Appraisal Practices in Web Archives

Ed Summers

Contents

Abstract **3**

I. Introduction **5**

II. Literature Review **10**

Appraisal 10

Archival Meta-theories 11

Governmentality 13

Tacit Appraisal 14

Realizing Appraisal 16

Documenting Society 19

Appraisal Critique 21

Strategy 22

Decentralization 24

Outside the Archive 25

Appraising Appraisal 26

Appraisal and Web Archives 27

Science and Technology Studies 29

Digital Libraries 30

Archival Studies . . . 32
Appraisal Practice . . . 34
Ethnography of Infrastructure . . . 36
Repair and Maintenance . . . 37
Software Studies . . . 38
Data Justice . . . 41

III. Methodology **43**
Algorithms as Culture . . . 44
Studying Up . . . 46
Relations All the Way Down . . . 48
Two Projects . . . 49
Appraisal Talk . . . 50
Implementing Appraisal . . . 51
Design . . . 53

IV. References **54**

V. Appendixes **74**
Study 1 Documents . . . 74
Consent Form . . . 74
Recruitment Letter . . . 77
Interview Protocol . . . 78
Study 2 Documents . . . 81
Consent Form . . . 81
Recruitment Letter . . . 84
Interview Protocol . . . 85

Abstract

The web is a site of constant breakdown in the form of broken links, failed business models, unsustainable infrastructure, obsolescence and general neglect. Some estimate that about a quarter of all links break every 7 years, and even within highly curated regions of the web, such as scholarly publishing, rates of link rot can be as high as 50%. Over the past twenty years web archiving projects at cultural heritage organizations have worked to stem this tide of loss. However, we still understand quite little about the diversity of actors involved in web archiving, and how content is selected for web archives. This is due in large part to how web archiving projects operate out of sight as complex sociotechnical assemblages at the boundaries between human and automated processes.

This dissertation explores appraisal practices in web archives from the perspective of Science and Technology Studies in order to answer two motivating research questions. 1) How is appraisal currently being enacted in web archives? 2) How do definitions of what constitutes a web archive relate to the practice of appraisal? Answering these questions will help better understand the dynamics that shape our memory of the past that is mediated by the web, and inform pedagogy in archival studies. Critical Algorithm and Data Studies provide a theoretical framework for examining how web archiving systems function as both computational and cultural objects that participate in a wide variety of social and political projects. As machine learners increasingly become readers of web archives the stakes for understanding the dynamics by which these collections are built could not be higher.

Interviews with web archives practitioners, and a year long field study at a government agency involved in archiving data from the web, will provide rich descriptive material for an ethnographic analysis of what current practices are, and how these practices are shaped by the ontological dimensions of web archives. Critical Discourse Analysis will be used to analyze interview transcripts and field notes from participant observation in meetings and in the work of the archive. In addition, methods drawn from Trace Ethnography will be used to analyze version control histories, and

ticketing systems that have been used to coordinate and assemble a web archive that challenges our notions of what constitutes a web archive.

I. Introduction

For a moment try to imagine your day to day life without the web and the underlying Internet. It is difficult to do because the web is a communications infrastructure that is completely enmeshed in global systems of capital, governance, science and culture. However, somewhat paradoxically, we experience the loss of the web on a daily basis, every time we click a link only to get a *404 Not Found* error. The architectural precarity of our "World Wide Web" is constantly being made visible to us. These quotidian breakdowns are so common that we have come to expect, or even anticipate them. The continual loss of the web and its broken links become part of the infrastructural background in the metaphor of the constantly changing cloud (Hu, 2015). But what remains of this backgrounding, or evaporation, is an archival anxiety that asks: What will we remember of our current epoch? Are we really living in what will be known as a *digital dark age*? (Hedstrom, 1991; Kuny, 1997) Or perhaps we are living in the ruins of a digital dark age right now?

Partly in response to this archival anxiety we have witnessed the emergence of *web archiving* as an increasingly significant activity. Web archiving is the practice of collecting content from the web for preservation, which is then made accessible at another part of the web known as a *web archive*. Web archiving is typically preformed by memory institutions (libraries, archives and museums), but increasingly it is individuals who take up the work themselves (Lee, 2011). Developing record keeping practices for selecting, collecting and preserving web content is proving to be an extremely important endeavor for historical production (Brügger, 2017) and for sustaining the networked public sphere (Benkler, 2006 ; Lessig, 2006). Archivists use the term *appraisal* to talk about the theories and methods that determine what records are collected in an archive. However, even with close to two decades of practice we understand surprisingly little about the day to day processes by which content is being appraised for web archives.

At the same time, our experience of using the web and the Internet is suffused with constant, and often invisible, data collection. These data flows pool into vast corporate and government data archives that have come to be referred to in shorthand as *Big Data*. For example, in 2015 Facebook

was collecting two billion photographs a day from its users, which at that time required hundreds of petabytes of storage (Bandaru & Patiejunas, 2015). Always-on mobile computing technologies, the *Internet of Things* and *smart cities* provide the infrastructure for a host of data capture and processing platforms that have become an essential parts of our daily lives. Hoskins (2018) calls these new data archives, and their attendant processual flows *shadow archives*:

> The archive has traditionally been seen (like other media) as separate and external to the self, as something with institutional status, as variously a place and space for the storage of artefacts of the past that give rise to remembering. Yet, the medial gathering and splintering of individual, social and cultural imaginaries, increasingly networked through sortable and pervasive digital media and communication devices, attach shadow archives to much of everyday life, that also blend and complicate that which was once considered as distinctly public and private. (p. 87)

Shadow archives are physically and conceptually remote, are often beyond our individual control, cognition, and are usually only readable in full by the entities that created them. These archives surface on the web in various ways, but are characterized not by an anxiety about what will be remembered, but rather by what will not be forgotten (Mayer-Schönberger, 2011).

Most importantly for my discussion here, these shadow archives are constructed both *in* and *of* the web, and operate as web archives just as much as the previously mentioned web archives operated by cultural heritage organizations. They are *shadow* archives in a second sense, in that they sit behind or to the side of normalized conceptions of what *web archives* are, as a specific deployment of software, hardware and networked infrastructure. In addition to manifesting as *Big Data* web archives can take the shape of thematically arranged websites (Fenlon, 2017) or *Small Data* (Abreu & Acker, 2013). Questions of what to collect, what not to collect, what to remember, and what to forget, are sites of controversy and anxiety, that surface on the web, and are encoded by the standards, protocols and infrastructure. How is it that our idea of what constitutes a *web archive* have become stabilized, and how does this stabilization relate to our decisions about what to archive?

In this paper I will explore the art and science of deciding what web archives collect (how they appraise) using the theoretical lens of Science and Technology Studies (STS). I will suggest that our anxieties around what web archives remember and forget, can be understood by attending to the specific material practices of people working with memory and machines. I will argue that it is useful to understand how these material practices operate within, in response to, or outside of conceptions of the archive as an instrument of *governmentality* (Foucault, Davidson, & Burchell, 2008 ; Lemke, 2019). But first, let me provide a thumbnail sketch of these areas of archival appraisal and practice and how they relate to web archiving.

Deciding what to keep, and what to discard, is a central theme in the field of archival studies– some even say it is the archivist's first and most important responsibility (Cox & Samuels, 1988) or "the most significant archival function" (Brichford, 1977). Over the past two centuries archival theorists have developed a body of literature around the concept of *appraisal*, which is broadly defined by the Society of American Archivists as the "process of identifying materials offered to an archives that have sufficient value to be accessioned". Document production, which began with the innovation of writing, and rapidly accelerated with the publishing technologies of moveable type, the printing press, photography, lithography, xerography, and computer automation has made it increasingly necessary for archivists to recognize their pivotal role in deciding what documents get to be called *archival records*.

As a practical matter, for an archive to exist, appraisal decisions must be made, which necessarily shape the archive over time, and by extension also shape our knowledge of the past (Bearman, 1989 ; Cook, 2011). It is in the particular contingencies of the historical moment that the archive is created, sustained and used (Booms, 1987 ; Harris, 2002). And yet the desire for a technology that will enable a complete archival record of the web, where everything is preserved and remembered, is a strangely persistent idea, or aspirational goal, with many social and political ramifications (Brothman, 2001 ; Mayer-Schönberger, 2011). Reviewing this literature of appraisal, with an eye to understanding the appraisal of content on the web is the first focus of this paper.

Part of the reason for the gap in our understanding about how web content is being selected for preservation is a matter of scale. Considered as a whole the web is an unfathomably large, decentralized and constantly changing information landscape. Unlike a box of photographs in an attic, that may find its way into a Hollinger box on a shelf in an archive, web content seems to come to us from *The Cloud* (Hu, 2015), and appears to resist the archival imagination that has traditionally focused on *information as thing* (Buckland, 1991).

The web is a site of constant breakdown in the form of broken links, failed business models, unsustainable infrastructure, obsolescence and general neglect. Ceglowski (2011) has estimated that about a quarter of all links break every 7 years. Even within highly curated regions of the web, such as scholarly publishing (Sanderson, Phillips, & Sompel, 2011) and jurisprudence (Zittrain, Albert, & Lessig, 2014) rates of link rot can be as high as 50%. Web archiving projects work in varying measures to stem this tide of loss: to save what is deemed worth saving before it becomes 404 Not Found. In this light, web archiving can be seen as a form of repair or maintenance work (Graham & Thrift, 2007 ; Steven J. Jackson, 2014) that is conducted by archivists, collaborating with each other, while also deeply engaged with tools and infrastructures and platforms that aid them in their work. Attention to issues of repair and maintenance and the larger field of Science and Technology Studies (STS) suggests an examination of web archiving as a set of material practices that includes activities such as website design, upgrades, storage backups, and the porting of content from one content management system to another. I will review how this lens of repair and maintenance helps us think about web archiving as *practice* forms the second part of my analysis.

The construction of web archives, and the maintenance of the web, entail each other, and present challenges and opportunities for archivists as they work with content creators, systems, services and other bespoke software. It is important that our knowledge of these systems be informed by an analysis of the social, technical and material practices by which web content is selected for an archive. Decisions about what to collect from the web in a web archive are *co-produced* by the technical means that are used to enact them (Jasanoff, 2006). Thus how appraisal is enacted in

web archives is fundamentally both an epistemological and an ontological question. How do web archives generate facts, evidence and knowledge? How does our idea of what constitutes a web archive and the record (Yeo, 2007 , 2008) shape that knowledge? Appraisal practices for the web manifest at the interface layer, which is itself a fractal of the infrastructure of the web itself. I will conclude the paper by outlining a research agenda for web archives that opens up from an analysis of appraisal, practice and the web. But first, let's examine what we know about the concept of archival appraisal and how it relates to web archives.

II. Literature Review

Appraisal

Generally speaking the field of archival studies is praxis oriented, in that it engages with issues of theory in the context of practice. The specific literature of archival appraisal is a prime example of this praxis orientation. The problem of appraisal is fundamentally concerned with the practical problem of how to select material for an archive given 1) an abundance of records, and 2) a finite amount of resources to store all of them. Cox & Samuels (1988) consider appraisal to be the "first responsibility" of the archivist, and define it broadly as:

> … any selection activity that enables archivists to identify recorded information that has enduring value, primarily for the documentation of modern society (p. 29)

In a recent review of the appraisal literature Anderson (2011) (p. 26) distinguishes between *appraisal* and *selection*, in order to provide clarity about when records are evaluated (appraisal) versus when they are chosen (selected), and when these activities take place relative to an archive taking custody of the records (accession). However for the purposes of this paper a more expansive, and generalist, view on appraisal is taken, which admits that appraisal is a process by which values are asserted about records. These assertions happen in multiple intersecting timelines and at varying scales that take shape due to repeated, atomic actions of selection. This more general view will be important later when appraisal activities are considered in light of the sociomaterial dimensions of the web.

While archives have existed for millenia (Posner, 1972), it is only over the past several hundred years that archivists have developed the concept of appraisal in order to manage the ever increasing deluge of records, that has been brought upon them, largely by the technologies of record production. Three years before the web was first prototyped by Tim Berners-Lee at CERN, Young (1985) at the Bentley Historical Library reviewed the literature of archival appraisal in the United States and found 178 monographs, articles, reports and unpublished manuals. A search in 2018 for *appraisal*

after 1985 in Library and Information Science Source yields some 300 more. A complete and exhaustive survey of this literature while possibly beneficial, does not serve my purpose here, which is to connect the literature of archival appraisal with practices of web archiving. To do this it is useful to examine meta-theories, or conceptual frameworks that have been used to talk about archival appraisal.

Archival Meta-theories

Eastwood (2002) outlines three strains of thinking about archival appraisal that have developed over the past few centuries of archival studies. The first and most commonly held view, is that appraisal is a vehicle for history: we must choose what to remember from the perspective of the future historian who is attempting to understand today (Schellenberg, 1956). The second view is that appraisal (the shaping of the archival record by archivists) is not a task for the archivist because it is the enemy of authenticity, evidence and the ultimately the archive itself (Duranti, 1994; Jenkinson, 1922). The third is that appraisal is an inherently political act, that necessarily carries with it the threat of erasure, while also providing opportunities for social justice, ethical engagement, and as sites for creative record creation (Harris, 2002 ; Punzalan & Caswell, 2016).

Eastwood's characterization of appraisal holds up well today, since it casts the three strands not as an evolution in time, but as a bricolage of approaches that coexist and function simultaneously. However it is important to note that Eastwood discusses these three strands of appraisal against a backdrop of Western democratic societies. Eastwood's thesis is that appraisal practices and theories are put to work in the service of democracy, and are to be understood ultimately as a tool for governance and accountability. This is a theme that I will return too shortly. But before I do that I want to examine another appraisal meta-theory.

In her recent study of appraisal practices Foscarini (2017) draws on the work of Cook (2013) to describe the discursive tensions that exist in the research literature about archival appraisal. Cook identified a general timeline of archival paradigms:

> I want to suggest that since the later nineteenth century, archival identity has shifted, or has been in the process of shifting, through four such paradigms or frameworks or mindsets, as it has struggled, and still struggles, with this memory-evidence tension. I am calling the four frameworks: evidence, memory, identity, and community. It is important to emphasize that these four accumulate across time; they do not entirely replace each other.

Foscarini takes up these four paradigms to examine how they manifest in various theoretical approaches to appraisal, which largely orient around what can be seen as top-down and bottom-up approaches. In top-down approaches (evidence and memory) the archivist intervenes as little as possible in the service of authenticity and integrity (Jenkinson, 1922), or they analyze organizational hierarchies, structures and activities in an attempt to document society as a whole (Cook, 2004). In the bottom-up approach (identity and community) appraisal is recognized as a creative activity, in which history, memory and social relations are assembled as part of a complex set of activities that are not necessarily centered on an institutional context. Foscarini goes on to note that these ideas about appraisal are not steps along a timeline, but are all still very much with us. For example, much of the last few decades of work in digital preservation have been focused on the design and development of technologies for ensuring authenticity and integrity of data, with the implicit and governing assumption that technology can, or should, allow us to perfectly collect everything.

In this paper I am offering my own meta-theory of appraisal specifically for web archives, which relies on Foucault's idea of governmentality as a framework for understanding the full scope of appraisal theories, in their social and historical contexts (Foucault, 1991). So before diving into these details it is important to first take a look at *governmentality*, which we will return to later as an example of a practice orientation to web archives.

Governmentality

The records from our earliest archives happen to be coterminous with the beginnings of recorded human history. This is no coincidence given that the methods of history depend in large part on the evidence of the past that survives, and this evidence is often found in archival repositories of various kinds. As such, archives are often seen as historiographical instruments. The disciplines of history and archival studies are twinned, but their relations are somewhat antagonistic. Historians are one of the most active users of archives. The records collected in archives are used by historians as evidence of previous events and activities, which are essential for piecing together and anchoring historical narratives in fact, or as close as can be got to fact many years later. Until the professionalization of archival studies, it was common for archivists to be trained historians, with invested knowledge of the events, people and organizations that the archive is ostensibly concerned with. Who better to tend to the records of an organization than the scholar who knows its history? It wasn't until archivists began to meaningfully grapple with the concept of *appraisal* that an understanding of the archive developed that was more than simply the tool of memory and history (Taylor, 1984), but also as an political and economic instrument of power, accountability, resistance and liberation (Jimerson, 2009).

The archival practices of southern Mesopotamia and ancient Egypt, which archival studies points to for its origin story (at least in Western traditions), are examples from the earliest known states used archives to manage records of laws, administrative activities, financial transactions, land ownership, and taxation (O'Toole, 2004; Posner, 1972). These archives reflected the interests of the earliest states in governing and controlling their populations. Similarly the gaps in records, and their silences, work to demarcate those on the periphery of the state, or outside it altogether (Scott, 2017). A significant portion of the stories of archives are thus bound up with the interests of institutions, states, governance and what Foucault (1991) calls *governmentality*.

For Foucault governmentality is a mode of rationality of governing through technologies of power, that reproduce themselves through specific practices. Governmentality is not simply the story of

the operations of the state, but is concerned with "the conduct of conduct", and can be found in the practices of individuals, households, families and communities. Foucault traces the emergence of the art of government and political economy, and situates it with the decline of the sovereign, the birth of the modern state, and more recently, neo-liberalism. Governmentality is a useful instrument because it provides a continuous field that includes many modes of archival production and appraisal. It provides a frame for looking at appraisal landscapes as a form of knowledge (savoir), that has particular rules of formation, transformation and correlation with other practices. As noted by Schwartz & Cook (2002):

> Archives have always been about power, whether it is the power of the state, the church, the corporation, the family, the public, or the individual. Archives have the power to privilege and to marginalize. They can be a tool of hegemony; they can be a tool of resistance. They both reflect and constitute power relations. They are a product of society's need for information, and the abundance and circulation of documents reflects the importance placed on information in society. They are the basis for and validation of the stories we tell ourselves, the story-telling narratives that give cohesion and meaning to individuals, groups and societies. (p. 13)

Foucault's idea of governmentality is useful for tracing practices of archival appraisal, or the practices of deciding what to remember and forget, because of the insights it brings into how appraisal practices function as part of the sociotechnical assemblages of archives and their manifestations in the web. But before turning to look at web archives I want to briefly outline a genealogy of appraisal practices that foregrounds governmentality.

Tacit Appraisal

Despite the millennial historical arc of archival practice, Western archival studies often locates its theoretical origins in the concept of *respect des fonds*, or more commonly, *provenance* from 19th century France. Respect des fonds is a practice of grouping records by their creator, rather than by

subject matter or some other predetermined taxonomy. In practice this means records are grouped together by the organization, family, or individual that created them. As Bailey (2013) describes, the introduction of provenance occurred in the aftermath of the French Revolution, which saw the simultaneous destruction and reconstruction of records; a truly monumental attempt to both erase and refashion the historical record in the newly established Archives Nationales. While attempting to deal with this project the librarian, archivist and historian Natalis de Wailly introduced the idea of *respect des fonds* in 1841 as a simplified practice for arranging the records:

> The principle was, in fact, a practical exigency, a method to simplify archival arrangement seen as more easily implemented by novice archivists than the more rigorous arrangement by classification. Bailey (2013)

It is significant to note that this use of provenance as a heuristic for organizing records was born amidst what was effectively a massive appraisal process, in which overtly political interests determined the preservation of pre and post-revolution records. However, the concept of appraisal was not itself explicitly part of the expression of respect des fonds. The question of what to keep and what to discard was subsumed into a practice for processing the pre-revolutionary materials.

Despite its partial application in France, respect des fonds proved popular as it spread to Belgium and the Netherlands, where it became part of the *Manual for the Arrangement and Description of Archives* (Muller, Feith, & Fruin, 1940), that was originally published in 1898. The so called *Dutch Manual* added to to the concept of respect des fonds an additional principle named *original order*, which stated that the arrangement of records should be the same as the original organization. In their manual the authors Muller, Feith and Fruin describe how records are created during the activities of administrative bodies or officials that operate using the metaphor of a life-form:

> … an archival collection is an organic whole, a living organism which grows, takes shape, and undergoes changes in accordance with fixed rules. If the functions of the body change, the nature of the archival collection changes likewise. The rules which govern the composition, the arrangement and the formation of an archival collection,

> therefore, cannot be fixed by the archivist in advance; he can only study the organism and ascertain the rules under which it was formed. (p. 19)

Even here the idea of appraisal, or the decision of what to keep and what to discard, is tacit. The concept of original order works to prevent the reordering or disruption of records that have *already* arrived as *archival*. The decision of what administrative bodies and officials to collect has already been made, and the principle of original order works to govern how the records are to be arranged and described once they arrive. The surprisingly fresh conception of records as both product and part of a form of life (organicity), still works to bracket off appraisal as something that has already happened out of band. The fact that an administrative body is generating the records is enough to transmute the documents from mere papers into *archival records* that are to be preserved for the long term. The process of governance is at work in the very conception of how the archive functions.

Realizing Appraisal

In the early 20th century, Hillary Jenkinson imported the concept of respect des fonds from continental Europe and fused it with existing English archival practices in his influential *Manual of Archival Administration*. Jenkinson was a medievalist by training, and stressed the importance of impartiality, authenticity and naturalness (similar to organicity) in archival practice. These three principles coordinate to position the archivist as the *keeper* of records, and proscribes the archivist from acting in any way to shape was is made archival. While Jenkinson admits that decisions need to be made about what to keep, these decisions are made by the record creator, before the records are added to an archive, and not by the archivist after the fact. More recently Duranti (1994) connects ideas around authenticity to the theory of *diplomatics*, the critical and forensic analysis of documents, which she traces back to the practice of Roman law in the 11th century. Notice here that the locus of appraisal is still in the hands of officials working within administrative bodies, in the bureaucracies of power.

Tschan (2002) suggests that much of the last century of archival thinking can be characterized as a sustained conversation between Jenkinson's Manual on the one hand, where archival appraisal is verboten and authenticity is paramount, and another manual *Modern Archives: Principles and Techniques* written by historian and archivist Theodore Schellenberg. Schellenberg's manual was written after World War Two, in a moment when record production was vastly outstripping the ability to store them. In his position at the US National Archives and Records Administration, Schellenberg recognized that pragmatic decisions needed to be made about what records to make *archival*, and that those decisions were essentially assertions about value, of which there were two kinds: primary and secondary.

Primary value was the value of the records to the record creator, in their active use. Secondary value on the other hand divided into two subtypes: *evidential* and *informational*. On the surface *evidential value* seems to be quite similar in principle to Jenkinson's interest in impartiality, authenticity and naturalness. However Schellberg makes clear that he is not concerned with the fidelity of the records as evidence, but with how well the records function as evidence of the organization and function of a particular government body:

> By evidential value I do not refer here to the value that inheres in public records because of the merit of the evidence they contain. I do not refer, in a Jenkinsonian sense, to the sanctity of the evidence in archives that is derived from "unbroken custody." I refer rather, and quite arbitrarily, to a value that depends on the importance of the matter evidenced, i.e. the organization and functioning of the agency that produced the records.

Schellenberg outlined a variety of criteria to use for assessing the evidentiary value of records, which crucially links the volume of records with organizational hierarchy:

In contrast, *information value* is determined by the archivist in a subjective way, that takes into account the historical moment that the records were created in, and often involves outside consultation with relevant subject matter specialists. Schellenberg, himself trained as a historian, notes that

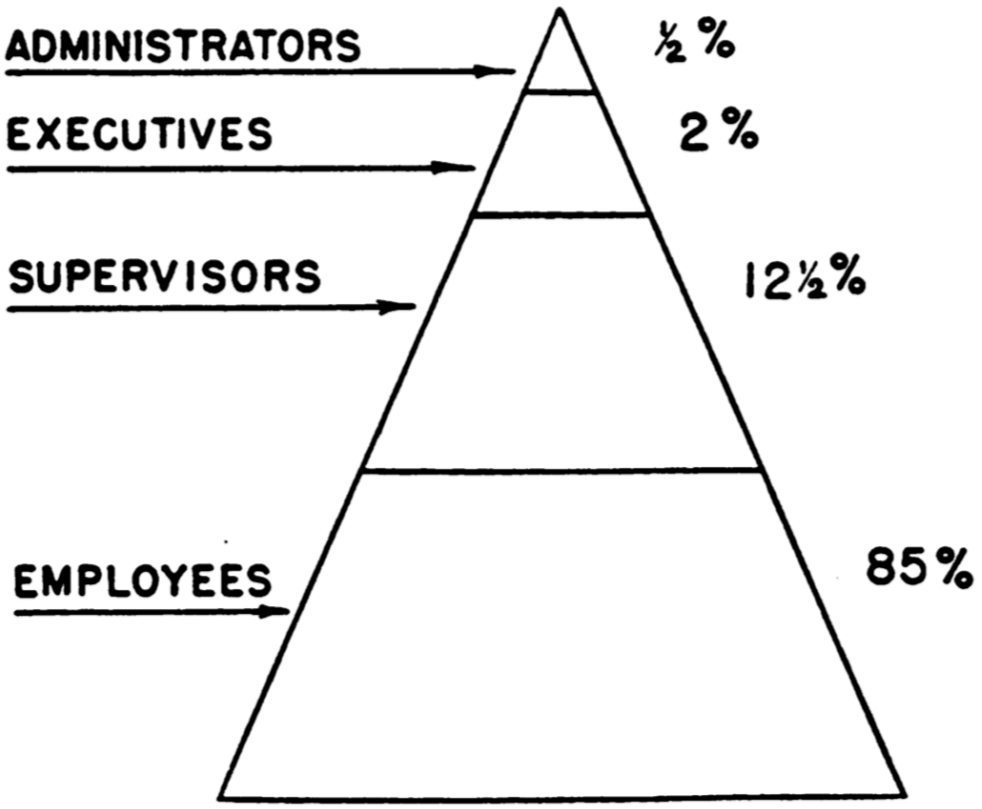

Figure 1: Personnel Distribution United States Civil Service Commission, p. 143

many archivists themselves are historians and thus are "competent to ascertain the historical values of public records" (Schellenberg, 1956, p. 150). As opposed to a highly structured approach, Schellenberg stressed that determinations of *informational value* by archivists resist consistency and systematization. As a result measures of *informational value* will be different in different contexts, and that "diverse judgments may well assure a more adequate social documentation".

Even with these allowances for contextual variation, Schellenberg's notions of evidential and informational are significantly framed by the institutional and governmental context. This is plainly seen in his use of the hierarchy of government as a mode of selection in Figure 1. In addition, judgments about historical significance made by archivists and other subject specialists, are key to establishing information value, but are themselves bound up and subject to less visible professional structures of power, which are left largely unexamined. Into this gap we see the final stage in top-down approaches, as the unit of analysis moves from the level of the state up yet further to that of society at large.

Documenting Society

The archival theorist Terry Cook was an eloquent archival theorist, most remembered for his ability to synthesize archival theory, and mobilize it for work in the present, particularly in the service of appraisal. His theory of *macroappraisal* provides a framework for shifting focus away from the value of records, and towards the values inherent in the functional context that records are created in (Cook, 2005). When using macroappraisal archivists study the functions and structures of government and organizations, while specifically attending to the effects that these institutions have on their citizenry and in aggregate, society as a whole.

> Macroappraisal is thus a provenance-based approach to appraisal, where the social context of the record's creation and contemporary use (not its anticipated research use) establishes its relative value. (p. 128)

Cooks' articulation of macroappraisal was born in the aftermath of Canada's Deschênes Commis-

sion on Nazi War Criminals, which uncovered how a large number of immigration records were inadvertently destroyed by the Canadian National Archives. The controversy sensitized Cook to the abject failure of Schellenberg's concept of *informational value*, or the historical determination of record value. It led him to instead focus attention on present value, instead of future anticipated use, and to anchor that value on a determination of how citizens were impacted by an organization's activities. Cook drew heavily on German archival theorist Booms (1987), who argued against state controlled appraisal, under repressive Soviet-style communism, and advocated for a view of archival appraisal that measured the interests of society as a whole:

> In our view, a legitimate value standard or principle for the archival appraisal process can only be derived from this kind of contemporary valuation. Such a standard is inherent in history itself, for it is a standard of the past. It is not the product of speculation or ideological beliefs; it does not do violence to source material by applying value standards of the present which in the near future may already prove to be inadequate. If there is indeed anything or anyone qualified to lend legitimacy to archival appraisal, it is society itself, and the public opinions it expresses- assuming, of course, that these are allowed to develop freely. (p. 104)

Here we come full circle with the idea of archival appraisal as reflective of the goals of democratic societies as discussed earlier (Eastwood, 2002). Even with its focus on the state, its citizenry and society as a whole, macroappraisal foregrounds a particular governance structure that supports the needs of a democracy and even (in the case of Booms) notions of social opinion and the marketplace of ideas. While fully aware of the political role that appraisal takes in shaping memory, Cook understands appraisal as a tuning of the machinery of the state, to bring it in line with the interests and well being of its citizenry.

Booms' emphasis on the importance of public opinion is certainly understandable, and even laudable, when faced with state sanctioned repression, violence and media censorship. However an unquestioned reliance on a market for determining public opinion, without a critical engagement

with the systems that generate those markets is problematic. Foucault's concept of governmentality is useful for unpacking and factoring these social and market forces, to see them not only at work in the functions of government, but also in the distributed material practices of individuals, and communities. Understanding how to measure popular opinion, or the impact of records on people, and some approaches that archives have taken to a form of appraisal that speaks directly to it political agency is where we turn next.

Appraisal Critique

In contrast to governance based approaches to archival appraisal are a group of divergent theories that speak directly to, and work to counter, the obscured, and often erased, influence and effects of power in archives. As we saw previously, structural approaches to appraisal often define appraisal implicitly by outlining principles of original order and provenance that map to an institutional or societal context. In emphasizing the central role of provenance in structuring the archive, certain assumptions are made about what records get to count as archival, who creates them, and what resources are required to mobilize them as an archive. Another group of appraisal practices in archives speak to the archive as a site for record *creation*, not simply preservation. Archival records are part of a larger landscape of memory in which archival records circulate. These bottom up style approaches to appraisal archivists fully engage, and take political responsibility for, their role as shapers of the archival record, and the limits that their actions impose.

In a memorable, and oft quoted, address to the Society of American Archivists in 1974, Ham (1975) issued a wakeup call for archivists and their practice of appraisal:

> Our most important and intellectually demanding task as archivists is to make an informed selection of information that will provide the future with a representative record of human experience in our time. But why must we do it so badly? Is there any other field of information gathering that has such a broad mandate with a selection process so random, so fragmented, so uncoordinated, and even so often accidental? (p. 5)

Ham highlighted the lack of theorization around appraisal in order to make a plea for increased cooperation, empirical analysis, funding, and intellectual openness for archival appraisal. In many ways Ham was echoing a similar call by historian Howard Zinn, who addressed the same body a few years earlier saying that the archivist had a responsibility to document the lives of ordinary citizens and political movements, to hold power accountable through transparency, and to consciously work against the status quo in archives where

> … the existence, preservation, and availability of archives, documents, records in our society are very much determined by the distribution of wealth and power. That is, the most powerful, the richest elements in society have the greatest capacity to find documents, preserve them, and decide what is or is not available to the public. This means government, business, and the military are dominant.

Zinn's insights here continue to reverberate in archival studies especially when considering the process of appraisal. He foregrounds the hegemonic shape of archives, and challenges the archival community to not only hold the disciplinary form of the archive accountable, but also to move archival work outside of its governmental enclosure. In many ways this recognition of the role of power in shaping archives marks the birth of appraisal theory. It is in the moment when one can see the dominant mold of archives and its historical form that it becomes possible to talk about how that power has manifested in the records that have been collected and cared for in an archive (Jimerson, 2009). Foucault's exploration of the role of governmentality is instructive here because he lets us see how power works through the state, but also through disciplinary forms such as the family, the school, the factory, and the prison.

Strategy

Documentation Strategies is an appraisal technique developed by Helen Samuels as she worked to document university life at the Massachusetts Institute for Technology. For Samuels the archivist works as an analyst in a *distributed network* to study the social forces that generate records, rather

than the records themselves. Samuels articulated four parts to developing a documentation strategy: 1) choosing a topic to be documented, 2) selecting the site for the strategy 3) examining the form and substance available documentation, 4) selecting and placing the documentation. This general strategy-for-developing-strategies, or meta-strategy, was distilled down to an appraisal technique known as Institutional Functional Analysis which bears some resemblance to Cook's macroappraiasl, particularly in its attention the functions and societal impacts of institutions.

However, Samuels' crucial insight was that a *network* of analysts and repositories would be needed to address the challenge. A documentation strategy was not concerned with the decisions of a single archivist working within the hierarchical structure of a single institution. This networked approach mirrors Samuels' interest in emerging practices around the use of automated computer networks such as OCLC, RLG and at the distribution of machine readable cataloging data at Library of Congress (Samuels, 1986). A key aspect to Documentation Strategies is deciding where records are to be held, which may or may not be at the archivists own institution.

Significantly, Samuels also recognized that a documentation strategy may identify gaps in the documentary record, and could in fact suggest the *creation* of records where none currently exist.

> While archivists acknowledge the overabundance of information, they also recognize that modern communication patterns and records-keeping practices leave gaps in the documentary record. Documentation strategies, however, are ongoing activities and provide the opportunity to intervene in the records creation process and assure the creation and retention of required information. (Samuels, 1986, pp. p121–122)

Seeing archival appraisal as an intervention and as a potentially creative act was a surprisingly radical move, especially given the vast number of records already in need of preservation, and the anxiety about preserving them.

Decentralization

This shift in attention outside the walls of the physical archive suggested by Documentation Strategies was in fact part of a longer movement in archival studies that often gets labeled *postcustodial*. The term postcustodial was first used by Ham (1984) in drawing attention to the ways in which information technology and automation were dramatically transforming the landscape of record production, and the concomitant need for the archival studies field to invest in researching these new forms of record production, in order to adequately preserve and provide access to them. A decade later Cook (1993) synthesized a significant body of work by Taylor (1988), Bearman (1989), Hedstrom (1991) under the rubric of *postcustodial*, which he aligned with post-modern theories of archives. For these postcustodialists a transformation of archival theory, and especially appraisal, was absolutely necessary because of the proliferation of electronic media, which resist the idea of a singular documentary artifact.

> In this fluid electronic environment, the idea of a record physically belonging in one place or even in one system is crumbling before new conceptual paradigms, where "creatorship" is a more fluid process of manipulating information from many sources in a myriad of ways, or applications, rather than something leading to a static, fixed, physical product. For information professionals, this signals that the custodial era is giving way to a postcustodial one, where the curatorship of physical objects will define our professions much less than will an understanding of the conceptual or virtual interrelationships between creating structures, their animating functions, programmes, and activities, the information systems, and the resulting records. (Cook, 1993, p. 424)

Postcustodial thinkers emphasize that appraisal is not simply a valuation of records during record acquisition, but is inherent in the design and construction of information systems: "archivists need to reexamine how information systems support organizational functions and relate to organizational structure within specific organizations and in a broader documentary context" (Hedstrom, 1991, p. 344). Postcustodial archival theory crystallized in the Records Continuum model, which dis-

mantles the idea of a linear life cycle for records in which the archive is the place where records come to rest. The Records Continuum instead stresses how archival records are part of multiple, recursive, processual flows, as they cycle between creation, capture, organization and pluralization (McKemmish, Upward, & Reed, 2010). In the Records Continuum appraisal is less concerned with ascertaining the value of records as it is with the design and implementation of *systems* that generate the records.

Outside the Archive

The need for appraisal to encompass the record creation process is also reflected in community archives approaches. Community archives extend and build upon postcustodial archival practice by situating the work of appraisal out in the world, in particular social contexts where records are created and actively used (Flinn, Stevens, & Shepherd, 2009). This movement allows communities to retain custody of their records, which affords more autonomy during the appraisal process. The decisions of what to archive are not being conducted solely by archivists, but also by members of a particular community of people, who are ultimately using the records. Rather than simply treating custody as something to decenter or transcend, because of the material configurations of computer networks and information technology, custody is directly linked to access and the use of records (Bastian, 2001). Community archives allow for records to take on new historical dimensions, interpretations and use for identity formation (Punzalan, 2009) and collective memory (Brothman, 2001 ; Jacobsen, Punzalan, & Hedstrom, 2013).

It is significant that community archives approaches often develop in response to the systematic erasure, marginalization, or disenfranchisement of particular groups, that is performed by traditional, institutional forms of archives (Flinn, 2007). In this way community archives speak to the archive as a site and instrument of power (Jimerson, 2013), for social justice (Punzalan & Caswell, 2016) and even activism (Cifor, Caswell, Migoni, & Geraci, 2018). In the context of social justice, archival appraisal fully admits, and even celebrates, its active political role in shaping and reshap-

ing collective memory and identity (Harris, 2002). For some, information technologies, such as the web and social media, offer new opportunities for participation, which broaden the set of actors who can perform appraisal, and thus the diversity of records (Gilliland & McKemmish, 2014; Huvila, 2008).

More recently there has been a move to position archival appraisal in terms of a feminist ethics of care, which de-emphasizes a rights-oriented, legalistic approach to social justice, which can inadvertently work to reinscribe the very same oppressive systems they are working to dismantle (Caswell & Cifor, 2016). The ethics of care approach expresses appraisal and other archival activities in terms of a network of mutual responsibility. These relationships become a conduit for appraisal as a measure of affective value, for deepening engagement with communities and furthering social justice (Cifor, 2016). We will turn to a deeper discussion of the ethics of care and its role in understanding web archiving as a set of repair practices shortly.

Appraising Appraisal

Admittedly, this cursory overview has glossed significant aspects of archival appraisal, while also failing to mention others. However my goal here was not to provide an exhaustive description of the field, but to describe a constellation of divergent and even opposed archival appraisal theories, in order to situate them with Foucault's notion of governmentality. Governmentality helps us to examine appraisal theories as a set of knowledge practices, that orient around organizational structures as well as to individual agency, and reproduce themselves at the levels of societies, states, communities, and even individuals. Appraisal is implicit in the duty to record the activities of the state or administrative body; it works along the grain of organizational hierarchies to document what is deemed most important; and it operates in the service of documenting society, and the interests of democracy and its citizenry. But in a counter movement appraisal also works to decenter the archive as institution, and to locate appraisal practices in systems of record creation and the design of information systems. Appraisal responds to the juridical form of the archive: to create

records where none exist, and to reinterpret existing records as part of recursive process that dismantles linear conceptions of the record lifecycle. Appraisal is a tool for social justice that speaks directly to power formations, in order to address oppression and marginalization, while imagining new frames for collective memory. Appraisal even attempts to reconcile the dialectical forces of structure and agency by centering the ethics of care and practice, and moving outside of western human rights based frameworks.

Appraisal and Web Archives

And so we return to the question that we started with: how are we deciding what to archive from the web? Or, how does the constellation of archival praxis around appraisal that we have just discussed meet with web archiving practices? You will notice that so far there has been very little discussion of the actual materials that are the subject of appraisal. It's almost as if our theories of appraisal are thought to function independently of the material being appraised; that in principle (if not in fact) appraisal can be applied as needed to all forms of media. Indeed, this abstractedness is part of appraisal's claim to theory in the first place, and forming part of an archival *science*.

However the birth of appraisal, or the awareness that records *must* be selected, in some fashion, by the archivist, occurs in a particular historical moment of profound material transformation, as the technologies of record production completely outstrip the archivists ability to process their outputs. This *overcoming* of the archivist, and the concomitant appraisal response, are not simply the result of an increased volume of records, or information overload (Ham, 1981). The concept of archival appraisal arrives as the centuries old archival technologies of paper, print and their containers are giving way to a proliferation of electronic media formations, which challenged and continue to challenge the archival imagination (Taylor, 1988). Cook (1994) suggests that appraisal is perhaps more of a sensibility than an abstract theory, for the way it offers an approach to practice, and a consolation amidst this transformation:

> Most important, for the first time, we are not producing, managing, and saving physi-

> cal things or artifacts, but rather trying to understand and preserve logical and virtual patterns that give electronic information its structure, content, and context, and thus its meaning as a "record" or as evidence of acts and transactions.

Here Cook casually deploys the idea of "virtual" records and the effects that they are having on archival practice. This virtuality derives at least in part from how computer technology collapses the media forms of word, picture, sound, video into a seemingly uniform binary representation–the so called ones-and-zeroes of digitality. However electronic records, especially born-digital-records, of which the web is a prime example, have actual material dimensions, and condense as the result of processual flows that involve platforms, infrastructures, networks, heuristics and algorithms. The virtuality of electronic records can be understood as what Kirschenbaum (2008) calls a "screen essentialism" or medial ideology that effaces the instrumentation that creates, and must constantly recreate, the experience of an electronic document. At the same time, archives of the web are particularly prone to a technological solutionism that treats preservation as a purely technical problem, where we simply need more efficient and less error prone storage, more comprehensive crawling strategies, or improved network protocols in order to "archive it all".

Despite efforts to archive the entire web (Kahle, 2007), the idea of a complete archive of the web remains both economically infeasible (Rosenthal, 2012), and theoretically intractable (Masanès, 2006b). Features of the web's Hypertext Transfer Protocol (HTTP), such as code-on-demand (Fielding, 2000), content caching (Fielding, Nottingham, & Reschke, 2014) and personalization (Barth, 2011), have transformed what was originally conceived of as a document oriented hypertext system into a complex multimedia information infrastructure that delivers content based on who you are, when you ask, and what software you use (Berners-Lee & Fischetti, 2000). As a result, the very notion of a singular documentary artifact, which has been under significant strain since the introduction of photography (Benjamin, 1999) and electronic records (Bearman, 1989), is now being pushed to its theoretical and conceptual breaking point. We rarely try to reason about *all* archival records that exist in the world as a singular assemblage. So why do we do this so naturally with the

network of networks that we call the Internet, or the massively distributed text that is the web? Dissolving the seeming virtuality of web archives, and understanding both the web and web archives as social, historical and technical artifacts is essential for being able to talk about how web content is being selected for preservation (and access) in an archive. Now that we are 30 years into the web's existence we are in a position to turn from the innovation-centric approaches to web archives, to the critical investigation of established practices.

Science and Technology Studies

In the early days of the web Hedstrom (1991) proposed a broad research framework for the study of *electronic records*, or what some today might call computational archival systems. This framework was based on a *sociotechnical* understanding of archives that recognized the need for archival studies research to move beyond technological determinist accounts, where society is shaped by technology, as well as overly socially determinist accounts, where technology is shaped purely by social and political interests. In *sociotechnical* accounts both technology and society must be understood in terms of a complex interplay between people, materials and processes, where the technical and the social mutually shape, or coproduce, each other. This line of inquiry is especially relevant today,and builds upon the rich empirical foundations of Science and Technology Studies (STS) (Felt, 2017; Latour, 1987 ; Pinch, 1987).

Hedstrom's framework for research into electronic records included 1) the relationship between automation, electronic records and organizational change; 2) new material forms of electronic records and their remediation of previous forms; 3) the design of new information systems with archival properties; 4) evolving markets for information technology and their impacts on archival records; and 5) the impact of electronic records on accepted norms and approaches to archival preservation (e.g. provenance and appraisal). Hedstrom stressed how attending to the social, historical and material dimensions of information technology, was the cornerstone of archival studies:

> The introduction of new forms of material and the simultaneous transformation of

> traditional forms into something new raises a series of questions about the relationship between forms of material and archival practice. When should new forms of material be managed differently from more traditional forms of documentation? Are there any archival principles that apply to all new electronic record types? What characteristics does an electronic memo share with a memo on paper? What does it have in common with other machine-readable records?

Hedstrom's framework holds up particularly well today, as many of these five areas have developed into full fledged fields of study. But this development, at least for the study of web archives, has mostly happened either in the domain of computer science where the concept of appraisal is infrequently used, or in the field of archival studies where the web is considered under the rubric of electronic records. Before discussing how STS can be applied to the study of appraisal in web archives I will briefly characterize these literatures and their perspective on web archives.

Digital Libraries

The computer science literature reflects a decade long interest in *digital libraries*, particularly the research output of Joint Conference on Digital Libraries (JCDL), that for the last 20 years has served as a research forum for all manner of investigations into novel methods for collecting, storing, indexing, accessing and preserving content collected from the web. When it comes to appraisal, digital library research on the subject of web archives has focused in large part on the problems of *harvesting* or *crawling* the web. Practices for crawling the web for the purposes of indexing and maintenance are nearly as old as the web itself (Fielding, 1994). However the actual dimensions of the web are still difficult to determine (Dobra & Fienberg, 2004), and the "deep web" presents challenges because of the way large regions of the web are hidden behind human driven query interfaces (Madhavan et al., 2008).

Substantial work has been done to measure the *archivability* of web content (Banos & Manolopoulos, 2015 ; Samar, Huurdeman, Ben-David, Kamps, & Vries, 2014) and to determine the age of web

resources (SalahEldeen & Nelson, 2013), which are important factors to consider when deciding what to archive. Another significant consideration is the expected disappearance of a resource from the web. Numerous studies have investigated the transience of web resources, so called link decay, link rot, or reference rot, in order to characterize the likelihood of certain types of web content disappearing. Measures of link rot can be used to mobilize efforts at increased web archiving generally, as well as specific areas such as social media (Salaheldeen & Nelson, 2013) , science (Hennessey & Ge, 2013), institutional repositories (Sanderson et al., 2011), and jurisprudence (Zittrain et al., 2014).

The problem of knowing what to collect from the web has also been treated in the digital library research community as a *focused crawling* problem. In focused crawling the goal is to collect content about particular topics (Risse et al., 2012), events (Klein, Balakireva, & Van de Sompel, 2018; Yang, Chitturi, Wilson, Magdy, & Fox, 2012), or to collect content that has a particular characteristic such as popularity (Page, Brin, Motwani, & Winograd, 1999), importance Baeza-Yates, Marin, Castillo, & Rodriguez (2005)] or social engagement (Gossen, Demidova, & Risse, 2015 ; Milligan, Ruest, & Lin, 2016; Nwala, Weigle, & Nelson, 2018). Generally speaking these approaches take the focus to be a topic, event, person, organization that can be qualified by the types of media (documents, audio, video). These criteria are taken as a priori, or a given, that is decided out of band, before the task of doing the crawling is undertaken. Similarly there is also a vast research literature on relevancy ranking in web search results that is dependent first on having already amassed a collection of web documents to index. However how that corpus of documents has been assembled is usually taken as a given. Despite the lack of the use of the term appraisal, this vein of digital library research speaks to Hedstrom's call for archival studies to directly engage in the design of record keeping technologies for the web.

Archival Studies

Within the field of archival studies there are numerous accounts of how appraisal applies to electronic records, but less so with regard to the web specifically. The term *electronic records* developed alongside computer automation to refer to the databases, electronic mail archives, and other types of data that accrue as *files* on a computer filesystem. These computer filesystems used the metaphors of paper based documents, files and containers for storing them, in order to make familiar a completely new information infrastructure. Designating these computer files as electronic records highlights how computer data is generated as part of the functioning of organizations, and provides evidence of that operation, in which authenticity, reliability and fixity are typically paramount (Cobb, Pearce-Moses, & Surface, 2005 ; Duranti, 2010 ; Harvey & Thompson, 2010). Discussion of electronic records also happens under the rubric of digital preservation (Rothenberg, 1999) or digital curation (Yakel, 2007). In addition to the preservation of computer files and filesystems researchers have worked on means of format identification (Hitchcock, Brody, & Hey, 2007) and pursued the use of software emulation to preserve and describe entire software systems (McDonough et al., 2010; Welch, Rehfeld, Cochrane, & Suchodoletz, 2012). Lee & Tibbo (2011) suggest that the term *digital curation* reflects a postcustodial response to digital preservation, that foregrounds the site of record production as outside the walls of the archive, instead of focusing entirely on the authenticity and fixity of data, and conventional ideas of the preservation lifecycle.

Electronic records, digital preservation and digital curation are largely predicated on prior custody of data, or the ready access to the computer systems (software and hardware) that the data has been produced on. These approaches tend to take as a given that the data is in fact available, and generally do not speak directly about the process of archival appraisal, or the social and technical means by which data is selected for preservation in an archive. As such the web presents a conundrum of sorts, where content *appears* openly available, and is often regarded as *public*. But even as they seem ready-to-hand, web documents also exist at a distance, sometimes in multiple locations, and are retrieved across blurred geopolitical boundaries, which the infrastructure of the Internet and the

architecture of the web (when working) makes instantly available in the web browser.

Many accounts of the appraisal of web content take a case study approach where web content is collected to document a particular event or category of content. For example Masanès (2006b) examines the identification and crawling of websites relating to the French presidential and parliamentary elections in 2002 by the Bibliothèque nationale de France. In another prominent study Schneider, Foot, Kimpton, & Jones (2003) detail approaches to creating a collection of web content at the Library of Congress related to the September 11 terrorist attacks. Changes to copyright law enabled national libraries to evaluate both broad and focused approaches to archiving entire top-level-domains (e.g. .fr or .pt) (Lasfrgues, Oury, & Wendland, 2008, and @Gomes:2006). The need for tools that allow *seed lists* (lists of URLs to archive) to be developed was articulated by Pearce-Moses & Kaczmarek (2005). Duncan & Blumenthal (2016) describe how networks, such as professional consortia, are useful resources for curating seed lists. In addition there has been discussion of the use of social media and their Application Programming Interfaces (APIs) as a means for discovering web content for archiving events such the Arab Spring (Arnold & Sampson, 2014) and the protests in Ferguson and the #BlackLivesMatter social movement (Rollason-Cass & Reed, 2015). Web archives are also considered as a site for critical engagement with issues of social justice (Aronson, 2017).

In one of the more conceptual models of appraisal in web archives Masanès (2006a) analyzed the process of *selection*, which is broken down into three different phases: preparation, discovery and filtering. Preparation involves establishing a focus for the collection, as well as selecting the technology for performing the collection. Discovery involves the act of collection itself, which leverages both the structure of the web (endogenous) as well as external resources such as search engines and link hubs (heterogeneous). Filtering is a process by which collected material is selected for an archive based on quality, subject, genre and publisher. Masanès discusses the degree to which automation can be used in these three phases, but unlike the previously mentioned case studies the discussion is prospective and largely unanchored from actual practice. For example, whether and

how archivists are involved in the filtering of collected web content is not explored.

The architecture of the web presents fuzzy geographies, where it is not always clear where one website ends and another begins. The idea of a singular document, which is central to information science (Buckland, 1991), is put under strain by the web's use of hypermedia, which allows documents to be dynamically composed from heterogeneous sources. Practices for appraising the web diverges significantly from electronic records in that it necessarily involves collecting or assembling content from the Internet. Appraising web content also requires the means and mechanisms for reassembly, or replay, of the content later (Andersen, 2013) in order to assess what has been collected. This recursive assembly and reassembly of content is tightly bound up with the appraisal process itself, and is achieved with specialized tools (Mohr, Stack, Rnitovic, Avery, & Kimpton, 2004), that have particular design assumptions, goals and affordances built into them.

Zooming out from the technical specifics, appraisal in web archives also can be seen as a form of virtual reunification (Punzalan, 2014) where who the various stakeholders are, their conceptions of process and product, and goals in creating the archive are made manifest. However in the case of web archives this reunification was always already virtual, at least in the sense that they were very often born digital, as part of the accrual of data and documents in the network. Unpacking the very real social and material processes and practices that underly the virtuality presented by the web archive is a key component to understanding how archival appraisal on the web works.

Appraisal Practice

In this paper I have outlined how ideas of archival appraisal have developed over time, with the goal of showing how these conceptions are used (and not used) in the research literature of web archives. In examining the research literature of appraisal I have drawn on Foucault's idea of *governmentality* as a way of understanding how a wide set of archival appraisal practices reflect, repeat and respond to concerns of governance. However, for Foucault governmentality is about much more than simply the affairs of government and state: it also includes the practices of individuals, families and groups

of all kinds:

> This word [government] must be allowed the very broad meaning it had in the sixteenth century. "Government" did not refer only to political structures or to the management of states; rather, it designated the way in which the conduct of individuals or of groups might be directed: the government of children, of souls, of communities, of the sick. It did not only cover the legitimately constituted forms of political or economic subjection but also modes of action, more or less considered or calculated, which were destined to act upon the possibilities of action of other people. To govern, in this sense, is to structure the possible field of action for others. (Foucault, 1982, p. 790)

Foucault developed the concept of governmentality by examining how institutions like hospitals, schools, prisons, barracks, and even archives (Foucault, 1986) work to discipline, and delimit the field of action. Foucault's examination of specific practices and their relations is considered by many to be part of a general turn to practice in social theory. Postill (2010) identifies Foucault as part of a first wave of practice theorists that also includes Ludwig Wittgenstein, Pierre Bourdieu, Anthony Giddens, and Michel de Certeau. He characterizes these thinkers as working to:

> … liberate agency – the human ability to act upon and change the world – from the constrictions of structuralist and systemic models while avoiding the trap of methodological individualism. These theorists regarded the human body as the nexus of people's practical engagements with the world. (p. 7)

Practice theorists are interested in resolving the dialectic between individual agency and the social and material structures that constrain and reproduce it, through a material examination of the processes and routines that center on our physical bodies and experience. Giddens (1984) uses the idea of *structuration* to propose that agency and structure are mutually constitutive entities with equal status, each of which recursively reproduces the other. The idea of structuration has already seen some purchase in the field of archival studies where it has been used for understanding appraisal (Brown, 1991), descriptive practices (Yakel, 2003), the space/time dimensions of records (Upward,

1996 , 1997), and the use of collaborative documentation networks.

Similarly Bourdieu (1977)'s idea of *habitus* identifies the site of our lived experience which is shaped by social structures, which generate individual action. Schwartz (1995) mobilizes Bourdieu when analyzing how archival photographs function as documents that both produce, and are the product of, social rules. Gracy (2007) has also drawn on Bourdieu's notion of field, habitus and social capital in researching how commercial and nonprofit film archives operate. In some significant recent work Ivanov (2017) conceptualizes archival theory in terms of practice theory, and uses both as a framework for examining record keeping practices in large news organizations.

At first it might seem strange to consider the physical body and its practices in relation to something as seemingly immaterial as the web. But the web and the underlying Internet are physical infrastructures that constantly generate records as we point and click in our browsers, swipe the screens of our smartphones, as we are quietly surveilled by the Internet of Things (Acker, 2016). Archives of web content sediment on servers as we document our lives using record *making* technologies such as Facebook Live, which aren't always record *keeping* technologies (Sheffield, 2018). Given this orientation towards practice and the general theoretical umbrella provide by Science and Technology Studies I will conclude by suggesting several promising areas for for future research into appraisal and web archives.

Ethnography of Infrastructure

The ethnographic study of infrastructures in terms of the human practices that play out in their construction and use is well developed in information studies (Bowker, 2005; Star, 1999 ; Suchman, 1985). For example Edwards, Mayernik, Batcheller, Bowker, & Borgman (2011) has examined how metadata practices work to shape knowledge production in the sciences. The mutually constitutive roles of policy, practice and design can shape infrastructures and define the controversies that surround them (Steven J Jackson et al., 2014). Social media applications like blogs have been studied for their role in shaping scholarly communication infrastructure (Burton, 2015). The *convivial*

decay of space science infrastructure has been investigated by Cohn (2016), who found that notions of repair, aging and multiple intersecting life stories of people & machines played a significant role in how infrastructure is designed, deployed, used and then dismantled. Karasti & Blomberg (2018) propose a methodological framework anchored in STS and anthropology for analyzing infrastructures over long time scales, that we see evidenced in the work of Ribes & Finholt (2009) on the *long now* of infrastructures.

The ethnographic study of web archiving infrastructure in particular has also begun, but more work remains. Ogden, Halford, & Carr (2017) provides a rare glimpse at how labor and infrastructure intersect in the practices of web archivists at the Internet Archive. Similarly Kosnik (2016) has performed a detailed ethnography of archival practices in fan fiction communities on the web. This recent work connecting web infrastructure, archives, and social practices suggests further ethnographic investigations of how web archives both shape and are shaped by ideas of appraisal, or what is deemed valuable in the web.

Repair and Maintenance

The turn to practice in Human and Computer Interaction (HCI) has been noted by Kuutti & Bannon (2014), who suggests that HCI's focus on user needs, or macro-level organizational systems are no longer sufficient for understanding the complex formations of users and systems we see today in digital environments. The study of computer systems in terms of practice offers a way of collapsing these scales. Kuutti points out that the turn to practice in HCI research takes shape around issues of performativity of social practices, materiality of human bodies and artifacts, agency, and knowledge generation.

The role of repair as a site for design, in contrast to more conventional and celebratory notions of innovation, is detailed in ethnographic work by Steven J. Jackson & Kang (2014), DiSalvo, Sengers, & Brynjarsdóttir (2010), Rosner & Ames (2014) and Maestri & Wakkary (2011). *Broken world thinking* (Steven J. Jackson, 2014) specifically recognizes that design happens not only in the

experimental setting of HCI research, but in our everyday lives as we learn to adapt and improvise solutions to conceptual and infrastructural breakdowns. Repair and breakdown are seen as two sides of the same coin of maintenance. Russell & Vinsel (2018) situate maintenance studies in terms of the history of technology, where the investigation of how standards are developed can help make sociotechnical assemblages legible (Lampland & Star, 2009; Russell, 2014).

With the maintenance perspective in mind it is possible to broaden the scope of what appraisal means in web archiving systems to include not only the design and use of web archiving systems, but also the material practices and labor that sustain the web, as software is upgraded, vendor contracts are abandoned and content is ported from one system to another. Tracking these processes of data migration as appraisal decisions in web archives is a viable and under explored avenue for archival studies research, particularly in connection with postcustodial theories of the archive.

Hedstrom (2002) suggests that the archivist's appraisal decisions are an essential part of the archive's *interface* with the past, found both on the screen, and embodied in the archivist. In some recent work Maemura, Worby, Milligan, & Becker (2018) has begun to explore how archival decisions manifest in the provenance of web archival systems, in terms of the documentation they provide and the sociotechnical means by which that documentation is generated and conveyed. Further surfacing the repair work of archivists, and the web as a site for continual repair and maintenance is a key area for future research into archival appraisal on the web.

Software Studies

The field of software studies provides a humanistic method for reading software and digital media systems as artifacts, with particular material, social and historical dimensions. For example, Manovich (1999) has studied the database as a narrative form whose ontology structures the cultural artifacts that it helps generate. Kelty (2008)'s pioneering ethnographic work has helped us understand open source software development communities as a *recursive public* that uses the infrastructure of the web and the Internet to develop and express what the Internet is, or can be.

Another class of research that fits (perhaps a bit less comfortably) under the rubric of software studies is work that explores the materiality of digital media in terms of inscription (Kittler, 1999), memory (Chun, 2011), digital storage and transmission (Kirschenbaum, 2008), protocols (Galloway, 2004), physical networks (Starosielski, 2015), documents (Gitelman, 2014; Levy, 2001), data representation (Dourish, 2017) and even archives (Ernst, 2013). In this varied literature there is a consistent engagement with how data systems are anything but neutral (Bowker, 2005 ; Gitelman, 2013 ; Walford, 2017), and are configured by the material that they are constructed from, and by the design decisions of their creators and maintainers.

One relevant area for understanding appraisal in web archives is the growing fields of platform and algorithm studies, which are closely aligned, but operate from different levels of abstraction and granularity. Platform and algorithm studies examine the social, political and cultural contingencies that the technologies provided by organizations like Google, Facebook, Twitter, Wikipedia coproduce with their publics and users. At the platform scale these contingencies manifest in established policies such as moderation rules, community guidelines, and terms of service documents, which are translated into actual practices, and in turn generate political economies (Gillespie, 2010 , 2018). Similarly algorithm studies, despite its usual association with computer science, provides a humanistic lens for studying the development, use and impact of computation in particular settings, enabled by specific practices, in order to achieve explicit or tacit ends (Gillespie & Seaver, 2015 ; Seaver, 2017). As such platform and algorithm studies fit within the scope of critical data studies (boyd & Crawford, 2012), which shifts the focus from policy and code to flows and accumulations of data, and their politics (Bratton, 2016; Zuboff, 2015).

As noted by Dourish (2017), it's awkward, and perhaps a bit misleading, to lump all these theoretical concerns into the category of *software studies*. However doing so, is a convenience that allow us to talk about the possible avenues of future research for web archives, especially with respect to appraisal. At the risk of introducing yet another category, Kitchin & Lauriault (2014) draws on STS terminology to mobilize the idea of *data assemblages*, which are heterogeneous,

sociotechnical constructs that resemble Foucault's *dispotifs* that act as bundles of "discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral and philanthropic propositions" (quoted in Kitchin & Lauriault, 2014). Kitchin suggests several methods for studying the complex phenomena that are data assemblages:

1. examining code artifacts, their data inputs and outputs as time bound processes
2. reflecting on the writing & design of code, heuristics and algorithms
3. reverse engineering algorithms, to intuit processes and architectures that are hidden from view
4. conduct ethnographies of design teams to ascertain the contingent, relational, and contextual way software is produced
5. widening the ethnographic lens to consider institutional and organizational forces at work
6. widening the ethnographic lens yet again to consider the work that the algorithmic systems do in the world, either intended or unintended, and their social and historical significance

In some significant recent work Ben-David & Amram (2018) used techniques from the study of *black box* algorithms (algorithms whose inner workings are secret, or so complex that they are unknowable) (Diakopoulos, 2014) to consider the epistemic role of web archives as fact or evidence producing systems. Ben-David specifically looks at the representation of the the North Korean top-level domain (.kp) in the Internet Archive's Wayback Machine, using traces of provenance information provided by the Wayback Machine's interface, in combination with the historical contingencies of DNS leak that occurred in 2016. Ben-David finds that the Wayback Machine's processes for acquiring data from the web are found to be iterative, top-down, bottom-up, and that they extend laterally outside of the Internet Archive organizational walls.

Similarly Schafer, Musiani, & Borelli (2016) employs techniques from STS to unpack the sociotechnical black-boxes of web archive infrastructure to stress the importance of web archives for Internet governance. Both pieces highlight how the quiet labor of web archives, performed by archivists in collaboration with, and sometimes in spite of, machines, are of historical and political significance.

More work remains for understanding the technological, social and historical contexts in how these software systems, standards and organizations have developed–particularly with regards to our understanding of what a web archive is, what it is not, and how the difference is negotiated. The time is ripe for these analyses since we have 30 years of web, and 20 years of web archiving practice to study.

Data Justice

And finally, a significant strand of work in the turn to practice has drawn attention to how the design of information systems can respond to the needs of social justice. The imperative for technology to be designed by those who are supposed to ultimately use it was initially developed in the 1960s in Scandinavia, and gave rise to the field of participatory or cooperative design, which has garnered sustained interest from the HCI and Computer Supported Cooperative Work research communities (Ehn, 1988 ; Kensing & Blomberg, 1998; Star & Ruhleder, 1996 ; Suchman, 1985). Related concepts such as Communities of Practice (Lave & Wenger, 1991) provide a rich methodological framework for understanding how social practices involving technology constitute the way we learn and work together in sustainable ways. Furthermore, the role of information technology infrastructures in measuring and making legible environmental change (Edwards, 2010 ; Steven J Jackson & Buyuktur, 2014) while also negatively impacting local and global environments (Cubitt, 2016) is an increasingly important area of concern. Indeed, the general thesis that there is no escaping the politics that are embedded in our technologies is a theme that is returned to again and again (Winner, 1980). More recent recent work has explored how data infrastructures are both the product of, and can give rise to capacity building for mobilizing change (Meng & DiSalvo, 2018; Tufekci, 2017). This work uses an ethnographic approach to consider how data practices are part of a constellation of other social and political practices.

This research literature bundled under the theme of social justice in information technology intersects broadly with archival studies in significant ways. As noted by Punzalan & Caswell (2016),

archives and memory studies have been perennially interested in the role of social justice. These researchers have worked to dispel the notion that archives are neutral in their representation of communities or society as a whole, and that they necessarily embody particular social and political values. Specifically the goals of community archives (Flinn, 2011) and participation in archival processes of appraisal and description have been marshaled from a variety of angles (Gilliland & McKemmish, 2014 ; Huvila, 2015) that mirror the goals of participatory design.

In terms of addressing web archives, and specifically the needs of archival appraisal, there has been some emerging work, but much remains. Goldman (2018) has begun studying how digital preservation frameworks such as the Open Archives Information System (OAIS), and principles such as Lots Of Copies Keeps Stuff Safe (LOCKSS), can introduce overheads that negatively impact the environment, and if widely practiced, could undermine the long-term sustainability of data archives. Christen, Merrill, & Wynne (2017) has explored how values expressed in open access licenses and metadata standards can be at cross purposes with the ethical curation and access to cultural heritage materials. These risks can be mitigated by engagement with local communities and community archives practices. Furthermore, in the wake of the 2016 US Presidential Election the Environmental Data Governance Initiative worked to preserve at risk data sets produced by the Environmental Protection Agency and other US federal agencies. This work was articulated as a web archiving project, and was performed by volunteer archivists, librarians, technologists, scientists and concerned citizens (Schlanger, 2017). Participants have reflected on how this *data justice* work fits into an evidence based accountability framework (Dillon et al., 2017). But understanding how these and other emerging data justice projects (Taylor, 2017) take place both in and through the infrastructure of the web, and express implicit or explicit appraisal decision is work that remains.

III. Methodology

In the previous chapter I outlined the various ways that archivists have historically conceived of *appraisal*, including its more recent expression in the domains of digital curation and digital preservation. This review was a necessary first step in beginning to pose my specific research question which concerns how appraisal is being enacted in web archives. What emerged from this review is that it is useful to understand archival appraisal not only as a product of governance (a concern of organizations, institutions and the state), but also through what Foucault calls *governmentality*, or the microphysics of power. While it is increasingly accepted to conceive of archives as a technology of power (Jimerson, 2009; Zinn, 1977), the exercise of archival agency is found not only in the top-down operations of governments and institutions, but also in every day practices at multiple, dispersed sites spread throughout society: in workplaces, families, communities and *collectives* of all kinds, and even by individuals (McKemmish & others, 1996). The concept of governmentality allows us to investigate these sites of archival appraisal as part of a continuum. Governmentality shifts our focus to the *practices* of appraisal and how they can enact power relations and resistance as well as social justice and collective memory (Jacobsen et al., 2013; Punzalan & Caswell, 2016).

An additional related theme that the previous chapter introduced is that the question of how appraisal operates in *web* archives requires an analysis that accounts for the role that technology plays in these practices. Of course, technology has always been an active agent in record keeping. This in itself is nothing new. And yet we are only a few decades into a continuing and profound shift to electronic records, in which centuries old archival technologies of paper have given way to a digital regime of databases, data processing, and computer networks (Bearman, 1989). As Hedstrom (1991) outlines in her *Framework for Research on Electronic Records*, the field of Science and Technology Studies (STS) provides a useful historical and theoretical framework for understanding archives as sociotechnical systems, without privileging either social or technical explanations. What gets deemed *archival*, and the very meaning of *preservation* and *access* are forged in the de-

sign and use of information processing systems, and attendant standardization practices. Similarly, the practices of *appraisal* in the context of the web, and even the definition of what constitutes a *web archive*, are the result of sociotechnical processes in which our design of algorithms, data formats and interfaces both produce and are produced by web technologies. STS provides both a theoretical and methodological framework for researching the ontological dimensions of appraisal.

So, my research project theorizes appraisal in web archives as a *sociotechnical practice*, where these practices can be understood using the framework of *governmentality*. This chapter outlines a methodology for investigating the question of how archivists decide what to archive from the web, and what significance these findings have for larger questions about memory, technology and the web. The processes of how we choose to remember with the web are critical for ascertaining what our web archives *mean* (Ketelaar, 2001 ; Maemura et al., 2018). However, it is important to remember that us humans are not the only readers of the records in web archives. The "meaning" of web archives is increasingly found in human-computer assemblages that are designed to "learn", or establish statistical patterns in archival data, so that very real decisions can be made in the world (boyd & Crawford, 2012 ; Mackenzie, 2017). Web archives are now assembled by machines, in order to be read by machines, for very human purposes. The philosophical question of whether machines actually "understand" such records or not is increasingly irrelevant, but the stakes for our understanding how records come to be in our web archives could not be higher.

Algorithms as Culture

As sociotechnical assemblages, web archives are complex sites where humans and computers collaborate to select web content for an archive,. Indeed, on close inspection it can often be difficult to untangle these relations and clearly demarcate where one begins and the other ends. Web archives are infrastructures of software and hardware that have been crafted by archivists and technologists over the past two decades to do something we've come to call "web archiving". Archivists and other types of users interact with these systems to collect particular regions of the web, and these

systems continue to change to help the further achievement of those goals. Finding techniques and methods for bringing these blurry boundaries and knotted relations of web archiving into focus is the task at hand.

Instead of being purely virtual, disembodied and abstracted, *algorithmic* processes are the result of distinct social and material practices, in very specific and highly contingent settings (Geiger, 2014). I have introduced the term *algorithm* here because, as we will see, the appraisal of web content is increasingly automated by crawling procedures or heuristics. I also want to use the methods of *Critical Algorithm Studies* to help answer the question of how appraisal operates in web archives. Critical Algorithm Studies centers the study of algorithms as material artifacts that participate in particular social settings.

Algorithms have traditionally been considered the sole domain of the computer scientist. For example here is a conventional definition of the term algorithm found in a popular undergraduate computer science textbook:

> … any well-defined computational procedure that takes some value, or set of values, as *input* and produces some value, or set of values, as *output*. An algorithm is thus a sequence of computational steps that transform the *input* into the *output*. We can also view an algorithm as a tool for solving a well specified computational problem. The statement of the problem specifies in general terms the desired input/output relationship. The algorithm describes a specific computational procedure for achieving that input/output relationship. (Cormen, Leiserson, Rivest, & Stein, 2009, p. 5)

Notice how this definition works to scope the concept of the algorithm to a particular setting: where an algorithm is a "tool", that helps solve a "problem" in terms of set of "inputs" and "outputs". The "problem" is computationally constrained, and presented out of band, almost as a given. The algorithmic problem is not to be questioned, problematized or investigated outside of its calculative dimensions – it is to be solved, almost like a puzzle. But how is the problem defined? How will the solution be recognized? How do the shape and content of inputs and outputs change as conceptions

of the problem and the solution are updated as the algorithm is implemented? These questions of process rather than simply processing push at the procedural definition of the algorithm, and invite us to consider how algorithms are positioned in larger sociotechnical settings that include material constraints such as energy, space and time, as well as the goals and politics of people, groups and organizations.

Seaver argues that algorithmic systems aren't simply black boxes, or sites that can be opened, analyzed and understood (Seaver, 2017). The study of algorithms requires a methodological approach that recognizes how algorithms are deployed in particular social settings, as part of specific material practices, that operate in the world as part of *culture*, in addition to (not in place of) their manifestation as computational processes. In a memorable turn of phrase Seaver advises, "If you cannot see a human in the loop, you just need to look for a bigger loop."

Algorithmic systems can be quite difficult to study because they don't live in a particular place, and often aren't known by a single individual. Algorithmic processes can be distributed between teams, systems and workflows that combine computation with people and their lived experiences. For Seaver the study of algorithmic systems in all these dimensions (including the computational) requires an approach that recognizes algorithms *as* culture, rather than algorithms operating simply *in* culture, both of which point to the use of *ethnography* as both method and theoretical orientation.

Studying Up

Ethnography usually involves some form of participant observation of people in a particular setting in order to understand social worlds, using data collection tools such as field notes, interviews and surveys. However the study of algorithmic systems is often challenged by practical barriers to data collection using traditional participant observation techniques. Attempts to understand algorithmic processes often put the researcher right into the beating heart of an organization, where information can often be guarded for competitive reasons, or because the information itself could allow the company's services to be subverted, gamed or otherwise critiqued. Technical jargon and expert

knowledge distributed across individuals in an organization can act as a foil for understanding the dimensions of algorithms. The ethnographer finds themself in a position of what Nader calls "studying up", where the researcher is disadvantaged as they attempt to access a site of power (Nader, 1972). For these reasons Seaver suggests that researchers employ Gusterson's method of *polymorphous engagement*, or "interacting with informants across a number of dispersed sites … collecting data eclectically from a disparate array of sources in many different ways" (Gusterson, 1997). Seaver refers to this using the shorthand of "scavenging" which is also echoed by Kitchin who suggests ethnography for the study of algorithms as sociotechnical *assemblages* using (by necessity) a wide variety of sources:

> Interviews and ethnographies of coding projects, and the wider institutional apparatus surrounding them (e.g., management and institutional collaboration), start to produce such knowledge, but they need to supplemented with other approaches, such as a discursive analysis of company documents, promotional/industry material, procurement tenders and legal and standards frameworks; attending trade fairs and other inter-company interactions; examining the practices, structures and behaviour of institutions; and documenting the biographies of key actors and the histories of projects. Kitchin (2016)

Ethnography of algorithmic systems is also challenged by the nature of observational data that the researcher encounters. Algorithms are created by people, but they are also performed as computational systems that, as our previous definition made clear, are a function of a particular set of inputs and outputs. These inputs and outputs happen in time and leave traces of their operation: be they transaction logs, database entries, status messages and the like. Geiger and Ribes method of *trace ethnography* is useful in these settings because it assists in "revealing the often invisible infrastructure that underlie routinized activities, allowing researchers to generate highly empirical accounts of network-level phenomena without having to be present at every node" (Geiger & Ribes, 2011). This opens the ethnographer up to studying data traces in files and databases, much as they might

also study participants use of language.

Relations All the Way Down

Before diving into the specific details of how I am proposing to researching appraisal in web archives it is worth briefly situating Critical Algorithm Studies in the broader tradition in Science and Technology Studies, which will be helpful for grounding my own research project.

As we have just discussed, algorithms operate as discrete computational processes that transform input into output; but at the same time they also function as part of an information *infrastructure*, in which the algorithms are defined in terms of the practices and relations they are embedded in. Susan Leigh Star, one key theorist of Infrastructure Studies, stresses that infrastructures are *relational*, that they are not constituted by a particular set of objects or artifacts so much as they are embedded in practices that happen in time:

> … we hold that infrastructure is a fundamentally relational concept. It becomes infrastructure in relation to organized practices. Within a given cultural context, the cook considers the water system a piece of working infrastructure integral to making dinner; for the city planner, it becomes a variable in a complex equation. Thus we ask, when–not what–is an infrastructure. (Star & Ruhleder, 1996, p. 4)

This attention to relations in infrastructure that Star identifies are a concern of Actor Network Theory (ANT) (Latour, 2005), or material semiotics more generally (Law (2009)). For Latour, ANT "define[s] the social not as a special domain, a specific realm, or a particular sort of thing, but only as a very peculiar movement of re-association and reassembling" (p. 7). These movements are identified by a broadening of the types of actors that can participate in relations to include so called non-humans as *actants*, which multiplies the types of and complexity of relations. Latour uses the term *symmetry* as a shorthand for this idea of granting both humans and non-humans with measures of agency. These relations are made durable in material, strategy, discourse, and perfor-

mance, which allows them to persist in time, or as Latour memorably says "technology is society made durable" (Latour, 1990).

In addition to tracing what is, and how these networks of relations are made durable, it is also critically important to consider the point of departure that is chosen for these descriptions: to factor in the role of power in whose stories we tell, and to account for how things "could have been otherwise" (Star, 1990). Foucault's analysis of the micro-physics of power and governmentality clearly connects here. Yet these are questions not only of politics, but of ontology, recognizing that ontology can function as a totalizing force, but also exists as what Mol calls a *multiplicity*, where distinct practices generate new ways of being that coordinate in time (Mol, 2002).

Two Projects

With these theoretical and methodological foundations in mind it is helpful to now return to my motivating research question before diving into how I plan to answer it. The general question of appraisal in web archives actually cleaves into two interrelated sub-questions:

RQ1: How is appraisal currently being enacted in web archives?

RQ2: How do definitions of what constitutes a *web archive* relate to the practice of appraisal?

RQ1 is *empirical* since it involves observation and data gathering to describe how appraisal is being performed in the field. The emphasis on *performance* here is intentional, since my purpose is not to make generalized and valid claims about of *all* appraisal practices in web archives. Instead I want to gain qualitative insight into the factors that motivate a discrete set of archivists in their decisions to collect content from the web, in order to better understand how appraisal is functioning in web archives.

RQ2 is *ontological* because it recognizes that decisions of what to collect from the web fundamentally define what a web archive is, and what it will become. At the same time, RQ2 also allows for appraisal decisions to be shaped by the material dimensions of web archives, or what web archives

are. To borrow a term from Jasanoff, the infrastructures of web archives and their affordances could (we will see) be *coproduced* by practices of appraisal (Jasanoff, 2006). For example standards can constrain and promulgate particular types appraisal practices in web archives. But these standards in turn reflect decisions made in the design of web archiving systems. Rather than one being a cause of the other, design decisions about how to assemble software, hardware, algorithms and computational resources in web archives are themselves a form of appraisal practice.

Law and Lien refer to this dual lens expressed by my two research questions as *empirical ontology* where questions of ontology are not concerned with describing a cosmology, uni-verse, or single dominant reality; but rather aim to describe how practices, or networks of actors, generate world views or multiple *ontologies* (Law & Lien, 2012). As previously discussed above, the rich descriptions generated by ethnography are particularly well suited to the task of exploring these questions of practice, especially in the context of algorithms and data as culture. I will conduct two case studies to help answer these questions. Rather than each study being tied to one of the research questions, these two sites provide a zooming function, where the practices of multiple informants at multiple sites are analyzed, before diving into a year long field study at a specific location, in order to reassemble the relations discovered by zooming out again (Nicolini, 2009).

Appraisal Talk

In the first research project I will conduct a series of semi-structured, ethnographic interviews with practicing archivists and archives adjacent actors in order to hear how they decide what parts of the web to collect. An initial set of practitioners will be selected from a list of attendees at the Web Archives conference in Ann Arbor, Michigan on November 12-13, 2015 ("Web archives 2015," 2018). This conference brought together a diverse set of individuals involved in web archiving: archivists, researchers who used web archives, administrators of museums and libraries, open source software developers, and vendor representatives. A semi-structured set of interview questions will be developed, which will be drawn upon to guide participants into sharing information

about how they select content from the web for their archives. But as a practitioner in the web archiving field myself, these will be conversational, unstructured, and will seek to elaborate participant's stories about how they have collected from the web. Interviews will be recorded, transcribed, and coded using inductive thematic analysis to derive key factors that influence and shape appraisal practices in web archives.

After identifying emergent themes using MAXQDA[1] I will use Critical Discourse Analysis (CDA) to examine in detail the words practitioners use, the conventions they have stabilized, the context they share, how they learn from each other in communities of practice (Lave & Wenger, 1991), and the political work that these communicative practices perform. CDA offers a theoretical framework grounded in critical theory for analyzing how participants' use of language reflects identity formation, figured worlds and social relations, while also addressing the larger sociotechnical context in which practice takes place. CDA helps examine how language use connects with issues of ideology and power, which are particularly relevant when considering archival appraisal as an inherently political act. The sociolinguist James Paul Gee has noted that "there are solid linguistic, even grammatical grounds, on which to argue that all language-in-interaction is inherently political." (Gee, 2004)

Implementing Appraisal

In my second case study I will conduct a year long field study at the National Software Reference Library (NSRL) at the National Institute for Standards and Technology (NIST) in Gaithersburg, Maryland. Since 1999 the NSRL has maintained the largest public repository of known software, out of which it publishes the Reference Data Set (RDS) which provides file profiles and file signatures for use by law enforcement and other organizations involved with computer forensic investigations. While not being a web archive in name, the NSRL contains software packages that have been actively collected from the web as data. This data is collected from the web using a variety of

[1] https://www.maxqda.com/

techniques and automated tools, and then is stored, indexed and processed in order to publish the RDS that documents the fixities of hundreds of millions of computer files. Crucially, the NSRL needs to decide what software to collect, and in doing so must balance the needs of their customers with the architectural constraints of their repository infrastructure, staff and funding mandate. The NSRL has developed practices, technologies and standards, because of its unique role, and placement within the federal government.

As part of my field study I will document my observations and experiences as a participant observer in the NSRL using field notes gathered during weekly staff meetings, and interactions with NSRL staff as they go about their work. NIST's online collaborative environments (Slack, Bugzilla) will be analyzed using trace ethnography (Geiger, 2014) in order to identify actors, systems and practices that are relevant for the process of appraisal. Techniques from Critical Algorithm and Data Studies will be used to analyze available source code, and database schemas (Thomer, Weber, & Twidale, 2018) to trace appraisal activities in the infrastructure of the NSRL. This data will be supplemented with in depth interviews with NIST employees that will provide additional materials for tracing the shape of appraisal in an infrastructural setting which does not necessarily fit with the traditional conception of a web archive. These interviews will focus on the personal stories of how these employees came to work at NIST with the NSRL, and explore how these individual stories work to form organizational narrative, memory and identity (Linde, 2009) that happen over long stretches of time (Cohn, 2013).

The resulting field notes and observational data will be analyzed using techniques from Actor Network Theory (ANT) which trace the activities of human and non-human agents in the performance of appraisal in the NSRL. Interviews will be coded using relevant themes, entities and relations that were identified in the previous CDA study. Field notes will be consulted in order to navigate what questions to ask of my informants, and how to relate to them, since they will be unstructured interviews. In addition, my field notes will help me analyze how my role in the NSRL, and my perspective on it and my research change during my time spent as a participant observer in the

NSRL. Reflexivity is a key component since the ethnographer become the instrument of research, in the stories they choose to tell.

Design

These two sites will allow me to connect how archivists talk about appraisal and also how they *do* appraisal. Both activities are integral parts of appraisal as a practice. The purpose is to gain greater insight into how data is collected from the web, because the ways in which these data lakes are assembled have crucial social and political implications for how that data is to be used. It is tempting to see a direct line of causation between how data is to be used, to how it is collected. However sociotechnical theory suggests that this line may not be as direct as we think, and may feed back on itself in important ways.

It also may be tempting to suggest that these two case studies will provide insights that can be used in the design of web archiving systems. While I don't want to rule this out entirely, it is not explicit purpose of my research project to derive new designs for web archiving systems. Implications for design may not be practical given how different these two case studies are. It is much more likely that implications for theory and/or method might be achieved. I hope that this project will shed light on how web archives themselves are conceived, and how a more enlarged and theorized concept of web archiving practice can enrich information studies pedagogy. As Dourish notes in his influential piece critiquing the role of ethnography in design:

> It is practice that gives form and meaning to technology; the focus of ethnography is the ways in which practice brings technology into being. From this perspective … we might suggest that what ethnography problematizes is not the setting of everyday practice, but the practice of design … What matters is not simply what those implications are; what matters is why, and how they were arrived at, and what kinds of intellectual (and moral and political) commitments they embody, and what kinds of models they reflect. (Dourish, 2006)

IV. References

Abreu, A., & Acker, A. (2013). Context and collection: A research agenda for small data. In *IConference proceedings* (pp. 549–554).

Acker, A. (2016). Radical appraisal practices and the mobile forensic imaginary. *Acrhive Journal*, (5). Retrieved from http://www.archivejournal.net/issue/5/archives-remixed/radical-appraisal-practices-and-the-mobile-forensic-imaginary/

Andersen, H. (2013). A website owner's practice guide to the wayback machine. *J. On Telecomm. & High Tech. L.*, *11*, 251. Retrieved from http://www.jthtl.org/content/articles/V11I1/JTHTLv11i1_Andersen.PDF

Anderson, K. D. (2011). *Appraisal learning networks: How university archivists learn to appraise through social interaction* (PhD thesis). University of California, Los Angeles.

Arnold, T., & Sampson, W. (2014). Preserving the voices of revolution: Examining the creation and preservation of a subject-centered collection of tweets from the eighteen days in egypt. *The American Archivist*, *77*(2), 510–533.

Aronson, J. D. (2017). Preserving human rights media for justice, accountability, and historical clarification. *Genocide Studies and Prevention: An International Journal*, *11*(1).

Baeza-Yates, R., Marin, M., Castillo, C., & Rodriguez, A. (2005). Crawling a country: Better strategies than breadth-first for web page ordering. In *Proceedings of the 14th international conference on world wide web*. Chiba, Japan.

Bailey, J. (2013). Disrespect des fonds: Rethinking arrangement and description in born-digital archives. *Archive Journal*, *3*.

Bandaru, K., & Patiejunas, K. (2015). Under the hood: Facebook's cold storage system. Facebook. Retrieved from https://code.fb.com/production-engineering/-under-the-hood-facebook-s-cold-storage-system-/

Banos, V., & Manolopoulos, Y. (2015). A quantitative approach to evaluate website archivability using the clear+ method. *International Journal on Digital Libraries*, 1–23. https://doi.org/10.1007/s00799-015-0144-4

Barth, A. (2011). *HTTP state management mechanism* (No. 6265). Internet Engineering Task Force. Retrieved from https://tools.ietf.org/html/rfc6265

Bastian, J. A. (2001). Taking custody, giving access: A postcustodial role for a new century. *Archivaria*, *53*, 76–94.

Bearman, D. (1989). Archival methods. *Archives and Museum Informatics*, *3*(1). Retrieved from http://www.archimuse.com/publishing/archival_methods/

Ben-David, A., & Amram, A. (2018). The internet archive and the socio-technical construction of historical facts. *Internet Histories*, *2*(1-2), 179–201.

Benjamin, W. (1999). The work of art in the age of mechanical reproduction. In *Illuminations: Essays and reflections* (pp. 211–244). Pimlico.

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. Yale University Press.

Berners-Lee, T., & Fischetti, M. (2000). *Weaving the web: The original design and ultimate destiny of the world wide web by its inventor*. San Francisco: Harper.

Booms, H. (1987). Society and the formation of a documentary heritage: Issues in the appraisal of archival sources. *Archivaria*, *24*(3), 69–107. Retrieved from http://journals.sfu.ca/archivar/index.php/archivaria/article/view/11415/12357

Bourdieu, P. (1977). *Outline of a theory of practice* (Vol. 16). Cambridge University Press.

Bowker, G. C. (2005). *Memory practices in the sciences* (Vol. 205). Cambridge, MA: MIT Press.

boyd, danah, & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–

679.

Bratton, B. (2016). *The stack*. MIT Press.

Brichford, M. J. (1977). *Archives & manuscripts: Appraisal & accessioning*. Society of American Archivists.

Brothman, B. (2001). The past that archives keep: Memory, history, and the preservation of archival records. *Archivaria*, *51*, 48–80.

Brown, R. (1991). Records acquisition strategy and its theoretical foundation: The case for a concept of archival hermeneutics. *Archivaria*, *33*.

Brügger, N. (2017). *The web as history*. (N. Brügger & R. Schroeder, Eds.). UCL Press. Retrieved from http://discovery.ucl.ac.uk/1542998/

Buckland, M. K. (1991). Information as thing. *JASIS*, *42*(5), 351–360.

Burton, M. (2015). *Blogs as infrastructure for scholarly communication* (PhD thesis). University of Michigan.

Caswell, M., & Cifor, M. (2016). From human rights to feminist ethics: Radical empathy in the archives. *Archivaria*, *82*, 23–43.

Ceglowski, M. (2011, May). Remembrance of links past. Retrieved from https://blog.pinboard.in/2011/05/remembrance_of_links_past/

Christen, K., Merrill, A., & Wynne, M. (2017). A community of relations: Mukurtu hubs and spokes. *D-Lib Magazine*, *23*(5/6). Retrieved from http://www.dlib.org/dlib/may17/christen/05christen.html

Chun, W. H. K. (2011). *Programmed visions: Software and memory*. MIT Press.

Cifor, M. (2016). Affecting relations: Introducing affect theory to archival discourse. *Archival Science*, *16*(1), 7–31.

Cifor, M., Caswell, M., Migoni, A. A., & Geraci, N. (2018). "What we do crosses over into activism": The politics and practice of community archives. *Acrhives and Public History*, *40*(2), 69–95.

Cobb, J., Pearce-Moses, R., & Surface, T. (2005). ECHO DEPository Project. In *Archiving 2005, final program and proceedings*.

Cohn, M. (2016). Convivial decay: Entangled lifetimes in a geriatric infrastructure. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing* (pp. 1509–1521). Association for Computing Machinery.

Cohn, M. L. (2013). *Lifetimes and legacies: Temporalities of sociotechnical change in a long-lived system.* (PhD thesis). University of California at Irvine.

Cook, T. (1993). The concept of the archival fonds in the post-custodial era: Theory, problems and solutions. *Archivaria*, *35*, 24–37.

Cook, T. (1994). Electronic records, paper minds: The revolution in information management and archives in the post/custodial and post/modernist era.[Based on a presentation delivered by the author during his november 1993 australian tour.]. *Archives and Manuscripts*, *22*(2), 300–328.

Cook, T. (2004). Macro-appraisal and functional analysis: Documenting governance rather than government 1. *Journal of the Society of Archivists*, *25*(1), 5–18.

Cook, T. (2005). Macroappraisal in theory and practice: Origins, characteristics, and implementation in canada, 1950–2000. *Archival Science*, *5*(2-4), 101–161.

Cook, T. (2011). We are what we keep; we keep what we are: Archival appraisal past, present and future. *Journal of the Society of Archivists*, *32*(2), 173–189.

Cook, T. (2013). Evidence, memory, identity, and community: Four shifting archival paradigms. *Archival Science*, *13*(2-3), 95–120.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms* (3rd

ed.). MIT Press.

Cox, R., & Samuels, H. (1988). The archivist's first responsibility: A research agenda to improve the identification and retention of records of enduring value. *The American Archivist*, *51*(1-2), 28–42.

Cubitt, S. (2016). *Finite media: Environmental implications of digital technologies*. Duke University Press.

Diakopoulos, N. (2014). *Algorithmic accountability reporting: On the investigation of black boxes*. Tow Center for Digital Journalism, Columbia University.

Dillon, L., Walker, D., Shapiro, N., Underhill, V., Martenyi, M., Wylie, S., … Initiative, G. (2017). Environmental data justice and the trump administration: Reflections from the environmental data and governance initiative. *Environmental Justice*, *10*(6), 186–192.

DiSalvo, C., Sengers, P., & Brynjarsdóttir, H. (2010). Mapping the landscape of sustainable hci. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1975–1984). ACM.

Dobra, A., & Fienberg, S. E. (2004). How large is the world wide web? In *Web dynamics* (pp. 23–43). Springer.

Dourish, P. (2006). Implications for design. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 541–550). ACM. Retrieved from http://www.dourish.com/publications/2006/implications-chi2006.pdf

Dourish, P. (2017). *The stuff of bits: An essay on the materialities of information*. MIT Press.

Duncan, S., & Blumenthal, K.-R. (2016). A collaborative model for web archiving ephemeral art resources at new york art resources consortium (nyarc). *Art Libraries Journal*, *41*(2), 116–126.

Duranti, L. (1994). The concept of appraisal and archival theory. *The American Archivist*, 328–344.

Duranti, L. (2010). Concepts and principles for the management of electronic records, or records

management theory is archival diplomatics. *Records Management Journal*, *20*(1), 78–95.

Eastwood, T. (2002). Reflections on the goal of archival appraisal in democratic societies. *Archivaria*, *1*(54). Retrieved from http://journals.sfu.ca/archivar/index.php/archivaria/article/viewFile/12855/14080

Edwards, P., Mayernik, M. S., Batcheller, A., Bowker, G., & Borgman, C. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, 0306312711413314.

Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. MIT Press.

Ehn, P. (1988). *Work-oriented design of computer artifacts* (PhD thesis). Arbetslivscentrum.

Ernst, W. (2013). *Digital memory and the archive*. University of Minnesota Press.

Felt, U. (2017). *Handbook of science and technology studies*. (U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr, Eds.) (4th ed.). MIT Press.

Fenlon, K. S. (2017). *Thematic research collections: Libraries and the evolution of alternative scholarly publishing in the humanities.* (PhD thesis). University of Illinois Urbana-Champaign.

Fielding, R. (2000). *Architectural styles and the design of network-based software architectures* (PhD thesis). University of California, Irvine. Retrieved from https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm

Fielding, R., Nottingham, M., & Reschke, J. (2014). *Hypertext transfer protocol (http/1.1): Caching* (No. 7234). Internet Engineering Task Force. Retrieved from https://tools.ietf.org/html/rfc7234

Fielding, R. T. (1994). Maintaining distributed hypertext infostructures: Welcome to momspider's web. *Computer Networks and ISDN Systems*, *27*(2), 193–204.

Flinn, A. (2007). Community histories, community archives: Some opportunities and challenges 1. *Journal of the Society of Archivists*, *28*(2), 151–176.

Flinn, A. (2011). Archival activism: Independent and community-led archives, radical public history and the heritage professions. *InterActions: UCLA Journal of Education and Information Studies*, *7*(2).

Flinn, A., Stevens, M., & Shepherd, E. (2009). Whose memories, whose archives? Independent community archives, autonomy and the mainstream. *Archival Science*, *9*(1-2), 71–86.

Foscarini, F. (2017). Archival appraisal in four paradigms. In H. MacNeil & T. Eastwood (Eds.), *Currents of archival thinking* (pp. 107–134). Libraries Unlimited.

Foucault, M. (1982). The subject and power. *Critical Inquiry*, *8*(4), 777–795.

Foucault, M. (1986). Of other spaces. *Diacritics*, *16*(1), 22–27.

Foucault, M. (1991). The foucault effect: Studies in governmentality. In (pp. 87–104). University of Chicago Press.

Foucault, M., Davidson, A. I., & Burchell, G. (2008). *The birth of biopolitics: Lectures at the Collège de France, 1978-1979*. Springer.

Galloway, A. R. (2004). *Protocol: How control exists after decentralization*. Cambridge: MIT Press.

Gee, J. P. (2004). Discourse analysis: What makes it critical? In R. Rogers (Ed.), *An introduction to critical discourse analysis in education* (pp. 49–80). Routledge.

Geiger, R. S. (2014). Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society*, 342–356. Retrieved from http://www.tandfonline.com/doi/full/10.1080/1369118X.2013.873069

Geiger, R. S., & Ribes, D. (2011). Trace ethnography: Following coordination through documentary practices. In *44th hawaii international conference on system sciences* (pp. 1–10). IEEE. Retrieved from http://www.stuartgeiger.com/trace-ethnography-hicss-geiger-ribes.pdf

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University

of California Press.

Gillespie, T. (2010). The politics of platforms. *New Media & Society*, *12*(3), 347–364.

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gillespie, T., & Seaver, N. (2015). Critical algorithm studies: A reading list. Retrieved from https://socialmediacollective.org/reading-lists/critical-algorithm-studies/

Gilliland, A., & McKemmish, S. (2014). The role of participatory archives in furthering human rights, reconciliation and recovery. *Atlanti*, *24*(1), 79–88.

Gitelman, L. (Ed.). (2013). *"Raw data" is an oxymoron.* MIT Press.

Gitelman, L. (2014). *Paper knowledge: Toward a media history of documents*. Duke University Press.

Goldman, B. (2018). It's not easy being green(e): Digital preservation in the age of climate change. In *Archival values: Essays in honor of Mark Greene*. Society of American Archivists. Retrieved from https://scholarsphere.psu.edu/concern/generic_works/bvq27zn11p

Gomes, D., Freitas, S., & Silva, M. J. (2006). Design and selection criteria for a national web archive. In *Research and advanced technology for digital libraries* (pp. 196–207). Springer.

Gossen, G., Demidova, E., & Risse, T. (2015). The iCrawl wizard – supporting interactive focused crawl specification. In *Proceedings of the 37th european conference on information retrieval*. Retrieved from http://www.l3s.de/~gossen/publications/gossen_et_al_ecir_2015.pdf

Gracy, K. (2007). Moving image preservation and cultural capital. *Library Trends*, *56*(1), 183–197.

Graham, S., & Thrift, N. (2007). Out of order understanding repair and maintenance. *Theory, Culture & Society*, *24*(3), 1–25.

Gusterson, H. (1997). Studying up revisited. *PoLAR: Political and Legal Anthropology Review*,

20(1), 114–119.

Ham, F. (1984). Archival choices: Managing the historical record in an age of abundance. *The American Archivist*, *47*(1), 11–22.

Ham, F. G. (1975). The archival edge. *The American Archivist*, *38*(1), 5–13. Retrieved from http://americanarchivist.org/doi/pdf/10.17723/aarc.38.1.7400r86481128424

Ham, F. G. (1981). Archival strategies for the post-custodial era. *The American Archivist*, *44*(3), 207–216.

Harris, V. (2002). The archival sliver: Power, memory, and archives in South Africa. *Archival Science*, *2*(1-2), 63–86.

Harvey, R., & Thompson, D. (2010). Automating the appraisal of digital materials. *Library Hi Tech*, *28*(2), 313–322.

Hedstrom, M. (1991). Understanding electronic incunabula: A framework for research on electronic records. *The American Archivist*, *54*(3), 334–354.

Hedstrom, M. (2002). Archives, memory, and interfaces with the past. *Archival Science*, *2*(1-2), 21–43.

Hennessey, J., & Ge, S. X. (2013). A cross disciplinary study of link decay and the effectiveness of mitigation techniques. *BMC Bioinformatics*, *14*(14).

Hitchcock, S., Brody, T., & Hey, J. M. N. (2007). Digital preservation service provider models for institutional repositories. *D-Lib Magazine*, *13*(5/6). Retrieved from http://www.dlib.org/dlib/may07/hitchcock/05hitchcock.html

Hoskins, A. (2018). Digital memory studies: Media pasts in transition. In A. Hoskins (Ed.) (pp. 85–109). Routledge.

Hu, T.-H. (2015). *A prehistory of the cloud*. MIT Press.

Huvila, I. (2008). Participatory archive: Towards decentralised curation, radical user orientation, and broader contextualisation of records management. *Archival Science*, *8*(1), 15–36.

Huvila, I. (2015). The unbearable lightness of participating? Revisiting the discourses of 'participation' in archival literature. *Journal of Documentation*, *43*, 29–41.

Ivanov, A. O. (2017). Practice theory: A new approach for archival and recordkeeping research. *Records Management Journal*, *27*(2), 104–124.

Jackson, S. J. (2014). Media technologies: Essays on communication, materiality and society. In P. Boczkowski & K. Foot (Eds.) (pp. 221–239). MIT Press. Retrieved from http://sjackson.infosci.cornell.edu/RethinkingRepairPROOFS(reduced)Aug2013.pdf

Jackson, S. J., & Buyuktur, A. (2014). Who killed waters? Mess, method, and forensic explanation in the making and unmaking of large-scale science networks. *Science, Technology, & Human Values*, *39*(2), 285–308.

Jackson, S. J., Gillespie, T., & Payette, S. (2014). The policy knot: Re-integrating policy, practice and design in CSCW studies of social computing. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 588–602). Association for Computing Machinery.

Jackson, S. J., & Kang, L. (2014). Breakdown, obsolescence and reuse: HCI and the art of repair. In. CHI; Association of Computing Machinery. Retrieved from http://sjackson.infosci.cornell.edu/Jackson&Kang_BreakdownObsolescenceReuse(CHI2014).pdf

Jacobsen, T., Punzalan, R. L., & Hedstrom, M. L. (2013). Invoking collective memory: Mapping the emergence of a concept in archival science. *Archival Science*, *13*(2-3), 217–251.

Jasanoff, S. (2006). *States of knowledge: The co-production of science and the social order*. Routledge.

Jenkinson, H. (1922). *A manual of archive administration including the problems of war archives*

and archive making. Clarendon Press.

Jimerson, R. (2013). Archivists and social responsibility: A response to mark greene. *The American Archivist*, *76*(2), 335–345.

Jimerson, R. C. (2009). *Archives power: Memory, accountability, and social justice*. Society of American Archivists.

Kahle, B. (2007). Universal access to all knowledge. *The American Archivist*, *70*(1), 23–31.

Karasti, H., & Blomberg, J. (2018). Studying infrastructuring ethnographically. *Computer Supported Cooperative Work (CSCW)*, 1–33.

Kelty, C. M. (2008). *Two bits: The cultural significance of free software*. Duke University Press. Retrieved from http://twobits.net/

Kensing, F., & Blomberg, J. (1998). Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW)*, *7*(3-4), 167–185.

Ketelaar, E. (2001). Tacit narratives: The meanings of archives. *Archival Science*, *1*(2), 131–141.

Kirschenbaum, M. G. (2008). *Mechanisms: New media and the forensic imagination*. MIT Press.

Kitchin, R. (2016). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 1–16.

Kitchin, R., & Lauriault, T. P. (2014). *Towards critical data studies: Charting and unpacking data assemblages and their work* (No. The Programmable City Working Paper 2). The Programmable City; The Programmable City Working Paper 2. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112

Kittler, F. A. (1999). *Gramophone, film, typewriter*. Stanford University Press.

Klein, M., Balakireva, L., & Van de Sompel, H. (2018). Focused crawl of web archives to build event collections. *arXiv Preprint arXiv:1804.01603*.

Kosnik, A. D. (2016). *Rogue archives: Digital cultural memory and media fandom*. MIT Press.

Kuny, T. (1997). A digital dark ages? Challenges in the preservation of electronic information. In *Proceedings of the 63rd international federation of library associations*. Retrieved from https://archive.ifla.org/IV/ifla63/63kuny1.pdf

Kuutti, K., & Bannon, L. J. (2014). The turn to practice in HCI: Towards a research agenda. In *Proceedings of the 32nd annual ACM Conference on Human Factors in Computing Systems* (pp. 3543–3552). Association for Computing Machinery. Retrieved from http://dl.acm.org/citation.cfm?id=2557111

Lampland, M., & Star, S. L. (2009). *Standards and their stories: How quantifying, classifying, and formalizing practices shape everyday life*. Cornell University Press.

Lasfrgues, F., Oury, C., & Wendland, B. (2008). Legal deposit of the french web: Harvesting strategies for a national domain. In *Proceedings from IWAW '08: 8th International Workshop for Web Archiving*. Aarhus.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Latour, B. (1990). Technology is society made durable. *The Sociological Review*, *38*, 103–131.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Law, J. (2009). The new Blackwell companion to social theory. In B. S. Turner (Ed.) (pp. 141–158). Oxford: Wiley-Blackwell.

Law, J., & Lien, M. E. (2012). Slippery: Field notes in empirical ontology. *Social Studies of Science*, *43*(3), 363–378.

Lee, C. A. (Ed.). (2011). I, digital: Personal collections in the digital era. In *I, Digital: Personal Collections in the Digital Era. Chicago, IL: Society of American Archivists*. Society of American Archivists.

Lee, C. A., & Tibbo, H. (2011). Where's the archivist in digital curation? Exploring the possibilities through a matrix of knowledge and skills. *Archivaria*, *72*(Fall).

Lemke, T. (2019). *Foucault's analysis of modern governmentality: A critique of political reason*. (E. Butler, Trans.). Verso.

Lessig, L. (2006). *Code: Version 2.0*. New York: Basic Books.

Levy, D. (2001). Scrolling forward. In. Arcade.

Linde, C. (2009). *Working the past: Narrative and institutional memory*. Oxford University Press.

Mackenzie, A. (2017). *Machine learners: Archaeology of a data practice*. MIT Press.

Madhavan, J., Ko, D., Kot, Ł., Ganapathy, V., Rasmussen, A., & Halevy, A. (2008). Google's deep web crawl. *Proceedings of the VLDB Endowment*, *1*(2), 1241–1252.

Maemura, E., Worby, N., Milligan, I., & Becker, C. (2018). If these crawls could talk: Studying and documenting web archives provenance. *Journal of the Association for Information Science and Technology*, *69*(10), 1223–1233. Retrieved from https://tspace.library.utoronto.ca/handle/1807/82840

Maestri, L., & Wakkary, R. (2011). Understanding repair as a creative process of everyday design. In *Proceedings of the 8th acm conference on creativity and cognition* (pp. 81–90). ACM.

Manovich, L. (1999). Manovich: Database as a symbolic form. *Convergence: The International Journal of Research into New Media Technologies*.

Masanès, J. (2006a). Selection for web archives. In J. Masanès (Ed.), *Web archiving issues and methods*. Springer.

Masanès, J. (2006b). Web archiving methods and approaches: A comparative study. *Library Trends*, *54*(1), 72–90.

Mayer-Schönberger, V. (2011). *Delete: The virtue of forgetting in the digital age*. Princeton University Press.

McDonough, J., Olendorf, R., Kirschenbaum, M., Kraus, K., Reside, D., Donahue, R., … Rojo, S. (2010). *Preserving virtual worlds final report*. University of Illinois at Urbana-Champaign; Citeseer.

McKemmish, S., & others. (1996). Evidence of me. *Archives and Manuscripts*, *24*(1), 28.

McKemmish, S., Upward, F., & Reed, B. (2010). Records continuum model. In M. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences*. Taylor & Francis.

Meng, A., & DiSalvo, C. (2018). Grassroots resource mobilization through counter-data action. *Big Data & Society*, *5*(2), 2053951718796862.

Milligan, I., Ruest, N., & Lin, J. (2016). Content selection and curation for web archiving: The gatekeepers vs. The masses. In *Proceedings of the joint conference on digital libraries*. Retrieved from https://cs.uwaterloo.ca/~jimmylin/publications/Milligan_etal_JCDL2016.pdf

Mohr, G., Stack, M., Rnitovic, I., Avery, D., & Kimpton, M. (2004). Introduction to heritrix. In *4th international web archiving workshop*. Retrieved from https://webarchive.jira.com/wiki/download/attachments/5441/Mohr-et-al-2004.pdf

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Duke University Press.

Muller, S., Feith, J. A., & Fruin, R. (1940). *Manual for the arrangement and description of archives*. The H. W. Wilson Company.

Nader, L. (1972). Up the anthropologist: Perspectives gained from studying up. In *Reinventing anthropology*. Pantheon.

Nicolini, D. (2009). Zooming in and out: Studying practices by switching theoretical lenses and

trailing connections. *Organization Studies*, *30*(12), 1391–1418.

Nwala, A. C., Weigle, M. C., & Nelson, M. L. (2018). Bootstrapping web archive collections from social media. In *Proceedings of the 29th conference on Hypertext and Social Media*.

Ogden, J., Halford, S., & Carr, L. (2017). Observing web archives. In *Proceedings of WebSci'17*. Troy, NY: Association of Computing Machinery. Retrieved from https://eprints.soton.ac.uk/410123/

O'Toole, J. M. (2004). Back to the future: Ernst posner's archives in the ancient world. *American Archivist*, *67*(Fall/Winter), 161–175.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The pagerank citation ranking: Bringing order to the web.* Stanford Infolab.

Pearce-Moses, R., & Kaczmarek, J. (2005). An arizona model for preservation and access of web documents. *DTTP: Documents to the People*, *33*(1), 17–24.

Pinch, T. J. (1987). *Social construction of technological systems*. (W. E. Bijke & T. P. Hughes, Eds.). MIT Press.

Posner, E. (1972). *Archives in the ancient world*. Harvard University Press.

Postill, J. (2010). Theorising media and practice. In B. Bräuchler & J. Postill (Eds.). Berghahn Books.

Punzalan, R. (2014). Understanding virtual reunification. *The Library Quarterly*.

Punzalan, R. L. (2009). 'All the things we cannot articulate': Colonial leprosy archives and community commemoration. In J. A. Bastian & B. Alexander (Eds.), *Community archives: The shaping of memory*. Facet Publishing.

Punzalan, R. L., & Caswell, M. (2016). Critical directions for archival approaches to social justice. *Library Quarterly*, *86*(1), 25–42.

Ribes, D., & Finholt, T. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems*, *10*(5).

Risse, T., Dietze, S., Peters, W., Doka, K., Stavrakas, Y., & Senellart, P. (2012). Exploiting the social and semantic web for guided web archiving. In *Theory and practice of digital libraries* (pp. 426–432). Springer.

Rollason-Cass, S., & Reed, S. (2015). Living movements, living archives: Selecting and archiving web content during times of social unrest. *New Review of Information Networking*, *20*(1-2), 241–247.

Rosenthal, D. (2012, May). Let's just keep everything forever in the cloud. Retrieved from http://blog.dshr.org/2012/05/lets-just-keep-everything-forever-in.html

Rosner, D. K., & Ames, M. (2014). Designing for repair?: Infrastructures and materialities of breakdown. In *Proceedings of the 17th ACM conference on computer supported cooperative work & social computing* (pp. 319–331). ACM. Retrieved from http://people.ischool.berkeley.edu/~daniela/files/cscw14-rosner-repair.pdf

Rothenberg, J. (1999). *Avoiding technological quicksand: Finding a viable technical foundation fo rDigital preservation* (No. ED426715). Council on Library; Information Resources.

Russell, A. L. (2014). *Open standards and the digital age*. Cambridge University Press.

Russell, A. L., & Vinsel, L. (2018). After innovation, turn to maintenance. *Technology and Culture*, *59*(1).

SalahEldeen, H. M., & Nelson, M. L. (2013). Carbon dating the web: Estimating the age of web resources. In *Proceedings of the 22nd international conference on world wide web companion* (pp. 1075–1082). International World Wide Web Conferences Steering Committee. Retrieved from http://arxiv.org/abs/1304.5213

Salaheldeen, H. M., & Nelson, M. L. (2013). Resurrecting my revolution. In *Research and ad-*

vanced technology for digital libraries (pp. 333–345). Springer.

Samar, T., Huurdeman, H. C., Ben-David, A., Kamps, J., & Vries, A. de. (2014). Uncovering the unarchived web. In *Proceedings of the 37th international acm sigir conference on research & development in information retrieval* (pp. 1199–1202). ACM. Retrieved from http://humanities.uva.nl/~kamps/publications/2014/sama:unco14.pdf

Samuels, H. W. (1986). Who controls the past. *The American Archivist*, 109–124. Retrieved from http://americanarchivist.org/doi/abs/10.17723/aarc.49.2.t76m2130txw40746

Sanderson, R., Phillips, M., & Sompel, H. V. de. (2011). Analyzing the persistence of referenced web resources with Memento. Open Repositories 2011 Conference. Retrieved from http://arxiv.org/abs/1105.3459

Schafer, V., Musiani, F., & Borelli, M. (2016). Negotiating the web of the past: Web archiving, governance and sts. *French Journal for Media Research*, *6*.

Schellenberg, T. R. (1956). *Modern archives: Principles and techniques*. University of Chicago Press. Retrieved from http://catalog.hathitrust.org/Record/003147122

Schlanger, Z. (2017). Rogue scientists race to save climate data from trump. Retrieved from https://www.wired.com/2017/01/rogue-scientists-race-save-climate-data-trump/

Schneider, S. M., Foot, K., Kimpton, M., & Jones, G. (2003). Building thematic web collections: Challenges and experiences from the september 11 web archive and the election 2002 web archive. In *3rd ecdl workshop on web archives*. European Conference on Research; Advanced Technology for Digital Archives.

Schwartz, J. M. (1995). ”We make our tools and our tools make us”: Lessons from photographs for the practice, politics, and poetics of diplomatics. *Archivaria*, *40*(Fall), 40–74.

Schwartz, J. M., & Cook, T. (2002). Archives, records and power: The making of modern memory. *Archival Science*, *2*, 1–19.

Scott, J. C. (2017). *Against the grain: A deep history of the earliest states.* Yale University Press.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2).

Sheffield, R. T. (2018). Facebook live as record making technology. *Archivaria*, *85*.

Star, S. L. (1990). Power, technology and the phenomenology of conventions: On being allergic to onions. *The Sociological Review*, *38*(S1), 26–56.

Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, *43*(3), 377–391.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, *7*(1), 111–134.

Starosielski, N. (2015). *The undersea network.* Duke University Press.

Suchman, L. (1985). *Plans and situated actions: The problem of human-machine communication.* Xerox Corporation.

Taylor, H. (1988). My very act and deed: Some reflections on the role of textual records in the conduct of affairs. *American Archivist*, *51*(Fall), 456–469.

Taylor, H. A. (1984). Information ecology and the archives of the 1980s. *Archivaria*, *18*, 25–37.

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, *4*(2), 2053951717736335.

Thomer, A. K., Weber, N. M., & Twidale, M. B. (2018). Supporting the long□term curation and migration of natural history museum collections databases. In *Proceedings of the association for information science and technology* (Vol. 55, p. 1).

Tschan, R. (2002). A comparison of Jenkinson and Schellenberg on appraisal. *The American Archivist*, *65*(2), 176–195.

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.

Upward, F. (1996). Structuring the records continuum (series of two parts) part 1: Post custodial principles and properties. *Archives and Manuscripts*, *24*(2), 268.

Upward, F. (1997). Structuring the records continuum (series of two parts) part 2: Structuration theory and recordkeeping. *Archives and Manuscripts*, *25*(1), 10.

Walford, A. (2017). Raw data: Making relations matter. *Social Analysis*, *61*(2), 65–80.

Web archives 2015: Capture, curate, analyze. (2018). University of Michigan. Retrieved from https://web.archive.org/web/20180307070842/https://www.lib.umich.edu/webarchivesconference

Welch, I., Rehfeld, N., Cochrane, E., & Suchodoletz, D. von. (2012). A practical approach to system preservation workflows. *PIK - Praxis Der Informationsverarbeitung Und Kommunikation*, *35*(4), 269–280.

Winner, L. (1980). Do artifacts have politics? *Daedalus*, 121–136.

Yakel, E. (2003). Archival representation. *Archival Science*, *3*, 1–25.

Yakel, E. (2007). Digital curation. *OCLC Systems & Services: International Digital Library Perspectives*, *23*(4), 335–340.

Yang, S., Chitturi, K., Wilson, G., Magdy, M., & Fox, E. A. (2012). A study of automation from seed URL generation to focused web archive development: The CTRnet context. In *Proceedings of the 12th ACM/IEEE-CS joint conference on digital libraries* (pp. 341–342). Association for Computing Machinery.

Yeo, G. (2007). Concepts of record (1): Evidence, information, and persistent representations. *The American Archivist*, *70*(2), 315–343.

Yeo, G. (2008). Concepts of record (2): Prototypes and boundary objects. *The American Archivist*, *71*(1), 118–143.

Young, J. M. (1985). Annotated bibliography on appraisal. *The American Archivist*.

Zinn, H. (1977). Secrecy, archives, and the public interest. *The Midwestern Archivist*, *2*(2), 14–26.

Zittrain, J., Albert, K., & Lessig, L. (2014). Perma: Scoping and addressing the problem of link and reference rot in legal citations. *Legal Information Management*, *14*(02), 88–99.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89.

V. Appendixes

Study 1 Documents

Consent Form

You are invited to be in a research study that explores the selection of Web content for preservation in an archive. You were selected as a possible participant because you have some expertise in either the selection of Web content for archival processing or the design of tools to assist in the archiving of Web content. I hope to interview you on these subjects.

Interviewing will possibly occur with real-time computer and code referencing, this is the considered the "observation" element of this study. Please feel free to ask me any questions before participating. Being in the study is voluntary and you are free to stop at any time. Refusing to be in the study or stopping study activity will involve no penalty or loss of benefits to which you are otherwise entitled.

The purpose of this document is to give you the information you will need to help you decide whether to be in the study or not. Please read the form carefully. You may ask questions about the purpose of the research, what I would ask you to do, the possible risks and benefits, your rights as a volunteer, and anything else about the research or this form that is not clear. When I have answered all your questions, you can decide if you want to be in the study or not. This process is called "informed consent." I will give you a copy of this form for your records.

Background This study is being conducted by Ed Summers in the College of Information Studies at the University of Maryland. Its purpose is to explore the current processes and tools used to select content from the Web for archiving. If you agree to be in this study, I ask your permission to a conduct face to face semi-structured interview.

Your decision on whether to participate will not affect your current or future relations with the

University of Maryland or your current employer, and you may withdraw at any time without affecting those relationships. The face to face interview usually takes no more than 1 hour, and there is no compensation for participating in the face to face interview. The observation component of this study may be part of the interview and is included in this 1 hour time period. I envision this as an active interview that possibly involves looking at Web content together, talking about Web archiving work, and looking online for examples to discuss.

The audio/video recordings will be transcribed and then coded to identify patterns and strategies for selecting Web content for archiving.

Risks There is no physical or medical component to this research, and there is no risk of physical injury. The identities and organizational membership of the interviewees will not be revealed in our published findings, and a pseudonym will be used.

Benefits There may be no benefits to you personally for participating in the current research; however, there may be some professional and societal benefits. This study will help the archival community better understand the processes by which Web content is selected for an archive. Currently very little research into this phase of the work of Web archiving has been done. The hope is that the findings of this study will help influence the design of tools that assist archivists in their work.

Confidentiality Confidentiality will be maintained by a) aggregating information and b) by assigning pseudonyms. I have an ethical and legal obligation to protect confidential information used or obtained in the course of research and all policies on confidentiality apply equally to data stored both in the computer and on paper records. Any non-disclosure agreements to which you are a party will be respected and maintained by the security of aggregation and pseudonyms.

Confidentiality will be maintained by aggregating information in tables and graphs that describe broad trends in attitudes and demographics across the population. Composite descriptive sketches

will not refer to the data collected from named individuals.

Confidentiality will be maintained by assigning you a pseudonym. Any taped conversations, paper notes, or other research materials associated with our exchanges will be identified with the pseudonym. The only code sheet identifying you with your pseudonym will be kept in locked storage over a mile away from the research materials. This code sheet will be destroyed at the end of the study, or by the end of 2016.

If any published material is going to include a quotation from the transcript I will notify you via email and request your approval. If I don't hear anything back in two weeks I will resend the notification. If there is no response I will publish the quotation. I will respect your wishes to either not include the quotation, or to provide clarification.

Contacts You will be given a copy of this form to keep for your records, and you may ask any questions you have now. If you have questions later, you may contact me by telephone at +1 (240) 478-7086 or by email at edsu@umd.edu.

If you have questions about your rights as a research participant or wish to report a research-related injury, please contact:

University of Maryland College Park Institutional Review Board Office 1204 Marie Mount Hall College Park, Maryland, 20742 E-mail: irb@umd.edu Telephone: 301-405-0678

This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.

Statement of Consent Your participation in this study indicates that you are at least 18 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you voluntarily agree to participate in this research study.

Recruitment Letter

Dear {Name}

I am a researcher at the University of Maryland, and I am investigating the appraisal process in Web archives. I am interested in interviewing you regarding your expertise in this matter and attached to this email is the formal interview request. If you agree the interview will last no longer than 1 hour. Please feel free to contact me with any questions you might have.

You were selected as a possible participant because you have some expertise in either selecting Web content for an archive, or in building tools/services that facilitate the selection process.

I will be conducting the interviews via Skype and recording the audio and video. This is considered the "observation" element of this study. In the analysis phase of the study I will personally be creating transcripts, and performing open coding on them for theory building about the Web appraisal process. Your name and your institution will not be mentioned in any published material from these transcripts.

I hope you volunteer to take part in this study. Please contact me if you are interested in participating or would like to ask me any questions about it. If you do not contact me, I will follow up with you once and then assume that not hearing from you means you are not interested in participating. Please note that communication sent via e-mail cannot be guaranteed confidential.

Please find attached an information statement with additional information on the interview procedure.

Sincerely,

Ed Summers

Interview Protocol

First, I want to thank you for participating in this interview today. I know you are busy and I really appreciate you taking the time to help me. Before we get started I thought I'd start by telling you a little bit about the study and what you can expect from this interview.

Just to review, the title of the study is Investigating Appraisal in Web Archives. As you may know there are many initiatives to archive parts of the Web. These can be found in libraries, archives, museums, businesses and government. Often these organizations have articulated collection development policies to help guide what Web content is collected. However the actual process for discovering websites and content that is relevant for a Web archive isn't well known. The Web is a big place, and even at the organizational level it can be difficult to know what needs to be collected and when.

In this study I'd like to talk to you about this process of selection or appraisal. The hope is that a better understanding of the decisions and mechanics of how archivists select content will help inform the design of new tools to assist archivists in their work. I'm really interested in the nitty gritty practicalities about how websites are added to an archive. I do have a short list of questions but this is going is a semi-structured interview because I would like the conversation to evolve organically. I'm pretty sure I don't have all the right questions!

I'm anticipating that the interview will take anywhere from half an hour to an hour. Feel free to say you don't know the answer to a questions, or you'd rather not answer it, and you can end the interview at any time. Please ask me to clarify any questions that aren't clear. I am recording the audio and video from the interview, which I will be transcribing and then doing content analysis to (hopefully) find patterns and themes. Your interview will be kept confidential and destroyed at the completion of my study. I will not mention you or your institution by name in my study except using a pseudonym. If I plan on quoting you at all in published material I will modify the excerpt to reduce the likelihood of identification. I will also contact you and ask for your permission to use the segment.

Do you have any questions about what I just covered or anything else?

Great, well lets get started then.

1. Could you tell me a little bit about your role at XXX? (Alt: how long have you been there? What are your responsibilities?)
2. Can you describe your work environment? (Alt: How many people do you immediately work with, what are their roles?)
3. Have you ever selected Web content to be archived? (Alt: do you do this on a regular basis?)
4. Try to recall a time when you selected content for archiving. What was it? Can you remember how you found it?
5. Can you think of another example?
6. Can you think of any other examples that seem different from these two?
7. What criteria do you think go into deciding whether a particular Web page or website is worth collecting? Can you give me examples?
8. How does the presence of the content in other Web archives factor into your decision if at all?
9. Do you consider whether it is important to collect the same web resource over time?
10. Do you record any information about why a particular website was selected? Do you think that could be useful?
11. How does the perceived size of a website factor into your decision to archive content?
12. Do you search for content to archive? How do you do it?
13. Do you consult with your peers when locating Web content? How does that work?
14. Do you talk to peers at other institutions or organizations when identifying web content?

15. Do you ever interact with the content owner/provider when selecting content for archiving?
16. Do people ever try to donate Web content? Can you describe an example of that?
17. Do researchers ever request that you archive particular Web content? How did that happen?
18. Have there been any requests to access archived content yet by internal or external people? How did that go?
19. Could an inventory of your Web archives contents be made available for a follow up study?
20. Does your organization have a collection development policy when it comes to collecting Web content? What kind of guidance does it provide? (Alt: are there any agreed on criteria for what Web content to archive?)
21. Is that policy available to the public? Can I get a copy?

Study 2 Documents

Consent Form

Project Title Enacting Appraisal: Investigating the sociotechnical factors of software selection in the NSRL.

Purpose of the Study This research is being conducted by d Summers at the University of Maryland, College Park. I am inviting you to participate in this research project because you have worked to help build and/or sustain the National Software Reference Library. The purpose of this research project is to better understand how decisions are made about how to build digital collections, particularly when they involve obtaining content from the web.

Procedures The procedure involves an unstructured interview with you which will be audio recorded and transcribed. Interviews will take between 60 and 90 minutes. If you wish you can choose a pseudonym that will be used in place of your name in all written materials. Content analysis performed on the transcripts will be used to derive emergent themes and issues that speak to the study's research question. There are no direct benefits to participants.

Potential Risks There is no more than minimal risk associated with participating in this study. However, every potential subject will have the option to refrain from participation. Additionally, all participants will be asked if they would like me to use a pseudonym instead of their name in order to protect their identities and minimize risk.

Potential Benefits There are no direct benefits to you. However, the goal of this study is to understand how content selection processes and technical infrastructures have coevolved in the NSRL. This may or may not be of interest to your work in or with the NSRL.

Confidentiality Any potential loss of confidentiality will be minimized by storing data in encrypted form in a private DropBox folder which will only be accessible by researchers. If I write a report or article about this research project, your identity will be protected to the maximum extent possible. I will also notify you before hand if I plan to use any quotes from your interview to give you an opportunity to clarify, or if you would prefer me not to use the quote.

Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law. If participants wish their interviews can be donated back to the NSRL as historical documents.

Right to Withdraw Your participation and Questions in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.

If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the investigator:

Ed Summers

Maryland Institute for Technology in the Humanities
University of Maryland, College Park, 20742
edsu@umd.edu
301.405.8927

If you have questions about your rights as a research participant or wish to report a research related injury, please contact:

University of Maryland College Park
Institutional Review Board Office
1204 Marie Mount Hall

College Park,

Maryland, 20742

Email: irb@umd.edu

Telephone: 3014050678

For more information regarding participant rights, please visit: https://research.umd.edu/irbresearchparticipants

Recruitment Letter

Dear _____

My name is Ed Summers, and I am a researcher from the College for Information Studies at the University of Maryland. I am spending a year working with the National Software Reference Library as part of NIST's Professional Research Experience Program (PREP). During this time I'm investigating the social and technical factors that help shape the construction of digital archives, in particular the content that they collect.

Given that the NSRL is a unique and long running example of a digital archive I was hoping to speak with you about your experience working in or with the NSRL. Hearing more about the types of work you do, and how you think about the activities of the NSRL would be extremely helpful to me as I conduct this research project.

All interview materials (transcripts and recordings) will be kept confidential. If you would rather not be named as a participant you can choose a pseudonym that will be used in all research materials. In addition if any quotations from the interview that are used in published materials they will be sent to you beforehand for clarification and approval. I expect the interview to last no longer than one hour. I hope we can find a time to talk.

Sincerely,

Ed Summers

Interview Protocol

1. How did you first come to work with the National Software Reference Library?

2. Can you describe what your usual working day is like? For example, what kinds of activities do you get up to, and who do you interact with the most, and has this changed over time?

3. What computer systems, applications or tools do you use most often in your work?

4. How is material being selected for the NSRL? What do you think

5. How has the NSRL Reference Data Set and the collection of software been used?

6. Is there anything you were expecting me to ask which I didn't?