships but links of influence and practice.

The first relationship, between organization A and organization B, is mediated by Jack's own professional history. He worked at both of these organizations and mentions them in order to highlight a specific mentoring relationship between himself and a colleague at University B, who had done significant archival work around documenting fracking. His description of this relationship as "close" and as a source of inspiration lacks the clinical tone used in descriptions of the previous hierarchical relationships. This mentoring connection knits archival practices for oral histories together with those for websites, using the shared interest in documenting environmental issues and activism.

There is also a relationship between Jack and an individual at University C; their shared interest in documenting fracking activated Jack's ability to begin work on the collection and also became the seed of a collaboration. It is important to note the implicit role that the Web plays in this collaboration. The distributed, globally accessible information space of the Web means that Jack and his collaborator at University C needed to partition their work geographically. Unlike physical collections, which can be in only one place at one time, the public Web is available to everyone who has a computer and an Internet connection. Jack is not stymied by University C's move to document fracking but is emboldened to participate. While a request for approval again moves up the organizational hierarchy, the initial impetus (archivalization) comes laterally, from a peer at another institution, and from the past in the form of his mentor. Rather than being a discrete event, the moment of archivalization actually involves an assemblage of actors removed in both time and space.

In the following excerpt, we continue to look at mentoring relationships as evidence of an emergent community of practice including archivists and web archivists. In this segment, we hear from Carly, who explains that she became involved in web archiving while working as an archivist for over a decade at several large research universities, where she spent a significant amount of time performing web archiving.

Line #	Speaker	Utterance
1	Carly	Yeah, that's kind of how I've always /
2		Back in the day when I first started in GovDocs /
3		one of my mentors /
4		she was a local docs librarian at University D /
5		and her approach /
6		and I feel I would love to figure out a way to do this better /
7		So she actually just got the three major newspapers in the county area /
8		and she=we would pile them up for her /
9		and she would just take like a day a week /
10		and she would go through and CLIP /
11		the news articles /
12		and then she would make sure she got /
13		the documents that were mentioned um /
14		in them //

Here, Carly is connecting her practice in appraising web content with her professional experience of working as a government documents librarian. Carly specifically uses "one of my mentors" to emphasize that she learned from a specific individual – and the plural form identifies this person as one of several mentors she has learned from during her career. References to these mentoring relationships suggest that archivalization draws on a network of learning through a community of practice. Carly's attention to specific details, such as the "three major newspapers," how her mentor "would pile them up" – as well as the stressed "clip" – recall the physical process of doing the work. The confusion in subjects in the "she=we" points to Carly's reconstruction of the scene for this appraisal work and indicates that she was one of several people working together as part of a team. It is also apparent from the repair in line 6 that Carly feels that this material process does not have a direct analogue in her current web archiving work (even though she goes on to talk about her use of email discussion lists, RSS feeds, and bookmarks later in the interview) and that it could be useful to find one.

Structures

There is something else going on in Carly's discourse, which may be apparent only to an outsider to the library and archives profession. To draw this out, it can be useful to follow the guidance in Gee's making strange tool: "For any communication, try to act as if you are an 'outsider.' Ask yourself: What would someone (perhaps even a Martian) find strange here (unclear, confusing, worth questioning) if that person did not share the knowledge and assumptions and make the inferences that render the communication so natural and taken-for-granted by insiders?"[65]

Specifically, the reference to "getting started in GovDocs" (line 2) and the expression "local docs librarian" (line 4) speak to a particular type of work that is not necessarily directly tied to the work of web archivists. Government documents librarians are trained librarians who focus on collecting, preserving, and providing access to documents published by federal, state, and local governments. This type of work came about in the United States after the establishment of the Federal Depository Library Program and the Government Printing Office by the *Printing Act of 1895*. As such, it is highly regulated work that is guided by policy. The work of scanning the "three major newspapers" and looking for references to "documents" was being done in the context of this highly politicized activity. How were the major newspapers selected, and what factors influenced their selection? What government documents – material output of governmental activities that present a view of society from the perspective of the state – were librarians looking for? This context for the newspaper clipping and note taking is also the experience that guides Carly as she decides what to archive from the Web. The scanning of newspapers for references to government documents is a precise moment in the process of archivalization.

We see this same political aspect at work more explicitly in this final excerpt from Jack, who is reflecting on his work to document fracking.

---

65   Gee, *How to Do Discourse Analysis: A Toolkit*, 19.

Line #	Speaker	Utterance
1	Jack	I really see like one of / my next curatorial responsibilities being um /
2		not really more crawling or more selecting /
3		but using the connections I've made here /
4		to get more contact and more dialogue going with um /
5		with the actual communities I've been documenting //
6		And I'm a little nervous about how it's gonna go /
7		because I went ahead and crawled a bunch of stuff /
8		without really doing that in advance //
9		I'm also a little nervous about it because /
10		through our biology librarian I did try to talk to um /
11		I did try to get more of the local expertise involved um /
12		in helping us scope out you know sites to crawl //
13		But the way she always sort of implemented that was /
14		she ended up setting us up with some of the very people who /
15		I think these activist groups feel were complicit in what's gone on since then //
16		And one of them I distinctly recall /
17		told me straight up in our meeting that he doesn't think I should be crawling these activist groups
18	Ed	Really?
19	Jack	Because he doesn't find them credible //

Using Gee's making strange tool again to uncover the context of this excerpt draws our attention to the use of the phrases "crawling and selecting" (line 2) and "scope out . . . sites to crawl" (line 12). "Crawling" refers to the behaviour of software used to collect content from the Web. This software is traditionally referred to as a *spider* because it automatically and recursively follows links in web content for an amount of time that is determined by the scope it is given. However, the software needs to be told where to start crawling. Jack sought advice from local domain experts (lines 10–11) in determining where the software should begin crawling and for how long, but he also indicates that he is planning to do more work with the activist communities he is documenting. Jack reflects that activist communities may be concerned with how he has selected content for the archive. He also significantly discusses credibility, in lines 16–19, as a measure of what should (and should not) be in the archive. The web archive is shown to be a contested political space.

We also see several relationships being teased out here: the relationship between Jack and other members of the university community; the relationship between Jack and the community of activists who are working to stop fracking in the state; the relationship between Jack and the software that is performing the archiving activity; and finally, the relationship between Jack and the interviewer (myself) as I orient to his description of how the credibility of the fracking archive had been called into question (line 18). The two communities, of activists and of the university, are presented as being at odds; but elsewhere in the interview, Jack talks about an overlap between them (university members who are also activists like himself). Tracing this network of political agendas and associations is tied up in the work of selecting which websites to archive and is part of the figured world that Jack and I are building in this interview. Doing the work of crawling the Web and appraising web content inscribes these political dynamics into the archive and presents an opportunity to reflect on what they are.

Discussion

Gee's seven building tasks provided a clarifying lens for studying the discourse that emerged from these interviews. Specifically, Gee's relationships building tool helped trace the connections between archivists, their colleagues, their institutions, and the creators of web content out in the world. These relationships mark

pathways of mutual engagement and illuminate how appraisal decisions, or acts of archivalization, are made as part of a community of practice.

Carly and Jack's comments about mentoring relationships are of particular interest because they present historical relationships that extended in time, rather than relationships anchored in individual workplaces. Both archivists drew on mentoring experiences when speaking about how they made appraisal decisions. It is interesting that, in both cases, the experiences involved not web archiving but the archiving of physical documents and oral histories. The appraisal decisions were oriented around the purpose or function of the archives as political agents rather than around the specifics of the Web as a medium. The archivists engaged in translation work to map their experience with archiving print material to help them make determinations about whether to archive web content.

The inherent political dimension to these relationships was another feature that emerged from the discourse. Hierarchical relationships within the organizations operated to buttress appraisal or validate appraisal decisions, but we also saw significant relationships between the archivists and the communities or individuals that were being documented. In Carly's case, this relationship was embedded in the professional discourse around government documents work and the operations of government. In the United States, government documents are produced within a legal framework where they are considered part of public discourse and the public domain.

In Jim and Jack's cases, there was an awareness of a need for more interaction with the creators of the documents being archived. The role of the archivist in relation to those being documented forms part of a complex terrain that the archivist must navigate in doing web archiving work. The organizational and community relationships intersected with each other to generate productive and destabilizing effects. Developing practices archivists can follow as they go about making these connections with content creators on the Web, or as Jack says, getting more "contact and dialogue" with content creators, is marked as a potential area for further methodological and design work, especially with regard to establishing trust relationships on the Web.[66]

---

66   Kathryn M. Neal, "Cultivating Diversity: The Donor Connection," *Collection Management* 27, no. 2 (2002): 33–42; Karen F. Gracy, "Documenting Communities of Practice: Making the Case for Archival Ethnography," *Archival Science* 4, no. 3–4 (2004): 335–65.

The approach of tracing relationships, exhibited here, has much in common with the methods offered by actor network theory (ANT),[67] with the important distinction that non-human actors are not part of this discussion. ANT suggests that one way to extend or enrich this work would be to explore how artifacts such as policies, software tools, standards, and services fit into this network of relationships and how they figure into moments of translation in the work of web archiving.

In addition, the concept of a community of practice provides guidance for mapping the interactions and practices of web archiving work. Wenger describes a community of practice as a process of "negotiated meaning" that is achieved through the participation of its members in some joint enterprise. But participation is not the whole story; negotiated meaning is also dependent on something Wenger calls *reification*, which he describes as "the process of giving form to our experience by producing objects that congeal this experience into 'thingness.' In so doing we create points of focus around which the negotiation of meaning becomes organized."[68]

Wenger's idea of reification and its role in building communities of practice suggest that consideration of the artifacts of web archiving could enrich this picture of the relationships involved in appraisal work in web archives. This move also nicely parallels Gee's inclusion of non-linguistic elements into the analytical scope provided by CDA. Recall Carly's memorable description of the work she did with her mentor to pile up newspapers and clip articles that held hints or clues about documents they needed to track down for the archive. Activities like this are examples of a shared repertoire that knits together participative and reifying elements of a community of practice. It is important to remember that Carly felt that she lacked an analogue to this practice in her work with web archives. Moving beyond the interview and into participant observation in the context of a case study is one way of exploring this gap.

---

67   John Law, "Actor Network Theory and Material Semiotics," in *The New Blackwell Companion to Social Theory*, ed. Bryan S. Turner (Oxford: Wiley-Blackwell, 2009), 141–58.

68   Wenger, *Communities of Practice*, 58.

Conclusion

Considering the political and the artifactual dimensions of *appraisal talk* suggests that an architectural shift has taken place in the movement from archiving physical media such as documents, photographs, and disks to archiving networked resources such as web pages, websites, and web platforms. Physical media require some form of hand-off, where an archives gains possession of material either through donation or by some other means. This often entails significant work by the archivist, who is often involved in the physical transfer of materials and the negotiation of a deed of gift that serves as a contract between the archives and the individual or organization that currently owns the material.

The architecture of the Web dissolves this traditional relationship because the content can be immediately acquired using the Internet and the Hypertext Transfer Protocol (HTTP). Additionally, web archiving software allows materials to be rapidly collected in bulk, often without any interaction with the publisher or owner of the content. On the Web, the idea that records will be transferred to the archives when they are no longer actively used no longer applies because it is precisely at the moment when records are removed from the Web that they become unavailable to the web archivist, at least with our current set of tools and practices. This revolution in record transfer technologies suggests an inversion of current web archiving practice and a realignment of traditional donor relations, in which web publishers reach out to web archives to have their websites collected prior to them being turned off.

The way web archivists talk about their appraisal processes shows that, despite their relative isolation, they work within dynamic and distributed communities of practice that are extensions of a longer trajectory of appraisal in archives. And yet, at the same time, the architecture of the Web and its affordances for access have disrupted the traditional relationship of trust between the donor and the archives. Access to the appropriate tools grants the archivist the ability to easily collect content for the archives with very little interaction with the content owner. This means it is more important than ever to consider the positionality of the archives in relation to the documented entity when deciding what to archive on the Web.[69] It also suggests that there are opportunities for bridging this gap by becoming participating members of the communities we document, including

---

69   Jimerson, *Archives Power*.

them in our communities of practice, and developing tools and strategies that help us establish these connections.

> Appraisal brings into sharpest focus the power wielded by archivists – the power of what the French philosopher Jacques Derrida calls consignation. Which stories will be consigned to the archive and which will not. This power of the storyteller is ultimately a political power. Which is why, in a democracy, society must find ways of holding archivists accountable for their appraisal decisions.[70]

As Verne Harris indicates here, accountability is an ever-important dimension to the work of an archives. But engagement in a community of practice that includes content creators as dynamic and complex participants presents challenges for the archivist who works with the Web. What are the pathways of trust in web archives? How do we enact and map them? While these have been perennial challenges for the archival community, they are placed into stark relief in web archives because of the modes of acquisition that often involve the record creator only minimally, if at all. Finally, echoing Maemura's point about the importance of documenting provenance in web archives, we must recognize these moments of archivalization as necessary elements of archival practice on the Web. Yet, while provenance looks backwards in time to reconstruct relationships between records and the world, appraisal looks forwards to actively construct them. Appraisal in web archives is charged with an architectural tension, as the Web's access protocols and global namespace collapse expected relations between archivists and records creators. The next step for researchers and archivists working in and with web archives must be to examine how current tools and practices can mend this architectural divide and to establish a social web of trust that determines how a particular set of records ends up in a web archives.

---

70   Verne Harris, "Postmodernism and Archival Appraisal: Seven Theses," *South African Archives Journal* 40 (1998): 48–50.

BIOGRAPHY   Ed Summers is research faculty at the Maryland Institute for Technology in the Humanities. He is also a PhD candidate in the College for Information Studies at the University of Maryland. Ed's research interests currently centre on issues related to archives and memory as they intersect with science and technology studies. Previously, Ed worked for eight years as a software developer in the Repository Development Center at the Library of Congress, where he helped develop the National Digital Newspaper Program and the Linked Data Service for publishing name and subject authority data on the Web. Prior to that, he worked at Old Dominion University, the University of Illinois, and Columbia University, where he focused on the use of metadata in information services.