The Promise of the Finding Aid: A Critical Approach to Finding Biographical (Linked) Data in the Archive

Rachel Pierce[1], PhD (Research coordinator)

[1]*KvinnSam, Humanities Library, University of Gothenburg, Renströmsgatan 4, 41255 Gothenburg, Sweden*

Abstract

Projects focused on biographical archival data have traditionally placed community-wide institutional lists at the center – often church records for earlier historical periods and census information for years dominated by the modern state. Yet these lengthy lists of individuals lack contextual material, which is often necessary for definitively identifying individuals and building argument-driven historical research projects. This article will address the digital finding aid as a potential source of biographical data grounded in the contextual information so central to humanities research, using a set of finding aids of a selection of personal and organizational collections from the Schlesinger Library at Harvard University and the Library of Congress as examples. While resurrecting the evidentiary lives of everyone is impossible, embedding digital technologies in finding aids may offer a more equitable and less hierarchical method of assembling biographical data, as well as new doors into physical archives. However, a critical perspective is key; the creation of linked data will create new hierarchies and silences within the world of archives.

Keywords

digitization, linked data, finding aids, representational equity, archives

1. Introduction

As feminist historian Joan Wallach-Scott has noted, "historians make death a minor episode, something that is transitory rather than final" 1, 144. And yet it is difficult – impossible even – to resurrect everyone, though historians have tried. Digitization is merely the latest method. Recently, projects have focused on the digitization and mining of censuses, church records, and judicial documents, which have long been considered the site of much bottom-up historical research. While implementation of linked data has received a fair amount of attention in the drive to save past actors from ignominy, scholars have discussed linked data almost in the context of digitized and born-digital primary source collections, although biographical dictionaries have also received substantial attention [2][3]. Strangely, the choice of text for linked data implementation has rarely been problematized.

Given the size of archival collections and tendencies within selection for digitization, there are considerable problems inherent in assuming that the material that has been digitized is

https://doi.org/10.3986/9789610508120_01

representative of physical archival holdings [4]. And in reality, biographical dictionaries have many of the same representational problems, privileging a definition of individual importance grounded in nineteenth-century historiographical traditions that are out-of-tune with current historical work on grassroots publics, movement politics, and the everyday [5][6]. This article will argue that implementing linked data in finding aids, which have long provided routes into the archive for researchers, is potentially a more effective, efficient, and historically useful way of facilitating historical research. These documents are structured well for digital mining but also place individuals within correspondence networks, family structures, and work hierarchies while providing information on concrete context such as dates and places. Further, and most importantly, finding aids do not reflect history as it has been written, but rather open up linked data into the history that could be written. This article will address the structure of finding aids, which offer good contextualized biographical information while exploring the potential difficulties of differences between personal and organizational archives.

Once properly digitized, technical opportunities for analyzing the history embedded in these documents are numerous. But linked data in particular is a relatively unexploited and flexible tool that would make embedded biographical information computationally digestible without losing the contextual structure that is so central to humanities research. As noted by digital humanists, the field has struggled to create solutions that preserve context while producing machine-digestible humanities data [7]. But digital technologies are not simply well-suited to enhancing finding aids. They also offer an interesting set of solutions to problems of representation inherent within archival collections. With the addition of linked keywords and item-level description, potentially crowdsourced from researchers, this approach would also help to address historical imbalances in archival description. Scholars have pointed out the possibilities of linked data for depicting relationships between individuals, groups, and organizations without ordering these relationships hierarchically [8]. Linked data within finding aids is one potential practical solution that is worth exploration.

In order to analyze the possibilities integrating biographical linked data into finding aids, this article will look in depth at finding aids from both women and gender-centered archives and national archives. Materials from the Schlesinger, a library and archive devoted specifically to women's and gender collections, are finding aids for the collections of (1) the turn-of-the-century activist, economist, and author Charlotte Perkins Gilman and (2) the Boston Women's Health Book Collective, the group that authored the seminal feminist text Our Bodies, Ourselves. From the Library of Congress, finding aids for the papers of (3) former feminist congressperson Patsy T. Mink, of Hawaii, and (4) jazz drummer Max Roach. These materials span a century and a half and cover individuals and groups with diverse biographical and contextual backgrounds. The finding aids are also well-developed while displaying a structural diversity that will be useful in discussing the possibilities and limits of linked data within this genre.

2. Digitizing Finding Aids: Identifying Current Challenges

What is a finding aid? Chung offers a loose description of the document as "any type of descriptive tool that provides intellectual access of the materials held in an archival or manuscript repository" 9, 147. This definition is necessarily broad, given that the descriptive practices

employed by archivists varied greatly from institution to institution and type of collection to type of collection for centuries before standardization began to take hold in the second half of the twentieth century [10]. Now, many countries have guides for archival description that prioritize and seek to preserve provenance and original order through the both the hierarchical organization and description of archival collections. Finding aids also often include information about the creation and historical use of collections. These are documents designed to both describe collections so that their contents are understood within a historical system of practices and provide a multiplicity of entry points into archival collections for the roving researcher. These two goals are not always synergetic, as multiple scholars have noted [11][12].

Archives began digitizing finding aids early in the history of the Internet, in order to make their collections more findable and provide information to researchers so that they could order sections of archival collections in advance of physical visits. This was a revolutionary shift in archival research practices that received little attention at the time [12]. The shift to digital (digitized) finding aids capped a broader shift towards standardized descriptive practices, making these now systematically structured documents open to the public. The implementation of encoded archival description (EAD) has dominated the transformation of finding aids to digital, computer-searchable documents. This description system is an international standard for encoding finding aids using XML, run by the Society of American Archivists and the Library of Congress. More recently, the RDF-based Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF) was developed to encode archival entities and enable the implementation of linked data.

Since its introduction at the end of the 1990s, the implementation of EAD has proceeded without much debate. Yet concerns about the usability of the finding aid have only increased, as digitization has shifted the definition of archival users from a physical group of presumed historians with a sprinkle of genealogists to a broader public. Simultaneously, web access to finding aids was not accompanied by digital access to archivists themselves – traditional repositories of information about finding aid navigation in general and individual collection navigation in particular. As Cox has argued, "Perhaps the real business of postmodern archival enterprise ought to be re-evaluating just what finding aids represent over time, studying them as a documentary source reflecting attitudes and practices of the archival community at various times" 10, 9. What is clear is that while digitization of finding aids made whole archives more discoverable, it did not necessarily make the search for particular topics or individuals within the Archive – across collection and institutional boundaries – more findable or more usable [1]. As Garmendia and Retter note, this is a far more complicated process than it might seem at first [14].

This situation has particular repercussions for scholars seeking materials about and by historically underrepresented groups. Feminist scholars have long argued that the structure of both archives and finding aids obscure or erase the presence and importance of women, people of color, the lower and working classes, and other historically disempowered peoples [15][16]. In the context of digital finding aids, these scholars demand two things. The first demand is the

---

[1]Here, I follow Carolyn Steedman's practice of differentiating between individual collections, archival institutions, archivist practices, and the Archive of the world's current and future cultural historical matter [13]. The last of these is as much an idea as a(n ever-changing) physical reality.

preservation of historical hierarchies that not only lies at the heart of archival order but also provides meaningful contextual information about the hierarchies that governed the creation of these collections. The second demand is the creation of tools that allow archival researchers to circumvent these hierarchies in searching for historically and archivally marginal groups and individuals. It is becoming clear that current trends within the digitization of finding aids fulfill the first goal while failing on the second.

So why finding aids, given these barriers? Until now, most digital humanities scholars have focused on implementing linked data for digitized cultural heritage materials or national biographies. Issues with digitized cultural heritage materials are clear – these collections of documents represent a scattershot array of archival materials, whose representativeness is both highly doubtful and impossible to estimate. Biographical dictionaries seem, upon first glance, to go beyond the specificity of archival documents in some of the same ways as finding aids. However, as Warren (2018), points out, the biographical dictionary exists somewhere in between the categories of primary and secondary source material, as a time capsule of the history that has been written. The relatively young sub-fields of women's, gender, indigenous, Black, queer, and social movement history (among many others) take up very little space in these texts, even as materials from these histories have flooded into preexisting archival institutions and prompted the widespread founding of new, activist archives that prioritize [17]. Meanwhile, the genre of biography continues to suffer from a tendency to prioritize the historically powerful [18]. If archives move slowly, biographical research moves even slower. And some scholars argue that the genre is structured to prioritize in this way, that even future historiography will not undermine the individualism and elitism of biography [19].

The digitization of finding aids – including for these new, introduces new possibilities for integrating digital tools into these documents. Yet development here has been slow and ignores the criticisms of archival scholars whose work focuses on the importance of understanding and making transparent the power dynamics that the archive reflects and institutionalizes. As previously noted, digital humanists have investigated the possibilities of linked data and archives but have focused largely on digitized and born digital primary source material, that is, material from the archive rather than material about the archive. This emphasis has resulted in archivist work on "knitting" relationships between finding aids and digitized documents to "improve use and discoverability" [20], though some projects such as LOCAH translated biographical EAD description to linked data (Niu 2016). Similarly, the arguments and projects presented at the 2021 workshop on linked data in archives focused on using linked data to reproduce archival information in linked data form, with little attention to archival researcher practices and inbuilt archival hierarchies (Lopes et al. 2021). As a result, this work does not expose, question, or disrupt the hierarchical structure of the archive.

The application of linked data within finding aids has not been discussed much, and as this article, will show, linked data is far from a panacea for the inequities of the archive or the internet. Critical approaches within the digital humanities raise concerns that mere application of digital tools does not erase or rectify pre-existing hierarchies and can, in fact, deepen these problems (Risam 2019). Humanists must, as multiple scholars argue, be attuned to the infrastructures that linked data sets, visualizations, and interfaces build on – these infrastructures are also objects in need of (humanistic) analysis [21][22]. Yet digital humanists interested in queering the field have focused on metadata as a potential site of creativity for a reason [8]. There are

opportunities here, if we are willing to exploit them in an imaginative and critical way. This article will present the possibilities, built-in biases, and potential pitfalls of linked data for finding aids, as a starting point for scholarly discussion and potential future implementation.

3. Case Studies from the Schlesinger and the Library of Congress

Four finding aids from two different institutions have been chosen for analysis. The Arthur and Elizabeth Schlesinger Library for Women's History is located at Harvard University, and two collections have been chosen for investigation, in order to look at the suitability of individual/familial papers versus organizational papers. The first case study is the finding aid for the papers of 19th century women's rights campaigner, economist, and author Charlotte Perkins Gilman [23]. The second finding aid covers the collections for the Boston Women's Health Book Collective, the 1970s feminist activist group that wrote the seminal reproductive rights text Our Bodies, Ourselves [24]. The last two finding aids are from the Library of Congress. The first is the finding aid of the papers of former Hawaii Rep. Patsy Mink, a legislator well-known for her work on women's and Asian American rights [25], and the second is the under development finding aid for bebop drummer Max Roach's papers [26].

These are well-developed finding aids that are far from representative. It takes extraordinary institutional resources to create this kind of document, underlining the resources that would be needed to create linked data while simultaneously highlighting the resources that would be saved if linked data was created as a part of finding aid creation. The need for crowdsourcing is acute here, and indeed crowdsourcing for linked data can provide not only a symbiotic relationship between archival scholars and archivists but a concrete introduction to digital tools for these scholars, who have long been seen as both ignorant of and resistant to digital tools [27].

The particular finding aids have also been chosen, somewhat ironically, to center a cross-section of underrepresented groups. But these finding aids should not be taken as representative. The scope of archival collections and the biographical histories that can be written based in these archives are neither representative of particular groups nor of the Archive in its voluminous, fractured, and fluid entirety. Instead, this article will use the four chosen case studies to illustrate more general possibilities and problems with using finding aids to create biographical linked data, with particular attention to issues of representation, classification, hierarchy, and binaries highlighted by feminist archival scholars.

3.1. Case Study 1: Papers of Charlotte Perkins Gilman, 1846-1961

A finding aid is a descriptive text and a map. These documents provide historical context for collections – who owned the materials over time, how and when they arrive at their final repositories – and the background information for the person(s) or organizations around whom/which the collections have been constructed. These introductory texts are minable for information that can become linked data. The Perkins Gilman finding aid from the Schlesinger Library at Harvard University is extensive and detailed, illustrating the potential for linked data implementation in well-developed finding aids.

This finding aid demonstrates that such documents – specifically those for individual and familial collections – are particularly well-suited for finding good, contextualized biographical linked data. The structure of the finding aid eases the process of finding this information. There are several places to look for such information, but in particular, the biographical note, the index, and the subject list all provide information on relationships between Perkins Gilman and a wide array of familial, work, and friendship relationships. The biographical note provides an abundance of potential well-contextualized linked data, including influences (such as Lester F. Ward and Sigmund Freud), friends (Grace Ellery Channing, for instance), and genealogical names (including her husband, artist Charles Walter Stetson) that can be linked to dates, places, and specific locations in the physical archival collection. The index and subject lists provide a total of 229 unique names that link these relationships to physical artifacts in the archive that provide evidentiary bases for these relationships. The finding aid structure also provides information on correspondence relationships, namely in Series II and III, where box-level description reveals familial, friend, and work relationships with substantial physical documentation. The Schlesinger has digitized much of this letter-writing, meaning that these relationships can also be linked to digitized material.

Perkins Gilman's biographical trajectory can also be traced via her position within a number of national and international women's rights and political organizations. These organizational names, which also appear in the index and subject list, provide important contextual information, placing her within a set of social movement networks that were the basis for mainstream women's rights politics of the 19th and early 20th centuries. Individuals are the basis of such organizations, as will be shown with the next section, and the implementation of linked data for charting organizational leadership and membership over time and space would provide important information on how social movements evolved and interacted with various people and institutions at specific points in time and in specific places.

Discussions of linked data in archives have until now not dealt with issues of hierarchy or representation, much less what is left out as these technologies are utilized. The assumption has been that linked data replaces hierarchies with graph-based relational representations that undermine archival hierarchical organization. But what this particular finding aid most reveals are the repercussions of selection in shaping the kinds of representations that are enabled with linked data implementation. This finding aid offers a fairly equal number of women and men, while class and race (amongst other kinds of) diversity is absent. The collection is populated with middle-class White figures. Those applying linked data in archival contexts should consider these qualities of each archival collection as well as what silences and new hierarchies are created as linked data is fashioned.

3.2. Case Study 2: Records of the Boston Women's Health Book Collective, 1905-2003 (inclusive), 1972-1997 (bulk)

Looking at the finding aid for the Boston Women's Health Book Collective provides a glimpse of how the biographical make-up of social movements might be expressed, as well as illustrating the difficulties such a finding aid poses for linked data implementation. This finding aid has a subject list but no name index, and the subject list contains only two names. However, both the "Scope and Contents" note and the sub-series-level description provides a wealth of names

(11), particularly relating to the founders of the collective, both of whom have contributed a full series of materials (Series V for Judy Norsigian and Series VI for Esther Rome) board work, and the outward-facing members who engaged in communication with the general public (see Series XXIV). The "History" section of the finding aid includes a list of founding members (14 women in total) noting that most of these women "remained together for more than twenty years, sharing their personal and professional lives" [24].

Understanding that individuals and collectives are interactive and interdependent historical actors is an important consideration. A major question with biographical linked data is how the concept "biography" is defined. Recent historical biographical research has pushed back against the traditions of biography, resulting in very interesting work on couples, families, and groups that challenges the individualism that has long undergirded the genre [18]. The finding aid and collections for the Boston Women's Health Book Collective challenge us to think about what we lose when we prioritize individual-oriented biographical data to the exclusion of group materials, particularly those related to social movements. It is here that less well-known individuals contributed to broad societal change through coordinated collective action. Being able to map the organizational relationships that sustained the activist-political careers of individual women over time would be an immense contribution to historical scholarship.

Actor-network theory treats people, events, objects, and sources as equal, providing a way of constructing biographical linked data that represents organizations and sub-units within organizations (for example, office staff or boards) alongside individuals [28]. This theory parallels SPARQL endpoints, which facilitate search for a multiplicity of objects, places, concepts, and other entities 29, 95. Such an approach to linked data would be highly appropriate, especially in the case of organizational archival collections. In the case of the Boston Women's Health Book Collective, relationships between different sectors of the collective (the board, marketing staff, grant writers, etc.) could be articulated. Further, relationships between the Collective and other organizations and projects were so central that they warranted their own series (Series IV). The founders of the Collective also worked in and with numerous other local and national organizations, functioning as links between various kinds of women's health work in and outside of government and across political, healthcare, and social movement environments. Representing these organizational synergies in linked data would be central to a good representation of the multifaceted networks that sustained the work of both the Collective and its individual members.

The finding aid for the Boston Women's Health Book Collective also illustrates the need to exploit physical archival labor in order to develop and implement new digital tools within the finding aid. Historians increasingly take photographs, create spreadsheets, and do other digital work as part of their archival research [27]. These digital skills are developed to contend with the administration of large quantities of information and documents, and their development offers an opening for synergy between physical archival research and the digitization of finding aids. But other options exist, particularly for archives located at universities, where students, professors, archivists, and digital resources staff can collaborate on the construction of linked data. Experiments of this kind will help to equip the next generation of archivists and researchers with useful digital skills [30].

3.3. Case Study 3: Patsy T. Mink Papers

Patsy Mink's collection is a lesson in humility for archival linked data enthusiasts. Mink represented Hawaii in the United States House of Representatives for 26 years, as well as cultivating a career in local and state politics. Her collection demonstrates the problems with large archives. Resources and archival traditions mean that finding aids describing large collections can be somewhat more concise and often fail to describe materials at the item level. That 24 individuals worked on constructing the finding aid alone is a sign of the kind of work that implementation of linked data would require [25]. These collections also are more likely to have vague titling practices that, for instance, denote policy areas, correspondence for particular years or specific document formats like photographic collections. This is a set of problems that gets more dire the closer one gets to the state, an irony that illustrates the flipped dynamics at work with linked data creation. Here, the congressperson is an actor within a state apparatus, literally a representative of collective interests. This situation produces an elision of actors in order to highlight the representative's interaction with and within a governmental organ.

But while working on this kind of linked data is far from easy, the result would provide important information on the nature of lawmaking, which is built on networking and personal relationships. Mink commanded a formidable staff of secretaries, lawyers, writers, and researchers on her personal and committee staffs. These individuals had relationships with staffs in other offices. These relationships are the groundwork upon which lawmaking is built, and yet these names are not elaborated in the finding aids for the collections of congresspersons, who were supposed to own the work of their offices. This is information that would need to be crowdsourced, and it is voluminous. There are 535 members of Congress every session and thousands of staffers who have passed through the offices of Capitol Hill.

Equally technically challenging but ethically fraught as well are two components of the Mink finding aid: constituent materials and genealogical information. If we consider provenance from a community archives perspective, these individuals should be considered co-creators of the archives that have heretofore been classified as congressional. And yet this is an authorship that they did not foresee or intend. We cannot assume that constituents wrote to their congressional representatives thinking that they were, in doing so, contributing to the historical record, although some letter-writers may have thought this way. Nor can we assume that these individuals would want this information published for the broader public, given that the historical record of archival collections has until recently been understood to be relatively closed and inaccessible. The genealogical information in Mink's collection presents further issues with explicitly restricted material, a not-uncommon occurrence, especially with more recent collections.

Linked data is, as Terhi Nurmikko-Fuller notes, an "information publication paradigm," and it is important to understand that this technology requires the publication of the personal data of individuals who considered their communications somewhere between public and private [31]. While scholars often position linked open data as the best-case scenario, it is important to remember that openness is not always the choice of those whose information is in question, and publication in any form may be against the wishes of archival contributors. Recent scholarship on community archives provides an important corrective to traditional archivist practices that simplified and individualized ownership of archival collections, and this scholarship has

repercussions for the implementation of linked data [17].

It is important to take these concerns into account, especially because it is individuals from historically disempowered groups who tend to be wary of depositing their papers with archival institutions and often place substantial restrictions on their papers when they are donated [32]. The implementation of linked data should always take place in line with the CARE data principles: Collective benefit, Responsibility, Authority to control, and Ethics. Though these principles were developed to protect indigenous community rights amidst digitization drives, the ubiquity of rights and access restrictions in the context of archives means that this issue must be met head-on regardless of where one chooses to start with linked data creation. The resolving of publishing rights for linked data will not be an easy task, as the FAIR principles of making data Findable, Accessible, Interoperable, and Reusable can end up conflicting with CARE [33]. Much work with data rights education will need to be done to ensure that scholars create linked data that does not infringe on the rights of individuals or leave out large swaths of the historical record.

3.4. Case Study 4: Max Roach Papers and Linked Jazz

The Max Roach Papers are currently being processed by the LoC, providing a look at how a less well-developed finding aid might still provide good quality linked data that enhances current linked data projects based in primary documents [26]. Roach was a renowned jazz drummer whose work spanned much of the second half of the 20th century. The Linked Jazz project began by mining oral histories for information about relationships between jazz musicians who, as Roach's own career amply demonstrates, were highly collaborative, taking inspiration from multiple sources and playing and recording in numerous configurations - for instance, Martin Luther King, Jr. is now included. Roach does in fact appear in the project, though he is currently somewhat peripheral within the network.[2]

What we see when we look at the Roach finding aid is new influences and collaborators that do not conform to the themed bounds of projects like Linked Jazz. For instance, there is currently no Linked Jazz triple elaborating the relationship between Roach and Nina Simone, a relationship important enough to show up in the shorter contextual summary of this in-process biographical section. But we also see a much wider influence and collaboration net that reflects time- and place-specific Black Power cultural creativity. Amiri Baraka (who wrote about jazz but did not play) and Maya Angelou (best known as an author, but also a singer who released one studio album in 1957) feature prominently, largely because of their similar sociocultural-political positioning. Baraka is absent from Linked Jazz, while Angelou is linked, tellingly, only to other women, including Roach's wife and collaborator Abbey Lincoln.[3]

Thematic projects – which include biographical dictionaries – tend to create boundaries and engage in selection and classification processes that obscure other facets of or ways of understanding the individuals featured in these projects. What finding aids give us is a set of relationships reflected in the stuff and structure of archival collections, which bridge divides between the personal and the political, the private and the public in ways that themed projects often (accidentally or not) reify. As Carolyn Steedman asserts, "The Archive is made from

---

[2]https://linkedjazz.org/network/ (accessed 15 Aug 2022)
[3]https://linkedjazz.org/api/relationships/all/nt (accessed 15 Aug 2022)

selected and consciously chosen documentation from the past and also from the mad fragmen-
tations that no one intended to preserve and that just ended up there" 13, 65. This mishmash of
intended and unintended, consciously kept and simply forgotten is not simply a goal of feminist
research. It is also a key driver of historical research 1, 145.

At the same time, the Linked Jazz project illustrates how biographical relationships can be
translated into linked triples via a combination of archival work, automation, and crowdsourcing.
Linked Jazz began with the use of a transcript analyzer to identify personal names, which were
then related to one another using the simple rel:knowsOf relationship, which crowdsourcing
then refined [34]. Finding aids have, as noted, very regular structures, making them ideal
documents for automated analysis. This approach allows for the identification of names and
relationships that can (or must, as the case may be) later be nuanced as researchers or archivists
use the collection. Given that finding aids are the door to archives, it is somewhat strange that
co-creation of finding aids has not been proposed as a key to making the archive more accessible.
Current archival science scholarship centers the archivist as non-neutral author [35], but future
finding aid construction will need to account for multiple authors, including machines.

The project also demonstrates how linked data can allow users to center historically under-
represented groups – if project creators have thought about and planned for this representation.
The addition of gendered attributes came late in project development and was conceived because
of the underrepresentation of women in the initial dataset. Attention to the smaller datasets
and the less well-represented or seemingly invisible individuals and groups is important, not
simply for producing linked data that is flexible and useful for lots of researchers but because
such disparities and silences are historically meaningful and often can provide us with new
perspectives rather than reproducing inequities that scholars already know exist. In particular,
as the Linked Jazz team notes, available linked data tools are not particularly well-representative
of the fluidity of sex and gender – an important area for future innovation 36, 922.

What an analysis of the Roach finding aid adds is a sense of how different the linked data
possibilities are when one starts with an archival collection rather than an oral history transcript.
Finding aids have the potential benefit of stretching across entire lives in a manner different
from oral histories. They often center familial and work relationships quite differently, since
finding aids are descriptive rather than narrative devices. Their subject indexes also – as Roach's
does – place individuals, groups (such as bands and recording constellations), and themes (like
"percussion music" or "race relations in mass media"). The Roach finding aid also highlights
the potential for cutting across documentary format, as materials belonging to these categories
are arrayed across photographs, musical scores, interview transcripts, and other document
types, which might help to rectify the historical privileging of text documents as the basis for
historical analysis [37]. These categories overlap with contributor names and the entities that
could be pulled from what is currently titled the "Summary" but will eventually end up in the
"Scope and Content" note. It is this messy array that makes the collection "a dynamic portrayal
of the personalities and cultural forces behind jazz in the United States and abroad in the latter
half of the twentieth century" [26] It is also this mess that potentially makes linked data such a
useful tool.

4. 8. The Challenges and Potential Gains of Finding Aid Linked Biographical Data

Christine Borgman has cogently argued that "data have no value or meaning in isolation. . . They exist within a knowledge infrastructure—an ecology of people, practices, technologies, institutions, material objects, and relationships" 38, 4 Any attempt to create biographical linked data should not forget that context is everything, and biographical data culled from archival collections has two important contexts: specific archives AND the historical context of the material itself. One of the potential promises of extraction of biographical linked data from finding aids is the preservation of both of these contexts. This approach to finding aids would provide a multitude of remixable routes into the physical and digital archives – an unfulfilled goal of digitizing archivists since the invention of the internet [12]. Crucially, linked data has the potential to break down the barriers between individual institutions that prevent researchers from seeing biographical networks that span archival collections and have the possibility to challenge rather than reify current and past historiography.

Perhaps most importantly, archival finding aids open up the array of primary materials in the Archive, facilitating future research in ways that bounded projects and biographical dictionaries cannot. There are vast oceans of material in archival collections that have not yet been researched. The limitations of biographical dictionaries for future historical research are clear; as historiographical, national documents, these dictionaries provide a window into what has been done and are not easily analyzed for what is not there but could be there. That much of history is simply inaccessible to us is a fact we must simply accept as the product of "archival silences," as Michael Moss and David Thomas put it [39]. Current linked data implementations are simply not designed to increase the routes into the physical archives, across institutional and national boundaries, as well as across well-trafficked and under-used formats, persons, and groups. If the goal is to spur new research questions and topics within the humanities and particularly within the discipline of history, finding aids are a highly appropriate starting place.

The technical or administrative aspects of this kind of work are far from easy. The creation of finding aids is undergoing revision as these documents and archival material itself moves online. Teaching archivists simple markup languages such as Text Encoding Initiative (TEI) would provide the groundwork for implementation of initial linked data structures within finding aids themselves. This kind of work would necessarily span institutional boundaries and reject narrower project-based focuses. Here, archivists would need to think locally, in line with postcolonial digital humanities practices that foreground "the particular over the universal," which via crowdsourcing and material selection can "offer [historically underrepresented groups] a more expansive humanities that takes advantage of the technical means of digital knowledge production to populate the digital cultural record with their own stories" 40, 9.

Defining what does and does not count as biographical data is key. Scholars have debated the relative merits of traditional biography, noting that the genre has certain conservative, "bourgeois" tendencies that deserve attention if not redress [19]. Given the Archive's traditional status as a repository for public-minded materials and the raced, classed, and sexed history of our definitions of "the public," it is difficult to reconstruct the biographies of individuals and groups who fell outside of this normative definition of the historically important [18]. An

analysis of the potential for linked data in finding aids demonstrates that there is material there that disrupts biography's tendency to use individuals with more substantial archives to represent groups. The structure of linked data embeds these individuals in sociopolitical networks, undermining the notion that great individuals – the ones with the archives – make history, while everyone else is at the mercy of historical forces. Treating organizations, spaces of interaction and invention, and other nonhuman entities as biographical would also do much to undermine this individualization of history, in no small part by providing representation to social movements.

However, the implementation of linked data cannot be seen as a panacea for the problems of representation and equitable engagement with archival materials. As the directors of the transnational digital African archive Aluka notes, "As it relates specifically to the domain of history as a discipline, the digitization process places contested archives in a cyberspace that is highly commodified" 41, 73. This commodification has numerous repercussions, primary amongst them a tendency to use digital tools to "sell" archival institutions, an imperative that results in the prioritization of flagship materials rather than the unexplored corners of collections. Good, transparent selection practices – which finding aids are chosen for processing and which sections are translated into linked (open) data – must be conducted together with affected groups and individuals. Processual openness is critical, if finding aid-based biographical linked data is to adhere to the CARE principles and serve as a means of making visible individuals and groups who relegated to the archival margins via traditional archival practices. But the fact of linked data as a salable tool inside and outside the academy will affect who and why this kind of work is considered valuable and useful.

Nor should we forget that archives have always served an exclusionary function. Death is only a "minor function" for the few. One of the most important issues for current digital humanities researchers is thinking about and finding ways to represent individuals, groups, and topics who/that are either underdocumented or evidence themselves as silences in the Archive. The goal is to research these hierarchies, but linked data has the potential to make such disjuncture and inequity less visible, as datasets tend to seem complete and neutral 42, 99. Linked data should be used to subvert but also make visible these exclusionary tendencies. New digital tools should be used to help us see some of the representational hierarchies built into the physical archives, while also allowing for a way of remixing archives to both center previously peripheral individuals and groups and make the silences and edges of archival collections more visible. Archival scholars and archivists themselves will then need to think about how to understand and interpret these silences and edges. Here, the humanities' emphasis on the need to account for doubt, multiplicity, and subjectivity is of critical value.

5. Conclusion

Linked data has attracted a great deal of attention from scholars and archivists who are interested in opening up archival collections to the general public, making materials more findable for researchers, and challenging institutional silos. There are other interesting gains identified here, amongst them a questioning of distinctions between individuals and collectivities, as well as a contextualization of biographies in the materials that undergird our understanding of historical

actors, as well as the provenance of these materials. Selection is key, as finding aids provide biographical networks from a particular person or organization's perspective. Last, linked data offers archivists an opportunity to test some archival science theories about centers, peripheries, and local knowledge (see for example [43], since it is potentially more interesting and fruitful to begin with smaller archival collections from individuals and collectives that have until now been viewed as less important. The implementation of linked data multiplies the doors into physical and digital archival collections, potentially changing the process of archival research in profound ways.

However, as with many technological "solutions," scholarship has focused on problems of implementation rather than the opportunities and limitations of these tools, as well as what such tools will mean from a user perspective. Digitization of various aspects of the Archive does not free information or materials from the hierarchies of the archive, but instead rearranges and subverts some of the power dynamics of archival collections while bolstering other inequities. While using finding aids as a jumping off point for biographical linked data creation is a potentially more equitable solution than starting with the "flagship" materials that have been prioritized for digitization, this method is not an end in itself. But to users, digital representations of the Archive can seem encompassing, even if digitized collections are incomplete or contain OCR-related problems, precisely because digitization is a process of decontextualization [44]. A feminist approach to issues of ethics and representation can help archivists to structure the deployment of linked data technologies in ways that undermine archival hierarchies while also illustrating the power dynamics and silences that are always at play in cultural heritage, digitized or not.

Acknowledgments

References

[1] J. Wallach-Scott, The Fantasy of Feminist History, Duke University Press, Durham, NC, 2011.

[2] A. Hawkins, Archives, linked data and the digital humanities: Increasing access to digitised and born-digital archives via the semantic web, Archival Science 22 (2022) 319–344. doi:10.1007/s10502-021-09381-0.

[3] M. Tamper, P. Leskinen, E. Hyvönen, R. Valjus, K. Keravuori, Analyzing biography collections historiographically as linked data: Case national biography of finland, Semantic Web 14 (2022). doi:10.3233/SW-222887.

[4] R. Pierce, Documents in the dynarchive: Questioning the total revolution of the digital archive, Proceedings from the Document Academy 8 (2021). doi:10.35492/docam/8/2/4.

[5] A. Fokkens, S. T. Braeke, N. Ockeloen, P. Vossen, S. Legêne, G. Schreiber, V. de Boer, Europa

baut auf Biographien: Aspekte, Bausteine, Normen und Standards für eine europäische Biographik, new academic press, 2017, pp. 193–227. doi:10.1007/3-540-09237-4.

[6] C. N. Warren, Historiography's two voices: Data infrastructure and history at scale in the oxford dictionary of national biography, Journal of Cultural Analytics 1 (2018) 1–31. doi:10.22148/16.028.

[7] D. Oldman, Information and Knowledge Organisation in Digital Humanities: Global Perspectives, Routledge, New York, NY, 2016, pp. 131–153.

[8] B. Ruberg, J. Boyd, J. Howe, Bodies of Information: Intersectional Feminism and Digital Humanities, University of Minnesota Press, Minneapolis, MN, 2018, pp. 108–127.

[9] S. K. Chung, Encyclopedia of library and information science, 3 ed., Taylor and Francis, New York, NY, 2010, p. 147–153.

[10] R. J. Cox, Revisiting the archival finding aid, Journal of Archival Organization 5 (2008) 5–23. doi:10.1080/15332740802153245.

[11] L. Freund, E. G. Toms, Interacting with archival finding aids, Journal of the Association for Information Science and Technology 67 (2015) 994–1008. doi:10.1002/asi.23436.

[12] C. B. Trace, A. Dillon, The evolution of the finding aid in the united states: From physical to digital document genre, Archival Science 12 (2012) 501–519. doi:10.1007/s10502-012-9190-5.

[13] C. Steedman, Dust: The Archive and Cultural History, Rutgers University Press, New Brunswick, NJ, 2002.

[14] J. Garmendia, A. Retter, Developing a pan-archival linked data catalogue, in: F. N. I. R. Carla Teixeira Lopes, Cristina Ribeiro, N. Freire (Eds.), Proceedings of Linked Archives International Workshop 2021, 25th International Conference on Theory and Practice of Digital Libraries, 2021, pp. 93–103.

[15] A. Burton, Dwelling in the Archive: Women Writing House, Home, and History in Late Colonial India, Oxford University Press, Oxford, 2003.

[16] A. Stoler, Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense, Princeton University Press, Princeton, NJ, 2010.

[17] J. Bastian, A. Flinn (Eds.), Community Archives, Community Spaces: Heritage, Memory and Identity, Facet Publishing, London, 2020.

[18] N. Chaudhuri, S. Katz, M. E. Perry (Eds.), Contesting Archives: Finding Women in the Sources, University of Illinois Press, Urbana, IL, 2010.

[19] S. Kracauer, Biography in Theory: Key Texts with Commentaries, de Gruyter, Berlin, 2017, pp. 107–111.

[20] L. Woolcott, A. Payant, S. Skindelien, Partnering for discoverability: Knitting archival finding aids to digitized material using a low tech digital content linking process, code4Lib Journal 34 (2016).

[21] J. Drucker, Graphesis: Visual Forms of Knowledge Production, Harvard University Press, Cambridge, MA, 2014.

[22] B. Mak, Archeology of a digitization, Journal of the Association for Information Science and Technology 65 (2014) 1515–1526.

[23] E. Mosley, J. Gotwals, Charlotte perkins gilman papers, 1846-1961, 1972, updated 2010. URL: https://hollisarchives.lib.harvard.edu/repositories/8/resources/4827, finding aid.

[24] K. Kraft, J. Carll, Boston women's health book collective records, 1905-2003, 2005. URL:

https://hollisarchives.lib.harvard.edu/repositories/8/resources/6841, finding aid.

[25] Patsy t. mink papers, 1883-2005, 2007. URL: https://hdl.loc.gov/loc.mss/eadmss.ms010008, finding aid.

[26] Max roach papers, 1880-2012, 2016. URL: https://hdl.loc.gov/loc.music/eadmus.mu016007, finding aid.

[27] I. Milligan, History in the Age of Abundance?: How the Web Is Transforming Historical Research, McGill-Queen's University Press, Montreal and Kingston, 2019.

[28] G. Durepos, A. J. Mills, Actor-network theory, anti-history and critical organizational historiography, Organization 19 (2011). doi:10.1177/1350508411420196.

[29] J. Niu, Linked data for archives, Archivaria 82 (2016) 83–110.

[30] J. Needham, J. C. Haas, Collaboration adventures with primary sources: Exploring creative and digital outputs, The Journal of Interactive Technology and Pedagogy 14 (2019).

[31] T. Nurmikko-Fuller, Teaching linked open data using bibliographic metadata, Journal of Open Humanities Data 8 (202). doi:10.5334/johd.60.

[32] D. G. White, Mining the forgotten: Manuscript sources for black women's history, The Journal of American History 74 (1987) 237–242.

[33] S. R. Carroll, I. Garba, O. L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, K. Raseroka, D. Rodriguez-Lonebear, R. Rowe, R. Sara, J. D. Walker, J. Anderson, M. Hudson, The care principles for indigenous data governance, Data science journal 19 (2020) 1–12. doi:10.5334/DSJ-2020-043.

[34] M. C. Pattuelli, M. Miller, L. Lange, S. Fitzell, C. Li-Madeo, Crafting linked open data for cultural heritage: Mapping and curation tools for the linked jazz project, Code4Lib Journal 21 (2013).

[35] M. Caswell, Dusting for fingerprints: Introducing feminist standpoint appraisal', in radical empathy in archival practice, Journal of Critical Library and Information Studies 3 (2021). doi:10.24242/jclis.v3i2.113.

[36] M. C. Pattuelli, K. Hwang, M. Miller, Accidental discovery, intentional inquiry: Levaraging linked data to uncover the women of jazz, Digital Scholarship in the Humanities 32 (2017) 918–924. doi:10.1093/llc/fqw047.

[37] J. Bryant, Ephemera, archives, and another view of history, Archivaria 40 (1995) 189–198.

[38] C. Borgman, Big Data, Little Data, No Data: Scholarship in the Networked World, MIT Press, Cambridge, MA, 2017.

[39] M. Moss, D. Thomas, Archival Silences: Missing, Lost, and Uncreated Archives, Routledge, Abingdon, 2021.

[40] R. Risam, New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy, Northwestern University Press, Evanston, IL, 2019.

[41] A. Isaacson, P. Lalu, T. Nygren, Digitization, history, and the making of a postcolonial archive of southern african liberation struggles: The aluka project, Africa Today 52 (2005) 55–77.

[42] E. Losh, J. Wernimont, L. Wexler, H.-A. Wu, Debates in the Digital Humanities 2016, University of Minnesota Press, Minneapolis, MN, 2016, pp. 107–111.

[43] D. Agostinho, Archival encounters: Rethinking access and case in digital colonial archives, Archival Science 19 (2019) 141–165. doi:10.1007/s10502-019-09312-0.

[44] N. Maurantonio, Archiving the visual, Media History 20 (2014) 88–102. doi:10.1080/