Context Lost: Digital Surrogates, Their Physical Counterparts, and the Metadata that Is Keeping Them Apart

Donald C. Force and Randy Smith

ABSTRACT

The digitization of archival collections has become ubiquitous in the modern age. Contrary to the prevalence and popularity of these virtual collections, they are not without their limitations. Archivists have not sufficiently addressed the relationship between digital surrogates and their original objects. This article reviews a project undertaken by the authors who examined forty-two digitized archival collections from seven midwestern states. The study sought to determine whether digital surrogates include sufficient metadata to enable the viewer to understand that the virtual object is a representation of a physical object, that the physical object may be accessed, and that the physical object is part of a larger collection. The article concludes that the metadata fields used to describe digital surrogates vary across repositories, as well as within the institutions; and that very little metadata provides strong connections between the virtual images and the physical materials they represent. The authors conclude by providing recommendations for how archivists might improve the linkages between digital surrogates and their physical counterparts.

KEY WORDS

Digitized archival collections, Finding aids, Metadata

Archivist[s] are active agents in constructing social and historical memory.

—Terry Cook[1]

Digitized archival collections have become ubiquitous; and it is rare to find an archives that does not display some of its archival holdings online. Researchers increasingly expect access to resources regardless of their location. Digitized archival collections may be the immediate or only source of primary and secondary information for some researchers.[2] Often, digitized archival collections offer only a glimpse of the archival materials. Accessing the materials online removes the chance encounters that occur when researchers conduct an in-person visit to a repository. When researchers view the materials in person, they may well be exposed to additional archival content that may not be available online. The very nature of an archives lends itself to support such serendipitous research experiences. Research conducted in digitized archival collections may result in a different experience than would have occurred had the same researcher conducted the same research in person.

Thus, while virtual collections may be an attractive option for researchers, the absence of adequate metadata can hinder their ability to understand the archival materials adequately or to recognize any potential distinctions between the digital surrogate and the original physical materials. The contextual information of the original physical objects—that is, the "historical trends, events, and figures related to the materials"—is essential for researchers to know that the records they are using are trustworthy; to understand the full context of both the digital and physical parts of the collection; and to identify any distinctions between the digital surrogate and the physical original.[3] This contextual information that an archivist provides about a collection provides not only descriptive data but also, possibly, "information on the attitudes reflected by the records."[4]

As examined in this article, this important contextual information may well be missing in a digital archival environment. Providing rich metadata associated with digital surrogates—digital images that represent original physical analog objects—enables users to view an original physical analog object without being in its presence. Context-rich descriptions, with detailed metadata for digital surrogates and clear linkages back to physical analog objects, are the exception rather than the rule for digitized collections. Whereas traditional finding aids often provide additional context about the physical body of records, digital surrogates that are not linked to finding aids, or that do not come with sufficient contextual metadata, lack the details and can provide inadequate information. This gap in descriptive information can potentially distort research, in large part because all physical archival items within most digitized collections will likely not be represented by equivalent digital surrogates.

This article presents the findings of a research project that considered this question of the relationship between digital and physical archives. In the project, we sought to answer three research questions:

1. Do archives provide links to finding aids in the metadata of their digital surrogates?
2. What metadata do archivists include with digital surrogates that enable users to understand that the digital surrogate is a representation of a physical object?
3. What metadata do archivists include with digital surrogates that facilitates the viewer's ability to access the physical object that is represented?

Based on a sampling of digital surrogates and an examination of their metadata, we discovered that digital surrogates often lack sufficient metadata to indicate to researchers that the repository contains the original object. Also missing is information about how researchers can contact an archives to see the original object and/or collection. We conclude that, although archivists have made great strides in enhancing the presence of their holdings online in the form of digitized archival collections, by doing so, they have potentially altered the context in which researchers access and interpret the objects they make available digitally.

Literature Review

As archivists strive to turn their analog holdings into digitized archival collections, there has been no shortage of discourse about the archival processes involved. Archivists have made it clear that three of the primary reasons for digitizing archival materials are, first, to provide access to specific materials; second, to increase exposure to the archival institution and its holdings; and third, to prolong the original materials through reduced usage. Being able to view archival materials remotely is clearly beneficial to many researchers but, as we discuss, some researchers still prefer to access the original object.

Discussions about specific digitization projects and descriptions of the processes used to create digitized archival collections dominate the archival literature.[5] Archivists have also explored more specific issues associated with creating groups of digital surrogates. For example, Krystyna Matusiak and Tamara Johnston discuss how their University of Milwaukee–Wisconsin project to digitize the American Geographical Society's nitrate negative collections helped to preserve and provide access to the materials.[6] Jean Dryden and Maggie Dickson discuss copyright challenges associated with digitized materials.[7]

Other research focuses on the digitized archival collections themselves, considering ways in which archivists can increase the discoverability of digital surrogates. Jane Zhang, for example, explores the relationship between analog materials and their digital surrogates, drawing on an examination of three digitized archival collections to argue that, while the provenance and content of archival materials may be sufficiently represented online, the order of the original physical archival materials "seems to be less crucial and less irreplaceable."[8] Surrogates of digitized archival collections are often placed online without reflecting their arrangement in the physical collection. They may also be brought together into artificial digital collections that do not indicate from which physical collection(s) the digital surrogates originated.

Jane Zhang and Dayne Mauney expand on Zhang's initial findings. They argue that archivists use two different types of archival description to describe digitized materials: traditional archival descriptive practices for contextual information about the objects and descriptive metadata for information about the digital surrogates. Zhang and Mauney contend that archival description standards are not yet designed to blend the two approaches. Therefore, both types of representation are needed, but the use of these two systems can affect how researchers interpret the digital objects. As Zhang and Mauney explain,

> not only are records reformatted into digital objects, metadata associated with digital objects may also result in structural changes in records. Newly assigned digital object metadata has the *potential* to enrich and expand the relationships among records, structurally as well as semantically.[9]

Studies also shed light on the information-seeking behaviors of historians who use digitized archival collections.[10] Many historians that Alexandra Chassanoff surveyed expressed some trepidation about relying entirely on digital surrogates; some preferred to view the physical objects represented in the digitized collections.[11] Her research mirrors sentiments expressed by several historians.[12] Donghee Sinn and Nicholas Soares, for instance, discovered that digitized archival collections are important for research, but the historians they surveyed stressed the importance of access to original documents.[13]

Historians have not reached the point where they rely extensively on digital sources. In her citation analysis of 150 published scholarly historical monographs, Kris Bronstad finds that fewer than 3 percent of references to archival materials are to digitized archival collections.[14] Suzanne Graham examines how historians use electronic resources and finds that "[f]ifty percent of historians in this sample do not believe that a digitized document is equivalent to the original."[15] She argues that this attitude may be attributed to the "associated loss of contextual information" about the original object because the participants

believe that "the complete artifact cannot be transferred into bytes."[16] Sinn and Soares draw a similar conclusion:

> Historians, no matter whether they prefer original forms of primary materials or digital collections, are similarly concerned about the disadvantages that accompany doing research exclusively with digital content. Digital content cannot provide field experience, which is considered a part of the authoritative historical research method.[17]

There are some indications that researchers will give more credence to digital surrogates as they become more prevalent.[18] Kim Martin and Anabel Quan-Haase interviewed twenty historians to understand their experiences of serendipity, or the "a-ha!" moment at which they discover a key piece of information for their research project. They learned that the physical environment, that is, the location where historians access the materials (e.g., the archives), plays an important role in how they construct their stories because the physical interaction that historians have with the materials can influence serendipity.[19] For example, one participant in the study remarked that there "is a serendipity to browsing the stacks, to being in the library and sort of seeing the book that's next to it. . . ." This ability to freely review adjacent materials helps the research process, both in the stacks of a library and in the reading room of an archives.[20] As archivist Ala Rekrut observes in her article about material literacy, "physical properties of records are a tangible site for interpretation of information from the many elements present—text, images, appearance, texture, smell, and historical context." Physical interaction with the objects, she argues, may help researchers understand them better.[21]

The importance of contextual information about any given object is not lost on historians; in fact, it is one of the pillars of their research. As Sherman Dorn remarks, the "critical traits of an archival resource for historians include custodianship and proper sourcing, and the critical traits of an online presentation of historical artifacts parallel those: care of the digital resource and clear provenance."[22] In addition to these two traits, he also stresses the importance of understanding the organization of the digital surrogates.[23] This is not to say that digital representations are detrimental to historical research. In fact, Stefan Tanaka argues that digital technology provides historians

> an opportunity to use tools that facilitate more complex, not complicated, narratives and stories of the past and how they continue to operate in our present. By bringing out such variability, we can show more of the operations of history, the stories embedded in primary data and the negotiations and decisions that lead to the structures, ideas, and social forms of our narratives.[24]

Research also establishes that users interact differently with certain types of digital objects than with physical objects. Anastasia Varnalis-Weigle finds that

digitized versions of "complex objects" (e.g., a 3-D rendering of a mask) result in the loss of some of that object's "unique intrinsic qualities" that affect the sensual, emotional, and spatiotemporal experiences of some users. Her research subjects spent a significant amount of time reviewing documentation associated with digitized objects to understand the content and context of the materials, not necessarily the object itself.[25]

Archivists argue that a significant reason researchers visit an archives and browse a collection in person is to discover what else is contained within the collection and understand how the items in the collection relate to each other.[26] This interest in understanding relationships and context applies whether the items are directly relational, that is, the items were created by a single source such as a collection of notebooks written by a single person for a single purpose (e.g., an author's multiple rough drafts of a story), or the items are indirectly relational, that is, they were created separately by more than one body but placed together in a single collection (e.g., a single person's collection containing a series of correspondence and notebooks originating from multiple people). These relationships might not be spelled out in the metadata of digital surrogates but might be observed when accessing a physical collection.

Poor documentation associated with digitized materials can prevent users from accessing the physical objects and, therefore prevent them from assessing their original context. For example, Andrea Johnson discovered that undergraduate students could not understand that a "digital object represented a 'real' object that could be accessed locally" because the digital object lacked sufficient contextual information about the original material.[27] In her study of the use of Dublin Core elements by nonarchival experts, Kathleen Fear learned that many of her participants "misunderstood" specific metadata terms, such as "relation," "publisher," or "source." She speculates that this misunderstanding can create barriers to the further exploration of archival materials and their use as evidence.[28]

Digitization undoubtedly provides access to archival materials for researchers who would not be able to visit an archives in person. And the use of digitized primary sources in combination with other technologies, such as geographic information system (GIS) tools, can offer unique and innovative perspectives on history.[29] Archivists, however, cannot overlook the fact that some researchers will continue to prefer to access the original object. In these cases, the digital surrogate may serve as a conduit for their visit to the archives. In the remainder of this article, we outline a research project designed to determine, through analysis of a sampling of digital surrogates, whether enough metadata is included in descriptions to clarify the relationship between a digital surrogate and its physical original and whether that information facilitated researchers' ability to access the physical object.

Methodology

To answer our research questions, we identified digitized archival collections from the thirteen states that comprise the Midwest Archives Conference (MAC).[30] We focused on collections from the state historical society and one midsized public university in each state. We chose state historical societies as opposed to local historical societies, on the assumption that the state institutions were more likely to possess the resources needed to create digitized collections. When selecting the academic institutions, we chose midsized universities with a student population ranging from about 10,000 to about 30,000 students—we wanted institutions large enough to undertake digitization initiatives but not so large that their initiatives would be out of scale with the other institutions studied.

To identify the universities, we used the National Center for Education Statistics' College Navigator database.[31] First, we identified the list of schools from each state that satisfied our criteria and exported the list into an Excel file. Using the "rand()" function in Excel, we assigned each school a randomized number and sorted each state's list of schools from smallest to largest randomized number. We then selected the school with the smallest randomized number and determined if that school had digitized any archival collections.

Out of a possible total of 26 institutions—one state historical society and one university from each of 13 states—we found that 21 institutions contained digitized collections: 12 universities and 9 state historical societies. We could not locate digitized collections in one university, nor could we locate digitized collections in 4 state historical societies.

To build our data set, we identified the titles and URLs of each repository's digitized archival collections, capturing the lists in Excel. We identified a total of 1,602 collections across the 21 institutions. The number of collections at each institution varied widely, from as few as 6 collections in one institution to as many as 163 in another. We discovered that 2 institutions do not organize their content by collection. Instead, they use subject terms to link objects to one or more subject categories. We retained these 2 institutions in our data set believing that the digital surrogates may still contain relevant metadata that we could use for this study.

Examining every digitized archival collection and every digital surrogate within each collection was well beyond the scope of this project. Instead, we conducted a preliminary assessment of the metadata practices of these midwestern institutions. We randomly selected 3 collections from each repository and then randomly chose 5 digital surrogates from each collection. To identify which digitized collections to examine, we assigned a randomized number to each one using Excel's "rand()" function and then sorted the collections from smallest to

largest randomized number. From these randomized lists, we chose the first 3 collections from each repository. We dismissed some collections if they did not contain materials that supported our research framework. For example, we did not include collections comprised entirely of digitized monographs or serial publications, and we did not include collections that only contain single objects, such as an artifact. We also excluded collections that consist entirely of born-digital objects and oral histories. We assumed it unlikely that finding aids would have been generated for these collections. When we encountered one of these types of collections among the first 3 collections on our list, we skipped it and moved to the next entry on the list.

We identified a total of 63 digitized archival collections, each of which contained from 5 to 112,897 digital objects. On average there were 4,212 objects in each collection. As shown in Table 1, most of the collections examined contained 1,000 or fewer digital surrogates.

Table 1. Number of Digitized Objects per the Number of Digitized Archival Collections Examined

Number and Percentage of Collections	Number of Objects
24 (40%)	<100
26 (41%)	101–1,000
10 (16%)	1,001–10,000
3 (5%)	10,001 or more

By examining 15 digitized objects from 3 different collections from each of the 21 institutions, we believed we could glimpse the metadata practices used by each institution. We theorized that each repository follows a consistent process for creating metadata for its digitized collections; therefore, the metadata would be mostly uniform within each institution's set of digitized collections. In other words, if one digitized collection does not contain a metadata field for a link to a finding aid, then it would be unlikely for any of the digitized objects from that repository to contain such a link.

In our analysis, we did not consider any metadata standards associated with the systems used to display digitized archival collections (e.g., Dublin Core); instead, we relied on only the metadata presented for the digital surrogates examined. To do this, first, we noted the total number of digital objects from each collection. Then, we used Excel's "randbetween" function to randomly generate 5 numbers within this range. For example, if the collection contained 235 objects, we used Excel to generate 5 numbers between 1 and 235 (e.g., 5, 45, 132, 200, and 233). Each number became a digital surrogate that we examined.

All the systems included for this study displayed their digital surrogates in a gridlike presentation, usually in rows of thumbnail images. The number of images on a screen vary based on the system and, in some cases, users can choose how many objects to display at once. To identify each of the 5 random numbers, we counted the images in the collection, starting with the first image on the screen and then moving horizontally across each row and then vertically through the page, until we reached the randomly selected number. For each digital surrogate examined, we copied the metadata and its content into Excel, which allowed us to conduct a quantitative analysis of the metadata. We removed elements that did not contain any content. Except for clearly misspelled terms, we did not normalize the list of metadata elements to avoid combining terms that may not be used with the same intention. This means, for example, that we did not combine the elements "Contributor," "Contributors," or "Contributor(s)" even though it would be fair to assume each represents the same or similar descriptive content.

Findings and Discussion

In this section, we provide contextual information about our data set, including a brief description of the elements found in the data set and the systems used to display the digitized archival collections. A review and discussion of each of our research questions follows.

Use of Elements

We identified 329 unique metadata elements from the 315 digitized archival collections examined. The number of metadata elements associated with the digitized objects ranges from 3 to 58. (See Appendix A for the complete list of elements identified.) On average, repositories use 18.7 of the metadata elements to describe their digital surrogates. When taking a closer look at these numbers, we see that historical societies average 21.3 elements, while digital surrogates from university collections average 16.7 elements.[32] Table 2 indicates the 10 elements most often used with these digital surrogates, the number of different repositories that use each element, and in how many collections they appear.

Table 2. Ten Most Frequently Used Elements and the Number of Times the Element Appeared among the Digitized Archival Collections Examined

Element	Total Times Used	Number of Repositories that Use Element	Number of Collections that Use Element
Title	351 (111%)	19 (90%)	53 (84%)
Description	223 (71%)	18 (86%)	44 (70%)
Rights	190 (60%)	12 (57%)	33 (52%)
Source	159 (50%)	11 (52%)	27 (43%)
Type	146 (46%)	11 (52%)	25 (40%)
Subject	144 (46%)	13 (62%)	28 (44%)
Date	140 (44%)	13 (62%)	29 (46%)
Creator	139 (44%)	16 (76%)	32 (51%)
Language	139 (44%)	13 (62%)	24 (38%)
Format	133 (42%)	11 (52%)	24 (38%)

No element is used in all 315 digital surrogates examined or by all the 21 repositories included in this study. The table also highlights another phenomenon: duplicated elements. Despite not being used in every repository or in every collection, the element "Title" exceeds the total number of objects within our data set. We found 11 collections that contain repeated instances of this element. As seen in Table 3, within our data set, the element "Title" is but one of 35 elements that appears 2 or more times among the metadata associated with individual digital surrogates; however, 26 of these elements are only duplicated once.

Table 3. List of Elements that Appeared 2 or More Times in the Metadata Associated with Individual Digital Surrogates and the Number of Times Each Element Was Duplicated within the Data Set

Element	Number of Times Duplicated	Element	Number of Times Duplicated
Title	11	Credit Line	1
Rights	4	Date	1
Format	3	Date of Original	1
Source	3	Decade	1
Description	2	Digital Format	1
Digital Identifier	2	Digital ID	1
Language	2	Digital Publisher	1
Ordering Information	2	Identifier	1
Subject	2	Object Type	1

Element	Number of Times Duplicated	Element	Number of Times Duplicated
Citation Information	1	Organization Name	1
City	1	Orientation	1
Collection	1	Publication Date-Electronic	1
Collection Name	1	Publisher-Electronic	1
Collection Website	1	Repository Institution	1
Content Statement	1	Searchable Date	1
Contributing Institution	1	Submitting Institution	1
Country	1	Type	1
Credit	1		

The institutions examined use different systems to display their digitized archival collections. As shown in Table 4, most of the repositories use CONTENTdm. For the 3 repositories that use a combination of databases, at least one of the collections appears in CONTENTdm, with at least one other appearing in a system that seems to have been created by the repository.

Table 4. The Average Number of Metadata Elements Used per Digital Surrogate According to the Type of System Used per Repository

System	Number of Repositories	Average Number of Elements per Digital Surrogate
CONTENTdm	13 (62%)	18.8
Internally Created	3 (14%)	12.3
Combination	3 (14%)	12.8
Digital Commons	1 (5%)	22.4
Flickr	1 (5%)	50.4

The high number of elements associated with the use of Flickr is the result of Flickr allowing users to access the technical metadata associated with a digital object.

The differences in metadata are not the result of the system used to display the digital surrogates. While Flickr and Digital Commons account for 2 of the collections with the highest variation in metadata used among the 5 digital surrogates examined, at least one collection in each system uses the same metadata for each of the 5 digital surrogates. Even those repositories that use CONTENTdm vary in the metadata associated with the digital surrogates. This indicates that even when repositories rely on a system that uses a standardized metadata schema (CONTENTdm uses Dublin Core), enough flexibility remains

to potentially negate the value of the standard, or the standard is not sufficient for describing digitized archival objects.

Links Between Metadata and Finding Aids

A link to a finding aid may be one of the clearest indications of the relationship between virtual and physical (as well as revealing other materials within the physical collection). We contend that one of the challenges with digitized archival collections is that viewers often understand the metadata fields in the context of the digital surrogate they see on the screen, but they may only associate the fields with that specific surrogate and not realize that they represent a physical object that might be materially different in shape, form, content, or context. Therefore, links to finding aids are important. Finding aids function as vital contextual resources for understanding physical collections; and the more precise the information in a finding aid, the easier it will be for researchers to make an accurate assessment of the value of the materials. Often, critical information is found in a collection's finding aid because the finding aid is designed to describe not only the materials within the collection but also contextual information about the origins and provenance of the records, along with other administrative information.

As Table 5 shows, our data set contain a very few references to the elements where we would expect a finding aid to appear. These include the elements "Finding Aid," "Collection Finding Aid," and "Archival Collection URL."

Table 5. Frequency of Elements Used that Reference Finding Aids or a URL to the Physical Archival Collection

Metadata Element	Total Times Used	Number of Repositories that Use Element	Number of Collections that Use Element
Finding Aid	5	1	1
Collection Finding Aid	7	1	1
Archival Collection URL	13	1	3

Similar to the instance of the "Title" element, the element "Collection Finding Aid" appears twice in one collection. These are not the only instances where links to finding aids appear. One repository includes a link to a finding aid in the description associated with the "Repository Collection" element for each of its 15 digitized objects.

One explanation for the general absence of links to finding aids may involve how the archival institutions compile their digitized archival collections. The majority of the digitized archival collections in our data set appear to be "artificial," that is, collections that include digital surrogates that represent

two or more physical archival collections, either within the hosting repository or from multiple repositories and/or different institutions. It is apparent that archivists designed these collections around a theme, topic, geographic area, time period, and so on. Archivists may focus on the content of the materials and their virtual relationship to each other at the expense of providing metadata that explains their contextual information (e.g., a finding aid). A finding aid to an individual collection that has little or no connection to a theme-based digitized collection may not be seen as an important element to associate with the individual digital surrogates.

Links between Metadata and Physical Object

While a link to a finding aid may be one of the clearest indications of this relationship between virtual and physical (as well as revealing other materials within the physical collection), other metadata fields could possibly indicate to users that the virtual object represents a physical object. From the complete list of metadata fields, we believe that the following 9 elements best indicate that a digital surrogate represents a physical object:

- Format of Original
- Format Original
- Original Item Condition
- Original Item ID
- Source Collection Number
- Original Item Medium
- Original Item Size
- Original Medium
- Source Collection Name

We then assessed which repositories use these fields, as shown in Table 6.

Table 6. Frequency of Elements Used that Indicate that a Digital Surrogate Represents a Physical Object

Metadata Element	Total Times Used	Number of Repositories that Use Element	Number of Collections that Use Element
Format of Original	29	2	5
Format Original	14	1	3
Original Item Condition	5	1	1
Original Item ID	5	1	1
Original Item Medium	10	1	2
Original Item Size	10	1	2
Original Medium	5	1	1
Source Collection Name	15	1	3
Source Collection Number	15	1	3

Although these elements are clearly used with more frequency than, say, the "Finding Aid" or "Collection Finding Aid" elements, their use is hardly widespread among the repositories examined. Based on these findings, we can conclude that archival institutions use metadata primarily to describe digitized objects with a minimal amount of metadata used to connect them to their original counterparts. In other words, contextual information about these digital surrogates, such as their provenance, is mostly absent, thereby potentially obscuring their true evidentiary value.

Metadata and Access to Archival Institution

Assuming that users know that the digital surrogate represents a physical object, the final research question we asked about these digitized collections was whether the metadata facilitates user access to the object in the archives. Although archivists should ensure that their websites are appropriately structured and offer information to be of value to researchers, not all researchers encounter digital surrogates via the digitized archival collection's main web portal or via the archives' homepage, so it is important to consider what metadata archivists might include in a description to enable users to access the physical object represented by the digital surrogate.[33] To answer our third research question, we looked at the use of four specific metadata fields from our data set:

1. Access Information
2. Address
3. Repository
4. Repository Institution

Table 7 summarizes the metadata fields identified and the number of digitized archival collections used in each field to capture repository information.

Table 7. Frequency of Elements Used that Provide the Archives Contact Information

Metadata Fields	Total Times Used	Number of Repositories that Use Element	Number of Collections that Use Element
Access Information	3	1	1
Address	3	3	3
Repository	92	8	17
Repository Institution	24	1	3

Overall, 12 of the 21 repositories use one or more of these 4 elements, but only 2 of the repositories use both the Address and Repository elements. For each of these 2 repositories, however, this combination of elements only appears in one of each of their 15 digital surrogates examined.

A closer analysis of the content for each of the 4 elements reveals some contact information but it is far from complete. For example, the 3 institutions that use the Address element only provide the street address, presumably of the archival institution that houses the original object. For the archival institutions that use Repository or Repository Information, 2 institutions include the name, city, and state of the repository, while all other uses of the elements mention the name of the institution and/or the university with which it is affiliated. Only the repository that uses the Access Information element (for only one of its 15 digital surrogates, we might add) includes contact information containing an email address and a phone number.

In none of these collections was a URL to the repository's homepage provided. In fact, we noticed that only 8 of the 21 repositories provide hyperlinks at the top of the webpage that direct users to the homepage (and one of the URLs did not work). Even clicking the logo of the institution (usually located at the very top of the screen) would often direct us to the repository's digital collection portal, not the homepage of the repository. The few URLs within the digital surrogates examined directed us to another point within that digitized archival collection, not to a page external to the collection.

The lack of consistent contact information about institutions is concerning because it may hinder access to the original object if researchers want to know more about it. Contact information is also important because many archival institutions use systems that aggregate digital surrogates from multiple digitized archival collections that may have been created by multiple institutions. When using these systems, researchers may be unclear which institution possesses the digital object, let alone its physical counterpart. Moreover, in some cases, aggregators may obscure the original content and context of the object because of the migration from one system to another. An aggregator may have a predetermined set of metadata fields and either convert nonconforming fields and content to the system's fields or hide the information altogether. Therefore, metadata that provides contact information of the institution that holds the original object and/or a link to a finding aid from the collection in which the object exists (which would also presumably include the contact information of the institution) become paramount in aggregators for digital collections so researchers can easily inquire about the virtual object and its physical counterpart when necessary.

Summary of Findings

Digitized archival collections are valuable to historical research, but they may also be detrimental to such research if the context of their creation and display is not properly understood. Users need to be made aware that what they

see online may not be the whole story of an object, and they need to be provided with adequate details to know what they are seeing, what they are not seeing, and where they can go for more information.

Archivists need to ensure that the descriptive fields associated with digital surrogates provide sufficient information about the relationship of the digital surrogates to their original physical objects. This does not mean archivists must scan all the materials in every collection; that is rarely possible or desirable. Archivists must make choices, but they also have a responsibility to make users aware of gaps in collections, which will affect users' understanding of the context of the archives in a digital environment. Metadata is essential to helping users overcome any barriers that might arise through a misinterpretation of the digital surrogate.

To help archivists provide this necessary context in the metadata of their digitized archival collections, the next section offers some recommendations regarding what metadata archivists should consider including. We intend these suggestions to help inform users that what they see online may be different from what they would see if they viewed the physical materials in person.

Recommendations

Based on the findings of this project, we present several recommendations for archival institutions involved with creating digitized archival collections or updating current digital collections. These guidelines are intended to facilitate access to the original physical object that a digital surrogate represents, so the suggestions do not consider all the metadata that might be captured for each digital surrogate. For example, we do not discuss metadata needed to capture information about the digital surrogate itself (e.g., digital format, resolution, date of the object, etc.), as those fields do not relate specifically to our priority here, which is to support the connection of digital surrogates with physical counterparts. Note that the recommendations offered here are still untested; future research should be conducted to explore further the metadata needed by users to link digital surrogates with physical counterparts and the archival actions that could be taken to support that need.

General Recommendations

Here are general recommendations for the use of metadata fields:

- **Use clear, concise, and consistent field names.** We found several examples of a lack of clarity about the purpose of the metadata field. For example, some repositories use the term "Source" to denote an

item identifier, but others use it to represent a collection number. This confusion could be alleviated by altering the field names to, for example, "Source ItemID" or "Source Collection Number" and adding metadata accordingly. We encountered several redundant field names. Clarifying the field names can help users understand the presented metadata. We realize that due to the limited fields available in Dublin Core, this may not be a possibility.

- **Minimize ambiguous fields.** Creating clear and concise field names should alleviate ambiguity, but it is worth repeating that each metadata field should correspond to only one concept. For example, the "Source ItemID" field for a digital surrogate might consist of a combination of the accession number and the box and folder numbers. This construction is fine, but if such a combination is utilized to create a "Source ItemID" number, then a separate metadata field should be included for the accession number alone and another one for the box/folder numbers. In other words, the accession, box, and folder numbers should still have their own unique metadata fields to alleviate ambiguity. Additionally, the digital surrogate itself can have its own ID in a field labeled "Digital ItemID."
- **Avoid duplicate elements.** To keep the metadata associated with individual digital surrogates as concise as possible, every effort should be made to avoid the duplication of metadata elements, for example, making sure "Title" only appears once. Reducing duplicate elements may also prevent the archivist from unintentionally associating different content for the same element within the metadata. We acknowledge that avoiding duplicated elements may not always be possible because of the system being used to display the virtual materials.

Recommendations for Collection-Specific Metadata Fields

Here are recommendations for the use of collection-specific metadata fields:

- **Accession number.** This number is primarily for the repository as users may not know what it indicates. It may not be necessary for a repository to include this number. If the number is included, it should have its own uniquely labeled field.
- **Collection number.** This number is more for the repository and less for users as they may not know its meaning. The number also implies that the digital surrogate represents a physical object because it exists within a physical set of materials. It should be given a field label such as

"Source Collection Number" to avoid the assumption that "Collection Number" implies "Digital Collection Number."

- **Box and folder number.** Similar to the collection number, this offers an additional piece of information that may facilitate retrieval of the original object that users may reference when contacting the repository, and it indicates the existence of a physical object. Such metadata can be displayed in two ways: either combined into a single field labeled "Source Box and Folder Number" or as two separate fields labeled "Source Box Number" and "Source Folder Number."
- **Format of original object.** This field indicates that a physical version of the digital surrogate exists. This field is important as it can help distinguish between three-dimensional objects and two-dimensional prints of objects. When users view a picture of an object on a screen, they are only seeing it as a flat two-dimensional digital representation. They do not necessarily know if the original is just a flat two-dimensional image of an object (e.g., a photograph) or a physical three-dimensional object due to the nature of digital images. A field labeled "Original Object Format" would help provide some context to what users observe when looking at a digital surrogate.
- **Number of objects digitized**. Understanding the context of a digital surrogate is essential information for researchers; without that information, they cannot understand whether they are seeing all or some of the digital version of a physical collection and which portions they are or are not seeing. To help researchers understand this context, archivists can provide information about how many digitized objects from the original collection have been scanned. This number may be an approximate percentage (e.g., 90%) or a precise number (e.g., 50 of 67).
- **Original collection name.** Several repositories provide a field called "Original Collection Name." This information is extremely valuable, especially in the case of "artificial" collections, where the name of the digitized archival collection may not mirror the name of the collection associated with the original physical objects. We recommend including the original collection name as its own field label as in "Original Collection Name."
- **Digital surrogate statement.** Each digital surrogate should contain a statement that indicates it represents a physical object, as opposed to being a born-digital object or an altered digital file (e.g., such as one that has undergone significant digital image manipulation or editing). This statement may be placed in a field generally utilized for notes about the object.

- **URL to finding aid.** When possible, the digital surrogate should include a hyperlink to a digital or digitized finding aid that identifies the location of the original physical object. We acknowledge that making this link may be challenging for repositories because URLs may change during system updates, effectively rendering the links moot unless the digital finding aid has an established DOI (digital object identifier).

Repository-Specific Metadata Fields

Here are recommendations for the use of repository-specific metadata fields:

- **Name of the repository.** A metadata field should allow users to know which repository is responsible for the digitized archival collection and (as appropriate) also responsible for the corresponding physical object. This information is especially helpful when digital surrogates are brought together from disparate digitized archival collections, often forming a new digital collection.
- **Repository contact information.** At the very least, a URL to the repository's homepage should be included in the metadata field that directs users to the "Contact Us" or "About Us" page. Other useful information includes the repository's phone number, fax number, email address, and street address. This information is beneficial if users are accessing the digital surrogate through a digital aggregator, or if another institution or individual is utilizing the digital surrogate as part of an artificial digital collection.
- **URL to collection landing page *and* archival repository.** Research indicates that if users cannot locate the information that they want after a few clicks, they may not be satisfied and may even abandon the search.[34] The use of multiple hyperlinks may be key to retaining users, allowing them to find information via different paths. Each URL used should be distinctly labeled to avoid confusion, and each URL should direct users to their chosen location easily. If the collection's landing page contains information about the digitized archival collection, including links to the repository and the relevant finding aids, then this should be the URL used. A combination of different metadata fields and URLs, each directing users to key information, may alleviate some challenges if one of the URLs become inactive.

Limitations

This study was not without its limitations. We identified a small sample size of repositories, digitized collections, and digital surrogates so we could gather as much research data as possible to generate discussion about the creation of digitized archival collections and their relationship to physical collections. Therefore, the findings of this study cannot be generalized to all digitized collections created by all archives.

We also assumed that every digital surrogate that we examined represented a physical object that belonged to a processed collection complete with a finding aid. Not only is it possible that our random sampling technique identified collections that do not truly represent archival collections, but that the physical collections to which the digital surrogates belong are not fully processed or have finding aids associated with them.

Finally, the study makes some assumptions about the users of digitized archival collections, including assumptions about the metadata fields users would want or need to draw connections between digital surrogates and physical objects. We also assumed that not all users understand that a virtual object is a representation of a physical object, or that the physical object may be viewed in person within the context of other materials, which may themselves also be digitized. No study has made this connection, and additional investigation is necessary to determine if archivists should address this finding in their descriptive and digitization work.

Conclusion

Digitized archival collections provide greater access to primary source materials, albeit in an altered form. Great potential exists for revealing more clearly the relationship of one original item to other items in the same collection, which can help researchers understand more fully the collection. Archivists, however, need to be more cognizant of the relationship between digital surrogates and the physical objects they represent. Our study indicates that some archival institutions either lack metadata fields that provide clear linkages to physical collections and finding aids, or do not use metadata fields consistently enough to ensure that users see accurate and precise information across collections or across institutions. In short, digitized archival collections rarely represent a clear one-to-one relationship with a physical collection.

Digitized archival collections have become an essential component of archival service. But the process of creating a digitized archival collection goes beyond simply scanning a physical object and placing it online. Providing context to digital surrogates is a key part of the process. The topic of describing

and managing digital surrogates has been raised before, as shown in the literature review, but in the intervening years since those findings were published, little has been done to address the problem of how metadata represents digital surrogates.[35]

We wonder if archivists take it for granted that the facsimiles they put online are easily understood as such, or as being part of a larger physical collection. Users of these digital surrogates may never be able to view the original archival objects in relation to other objects within a corresponding physical collection. At the least, users should be made aware that the digital surrogate is a representation of a physical object from an archival collection, and, as this study notes, this metadata is often lacking with digital surrogates.

Archivists do not yet know enough about how users access digitized archival collections. Whether or not they encounter digital surrogates through an archives' main portal makes a difference in the type of information they find. Archivists cannot assume that users will retrieve information only through the archives' webpage, or by some other path, and whether users will develop a good understanding of the context in which the digital surrogates exist. Various websites and search engines may link directly to the digital surrogates themselves; when this happens, users may not see or have easy access to the main description for the digital collection. As a result, the digital object loses not only the context the original, physical collection provides, but also the context other digital surrogates in the collection provide.

Not only do archivists know very little about their researchers' use of digitized archival collections, they also lack knowledge about the metadata needs of their users. Further research is needed to determine whether more robust metadata records make them more accessible and usable, or whether this even leads to clearer contextualization between the digital surrogates and their physical counterparts.

Users should be reminded that the digital surrogates they encounter may only be components within a broader set of physical materials. By providing additional contextual information about the original physical objects, archivists can help increase the awareness, value, and use of archival materials—both digital and physical. Linking digital surrogates clearly to a digital finding aid using accurate and precise metadata fields is one way to provide this awareness.

Appendix A: Metadata Elements from the Digitized Archival Collections Examined for this Study

About
Academic Year
Access Information
Accession Number
Additional Authors/Creators
Additional Information
Address
Alternative Title
Application Record Version
Architect
Archival Collection
Archival Collection URL
Archival Number
Archivist Notes
Artist
Athletic Program
Author
Author(s)
Author/Creator
Biography/History
Bits Per Sample
Building Type
Built Work Geographic Location
Built Work Name
Built Work Street Address
Buy a Copy
Call Number
Caption
Caption Writer
Capture Specifications
Categories
Circus
Citation
Citation Information
City
City/Place
Client
Collection
Collection Finding Aid
Collection Name
Collection URL
Collection Website
Color Mode
Color Space
Color/B&W
Community
Components Configuration
Compression
Content Statement
Continent
Contributing Institution
Contributor
Contributor(s)
Contributors
Copyright
Copyright Flag
Copyright Notice
Copyright status
Country
Country/Nation
County
Coverage
Creation Date
Creator
Creator Name
Creator Role
Creator Tool
Creator(s)
Credit
Credit Line
Culture
Date
Date and Time (Digitized)
Date and Time (Modified)
Date Created
Date Digital
Date Digitized
Date of Drawing Execution
Date of Original
Date of Photograph

Date of Work
Date Original
Date Published
Decade
Derived From Document ID
Derived From Instance ID
Description
Destination IN Journey
Digital Collection
Digital Collection Name
Digital Collection Number
Digital Content Type
Digital Date
Digital File Format
Digital Format
Digital ID
Digital Identifier
Digital Publisher
Digitization Equipment
Digitization Specifications
Dimensions
Dimensions of Original
Directory.Related work
Document ID
Document Title
Drawing Title
Drawn By
Duration
Edition
Envelope metadata
Exif Version
Extent
File Format
File Name
File Properties
File Size
Files
Finding Aid
Flashpix Version
Folder Description
Format
Format (Digital)
Format Digital
Format of Original
Format Original
Full Text
General Region
General Subject
Genre
Geographic Coverage
Geographic Feature
Geographic Location
Geographical Area
Global Altitude
Global Angle
Holding Institution
ICC Profile Name
ID Number
ID Number Note
Identification #
Identifier
Image Count
Image Description
Image Height
Image ID
Image Materials/Technique
Image Number
Image Size
Image Width
Inscription
Instance ID
IPTC Digest
Issue Date
Issue Day
Issue Month
Issue Present
Issue Year
Item ID
Item Identifier
Item Number
Item Type
JFIF Version
Job Number
Keywords
KSHS Identifier
Language

- LCCN
- Learn More
- Legacy IPTC Digest
- License
- Local Identifier
- Locally Assigned Subject

Headings

- Location
- Location Depicted
- Location Letter Sent From
- Marked
- Material
- MDL Identifier
- Measurements
- Media Type
- Medium
- Metadata Date
- Minnesota Digital Library
- Assigned Subject Heading
  - Month
- Name
- Native Digest
- Notes
- Object Name
- Object Type
- OCLC number
- Online Provider
- Ordering Information
- Organization
- Organization Name
- Organizations Depicted
- Orientation
- Original Collection
- Original Collection Name
- Original Date
- Original Dimensions
- Original Document ID
- Original Format
- Original Format Number
- Original Format Type
- Original Item Condition
- Original Item ID
- Original Item Location
- Original Item Medium
- Original Item Size
- Original Medium
- Original Source
- Other Title
- Owning Institution
- OwnInst
- Page
- Page Number
- Part of
- PDF Pages
- People Depicted
- People Pictured
- Performer
- Personal Name
- Photo from album
- Photographer
- Photographer's Note
- Photometric Interpretation
- Physical Description
- Physical Dimensions
- Physical Format
- Pixel Height
- Pixel Width
- Place
- Place of Publication
- PlaceKept
- Places Depicted
- Preservation File Name
- Project Number
- Provenance
- Province
- Publication Date
- Publication Date-Electronic
- Publication Name
- Publisher
- Publisher of Original
- Publisher-Electronic
- Quantity
- Recommended Citation
- Region
- Related Documents
- Related Items

Related Material
Relation
Repository
Repository Collection
Repository Institution
Representation/View 1
Representation/View 2
Reproductions and permissions
Required Credit Line
Resolution
Resource Identifier
Resource Type
Rights
Rights and Permissions
Rights Management
Rights Statement
RightsMgmt
Roster
Samples Per Pixel
Scale
Score
Searchable Date
Series Description
Size of Original
Snyder ID No.
Software
Source
Source Collection Name
Source Collection Number
Source Creation Date
SSID
Stadium
State
State/Province
Steward
Subject
Subject (TGM)
Subject (AAT)
Subject (LCSH)
Subject (LCTGM)
Subject (Local)
Subject (Names)
Subject (TGM)
Subject Headings
Subject Term (LOC)
Subject Term (Local)
Subject Term (TGM)
Subject(s)
Subject-Persons
Subjects
Subject-Topics (TGM)
Submitting Institution
Supplemental Categories
Technique
Time Period
Title
Title Number
Title Volume
Title(s)
Title.Alternate
Title.Alternate2
Topic Code
Transcript
Type
Type (DCMI)
Type of Material
URL
Use Statement
UseFormat
Variant Title
Way Number
Web Statement
Work Type
Writer
XMPToolkit
YCbCr Positioning
Year

Notes

[1] Terry Cook, "Remembering the Future," *Archives, Documentation, and Institutions of Social Memory*, ed. Francis X. Blouin Jr. and William G. Rosenberg (Ann Arbor: University of Michigan Press, 2009), 170.

[2] William Landis, "Archival Outreach on the World Wide Web," *Archival Issues* 20, no. 2 (1995): 129–30.

[3] Abigail R. Griner, "Where's the Context? Enhancing Access to Digital Archives," *Provenance, Journal of the Society of Georgia Archivists* 26, no. 1 (2008): 60–61. See also Matt Gorzalski, "Archivists and Thespians: A Case Study and Reflections on Context and Authenticity in a Digital Project," *American Archivist* 79, no. 1 (2016): 171, https://doi.org/10.17723/0360-9081.79.1.161.

[4] Kathleen Roe, *Arranging & Describing Archives & Manuscripts* (Chicago: Society of American Archivists, 2005), 15.

[5] Andrea Watson and P. Tony Graham, "*CSS* Alabama *Digital Collections*: A Special Collections Digitization Project," *American Archivist* 61, no. 1 (1998): 125, https://doi.org/10.17723/aarc.61.1.j037j5v9q78x9012; and Ben Walsh and Andrew Payne, "Stories, Sources and New Formats: The Challenges of Digitising Large Archive Source Collections," *Historian*, no. 129 (2016): 28–32.

[6] Krystyna Matusiak and Tamara Johnston, "Digitization for Preservation and Access: Restoring the Usefulness of the Nitrate Negative Collections at the American Geographical Society Library," *American Archivist* 77, no. 1 (2014): 241–69, https://doi.org/10.17723/aarc.77.1.gr125w680h64u277.

[7] Jean Dryden, "Just Let It Go? Controlling Reuse of Online Holdings," *Archivaria* 77 (Spring 2014): 43–71; and Maggie Dickson, "Due Diligence, Futile Effort: Copyright and the Digitization of the Thomas E. Watson Papers," *American Archivist* 73, no. 2 (2010): 626–36, https://doi.org/10.17723/aarc.73.2.16rh811120280434.

[8] Jane Zhang, "Archival Representation in the Digital Age," *Journal of Archival Organization* 10, no. 1 (2012): 45–68.

[9] Jane Zhang and Dayne Mauney, "When Archival Description Meets Digital Object Metadata: A Typological Study of Digital Archival Representation," *American Archivist* 76, no. 1 (2013): 190, https://doi.org/10.17723/aarc.76.1.121u85342062w155. Emphasis added.

[10] Wendy M. Duff and Catherine A. Johnson, "Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives," *The Library Quarterly: Information, Community, Policy* 72, no. 4 (2002): 472–96; Helen R. Tibbo, "Primarily History in America: How US Historians Search for Primary Materials at the Dawn of the Digital Age," *American Archivist* 66, no. 1 (2003): 9–50, https://doi.org/10.17723/aarc.66.1.b120370l1g718n74; Paul Conway, "Modes of Seeing: Digitized Photographic Archives and the Experienced User," *American Archivist* 73, no. 2 (2010): 425–62, https://doi.org/10.17723/aarc.73.2.mp275470663n5907; Hea Lim Rhee, "Unique Qualities of Historians' Information-Seeking Behavior in Historical Research," *Proceedings ASIS&T '10 Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem* 47 (2010): 1–2; Paul Darby and Paul Clough, "Investigating the Information-Seeking Behaviour of Genealogists and Family Historians," *Journal of Information Science* 39, no. 1 (2013): 73–84; Hea Lim Rhee, "Modelling Historians' Information-Seeking Behaviour with an Interdisciplinary and Comparative Approach," *Information Research* 17, no. 4 (2012), http://www.informationr.net/ir/17-4/paper544.html#.Wd5MNtOPJPY; and Alex Poole, "Archival Divides and Foreign Countries? Historians, Archivists, Information-Seeking, and Technology: Retrospect and Prospect," *American Archivist* 78, no. 2 (2015): 375–433, https://doi.org/10.17723/0360-9081.78.2.375.

[11] Alexandra Chassanoff, "Historians' Experiences Using Digitized Archival Photographs as Evidence," *American Archivist* 81, no. 1 (2018): 147, https://doi.org/10.17723/0360-9081-81.1.135; and Alexandra Chassanoff, "Historians and the Use of Primary Source Materials in the Digital Age," *American Archivist* 76, no. 2 (2013): 469–71, https://doi.org/10.17723/0360-9081-81.1.135.

[12] Jen Boyle, "Treading the Digital Turn: Mediated Form and Historical Meaning," *Journal for Early Modern Cultural Studies* 13, no. 4 (2013): 79–90. See also Mary Hammond, "Tracking Pirates through the Digital Archive: The Case of Dickens," *The Yearbook of English Studies* 45 (2015): 178–95, https://doi.org/10.5699/yearenglstud.45.2015.0178; Ellen Cushman, "Wampum,

Sequoyan, and Story: Decolonizing the Digital Archive," *College English* 76, no. 2 (2013): 115–35; and Richard Abel, "The Pleasures and Perils of Big Data in Digitized Newspapers," *Film History* 25, nos. 1–2 (2013): 1–10.

[13] Donghee Sinn and Nicholas Soares, "Historians' Use of Digital Archival Collections: The Web, Historical Scholarship, and Archival Research," *Journal of the Association for Information Science and Technology* 65, no. 9 (2014): 1801.

[14] Kris Bronstad, "References to Archival Materials in Scholarly History Monographs," *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage* 19, no. 1 (2018), https://rbm.acrl.org/index.php/rbm/article/view/16982/18723.

[15] Suzanne R. Graham, "Historians and Electronic Resources: Patterns and Use," *Journal of the Association for History and Computing* 5, no. 2 (2002), http://hdl.handle.net/2027/spo.3310410.0005.201.

[16] Graham, "Historians and Electronic Resources."

[17] Sinn and Soares, "Historians' Use of Digital Archival Collections," 1806.

[18] Sinn and Soares, "Historians' Use of Digital Archival Collections," 1799–1800; and Jennifer Rutner and Roger C. Schonfeld, "Supporting the Changing Research Practices of Historians," final report, ITHAKA S+R (2012), 40–41.

[19] Martin Kim and Anabel Quan-Haase, "The Role of Agency in Historians' Experiences of Serendipity in Physical and Digital Information Environments," *Journal of Documentation* 72, no. 6 (2016): 1017.

[20] Kim and Quan-Haase, "The Role of Agency in Historians' Experiences of Serendipity in Physical and Digital Information Environments," 1017.

[21] Ala Rekrut, "Material Literacy: Reading Records as Material Culture," *Archivaria* no. 60 (Fall 2005): 31.

[22] Sherman Dorn, "Is (Digital) History More than an Argument about the Past?," in *Writing History in the Digital Age*, ed. Jack Dougherty and Kristen Nawrotzki (Ann Arbor: University of Michigan Press, 2013), 28, http://www.jstor.org/stable/j.ctv65sx57.7.

[23] Dorn, "Is (Digital) History More than an Argument about the Past?," 33, fn 25.

[24] Stefan Tanaka, "Pasts in a Digital Age," in *Writing History in the Digital Age*, 35–46, http://www.jstor.org/stable/j.ctv65sx57.8.

[25] Anastasia S. Varnalis-Weigle, "A Comparative Study of User Experience between Physical Objects and Their Digital Surrogates," *Journal of Contemporary Archival Studies* 3, no. 1 (2016): 15–17. See also Daniel G. Dorner, Chern Li Liew, and Yen Ping Yeo, "A Textured Sculpture: The Information Needs of Users of Digitised New Zealand Cultural Heritage Resources," *Online Information Review* 31, no. 2 (2007): 173–74, https://doi.org/10.1108/14684520710747211.

[26] Luciana Duranti, "The Archival Bond," *Archives and Museum Informatics* 11 (1997): 215–16.

[27] Andrea Johnson, "Users, Use and Context: Supporting Interaction between Users and Digital Archives," in *What Are Archives? Cultural and Theoretical Perspectives: A Reader*, ed. Louise Craven (Burlington, VT.: Ashgate Publishing Limited, 2008), 146.

[28] Kathleen Fear, "User Understanding of Metadata in Digital Image Collections: Or, What Exactly Do You Mean by 'Coverage'?," *American Archivist* 73, no. 1 (2010): 51, https://doi.org/10.17723/aarc.73.1.j00044lr77415551.

[29] Stephen Robertson, "Putting Harlem on the Map," in *Writing History in the Digital Age*, 186–97, http://www.jstor.org/stable/j.ctv65sx57.20.

[30] The Midwest Archives Conference (MAC) region consists of "13 heartland states: Illinois, Indiana, Iowa, Kansas, Kentucky, Michigan, Minnesota, Missouri, Nebraska, North Dakota, South Dakota, Ohio, and Wisconsin," https://www.midwestarchives.org/what-is-mac.

[31] See the National Center for Education Statistics' College Navigator database at https://nces.ed.gov/collegenavigator.

[32] The high number of elements associated with the state historical societies occurs because one repository uses Flickr, which makes the technical metadata available for each surrogate. Flickr-based surrogates averaged 50.5 metadata elements. The removal of this institution from the data set reduces the average number of elements for the state historical societies examined to

17.7 elements, which is more in line with the number of elements from the digitized collections examined from universities.

[33] Hyejung Han and Dietmar Wolfram, "An Exploration of Search Session Patterns in an Image-based Digital Library," *Journal of Information Science* 42, no. 4 (2016): 13–14; Chassanoff, "Historians' Experiences Using Digitized Archival Photographs as Evidence," 146–47; and Fear, "User Understanding of Metadata in Digital Image Collections," 47.

[34] Scott P. Pitol, "Evaluating How Well an Archival Website Allows a Researcher to Prepare for an On-Site Visit," *American Archivist* 82, no. 1 (2019): 137–54, https://doi.org/10.17723/0360-9081-82.1.137; and Daniel Belanche, Luis V. Casaló, and Miguel Guinalíu, "Website Usability, Consumer Satisfaction and the Intention to Use a Website: The Moderating Effect of Perceived Risk," *Journal of Retailing and Consumer Services* 19, no. 1 (2011): 124–32.

[35] See Zhang, "Archival Representation in the Digital Age," and Zhang and Mauney, "When Archival Description Meets Digital Object Metadata."

ABOUT THE AUTHORS

Donald C. Force is an associate professor at the University of Wisconsin–Milwaukee School of Information Studies, where he teaches courses on archives and records and information management. His areas of research involve archives and records management pedagogy, the use of digital archival collections, and legal issues associated with records management practices in North America. He earned his PhD from the University of British Columbia and his master of library science and master of information science degrees from Indiana University Bloomington. He also holds an MA in history from Southern Illinois University, Carbondale.

Randy N. Smith is the preservation and access manager for the William R. & Norma B. Harvey Library at Hampton University. Previously, he worked at the Peter H. Raven Library at the Missouri Botanical Garden as a metadata specialist and rare book cataloger, and there he received preservation training. Smith received his MLS specializing in archival theory from the University of Wisconsin–Milwaukee in 2015. He also holds a certificate in cataloging from Rare Book School and assists with rare book cataloging at Hampton University.