When Archival Description Meets Digital Object Metadata: A Typological Study of Digital Archival Representation

Jane Zhang and Dayne Mauney

ABSTRACT

The relationship between archival description and descriptive metadata of digital objects has not been explicitly discussed in the literature. The discussion will enhance our understanding of the relationship between archival context and digital content, a significant topic in a networked digital environment. The data collected in this study show that archivists have made conscious efforts to build connections between archival description (context) and digital items (content), and, as a result, distinct representation models have emerged from digital archival practice. However, at the level of integration of archival context and digital content in digital archival representation, archivists are challenged to achieve an ultimate goal of making digital archives more accessible and better contextualized in the digital world.

KEY WORDS
Description, Metadata, Online Collections

An archival record, by nature, is not meant to exist on its own. The message it carries, the action it entails, and the consequence it incurs can only become fully comprehensible when the record has been associated with the juridical, administrative, and social context responsible for its creation, the function(s) it supports, and other records to which it relates in the business process. Archival science, therefore, has traditionally been known for "its analysis of aggregates of records in terms of their documentary and functional relationships."[1] In practice, the traditional aggregate-based archival approach is embodied in its modern formula of five-level representation hierarchy,[2] in multilevel archival description rules highlighted in the *General International Standard Archival Description* (*ISAD/G*),[3] and in the Encoded Archival Description (EAD) schema—the first data structure that accommodates a hierarchical structure for the presentation of online finding aids.[4]

Archival multilevel description featuring record aggregates provides a holistic representation of the context and content of an archival collection, pointing users from the descriptions of its creators and its organizational structures to its contents. Multilevel description proceeding from the general (highest) level to successively more specific (lower) levels ensures that records in an archival collection are described collectively and placed in the context in which they are created and used. In this practice, archival description, despite its name, does less to describe records, but more to contextualize them. Records as individual items are usually not the focus of attention in multilevel archival description.

The digitization movement, to some extent, has shifted that attention. In the digitization process, archival records are turned into digital objects for archivists to handle as individual items (i.e., manage, describe, preserve, provide access to), which is different from the tasks they traditionally perform. How do archivists describe digitized archival items so as to enhance their findability and discoverability? What major approaches do archivists adopt to incorporate item-level metadata into archival description in digitized archival collections? How would the well-established archival tradition be adjusted (or not) to absorb the newly created descriptive metadata of digitized archival collections? A unique perspective of multilevel archival description in relation to multifaceted digital object metadata in this study will help enhance our understanding of the relationship between archival context and digital content, a significant topic in a networked digital environment.

Literature Review

The relationship between archival description and descriptive metadata of digital objects has not been explicitly discussed in the literature. Early literature on electronic records highlights the contextual and structural value of metadata

associated with electronic records, as shown in David Bearman's discussion of item-level control and electronic recordkeeping,[5] as well as Wendy Duff's, Heather MacNeil's, and David Wallace's discussions on the pros and cons of metadata replacing archival description.[6] Research initiatives continued to focus on the identification, authentication, and preservation of the contextual relationships of records by means of recordkeeping metadata.[7] Adrian Cunningham cited the previous research to define recordkeeping metadata as "event-oriented metadata in an object-oriented world" and concluded that recordkeeping metadata is "fundamentally different from and infinitely more complex than resource discovery metadata and preservation metadata."[8] Recent archival literature also reiterated the recordkeeping value of item-level metadata, as shown in Greg Bak's article calling for "a transformation of recordkeeping and archival practice through an expanded definition of records classification and through item-level management of electronic records."[9]

The lack of discussion of the relationships between archival description and digital item-level descriptive metadata may have also resulted from an economical approach advocated in the "More Product, Less Process" (MPLP) movement to revamp traditional archival processing. In their initial 2005 article, the authors of MPLP conceded that digitization may generate "more detailed description—the closer to item level the better" in efforts to locate and digitize discrete items, but the collective descriptive rules should not be changed if the goal of digitization is to "identify whole collections, or whole series, that might warrant digitization as Web-accessible research material."[10] Greene expanded the idea in a second article dismissing the false assumption that "digitization must—or even should—be focused on individual items" and advocating the application of MPLP to digitization.[11] At the same time, Erway and Schaffner advised the special collections community to embrace "the standards and practices of managing collections and hierarchies, not necessarily items" to scale up digitization of special collections.[12]

The theoretical and empirical implications of recordkeeping metadata and MPLP-based digitization emphasized the contextual and collective management of electronic records and digital collections. Very few studies have been developed to investigate the necessity, possibility, and level of integration between archival description (context) and item-level descriptive metadata (content) of digital objects. Researchers and practitioners note the importance of the topic and discuss it briefly when they report their research on related topics such as implications and impact of EAD,[13] application of shareable metadata principles to archival description,[14] and interoperability (crosswalk) between archival (EAD) and bibliographic (MODS) metadata.[15]

As a key architect for EAD, Steven Hensen emphatically stated its value as "a fully-realized resource discovery mechanism for archival materials" in the

hyperlinked environment on the Internet that would "move seamlessly" from the very general level of collection descriptions to the very specific level of item-level descriptions.[16] He also recognized the emergence of "the potentially anomalous relationship" between traditional collection-level archival description and "the increasingly dense item-level descriptions found in deeply encoded EAD instances," especially "where that instance serves as a platform for the presentation of digitized collection materials."[17] More importantly, he noted that divergence has been developed in digital archival practice:

> Institutions that are using EAD as a platform for digital collection presentation . . . have concluded that traditional folder or item captions from finding aids are usually insufficient for searching, description, and identification of the digital items. . . . The result is that the descriptive metadata for digitized collection material has evolved into elaborate database records that fully describe the contents of each item and greatly facilitate complex searching both within and across collections.[18]

Hensen distinguished this new indexing practice from traditional "excesses of the item-level manuscript cataloging of years past" because of some important differences. First, digital objects require more descriptive metadata than simple folder or item caption description to make them searchable on the Internet. Second, the digitization process facilitates automatic, routine, or systematic capture of digital object metadata. He also found the current practice more justifiable for its focus on access than the traditional item-level cataloging of the past that "drew heavily on some of the traditions of rare book cataloging and had its focus almost entirely upon the physical and artifactual characteristics of the item." He called for more explicit metadata standards for describing digital objects, but made no further comment on how they would work along with archival descriptive systems or EAD instances that he discussed earlier.[19]

In their 2009 article, Jenn Riley and Kelcy Shepherd made a strong argument for archivists to create shareable metadata records that are openly available for use and reuse if archives are "to remain viable in a continually evolving information environment."[20] Among the challenges archivists need to address in making archival data more shareable, context and content are the most significant. While sharing context requires a reasonable balance in shared records of sufficient context data and full multilevel description, sharing content once again brings archivists' attention to item-level descriptions:

> Content is the second of the features of the shareable metadata framework that are most relevant to archives. Selecting the appropriate level of granularity for shared records is the greatest significant challenge facing archives that are making their descriptive metadata openly available for reuse. Most discussion and applications of shareable metadata in the cultural heritage sector have occurred in the digital library rather than in the archival arena. The most

notable evidence of the library-style approach is a general assumption that resources to be shared have item-level descriptions, an assumption that does not necessarily hold true in the archival world, where rich collection-level and multilevel descriptions are available more often than item-level descriptions.[21]

Riley and Shepherd provided some insights about the incompatibility between context represented in the encoded archival description and content represented in descriptive metadata. They differentiate EAD, which is designed to capture and represent archival description as a single hierarchical document, from a metadata structure standard as demonstrated in Dublin Core or MODS.[22] The document-centric approach in EAD thus contrasts with the data-centric approach in other descriptive metadata in a significant way that affects interoperability between the two:

> EAD contains features of markup languages, such as mixed content model, formatting information, and structural information such as lists and paragraphs, that are not found in other descriptive metadata structure standards. An EAD-encoded finding aid is therefore both metadata about an archival collection and its contents, and a document in and of itself. The finding aid is not just a simple inventory—it is a full narrative, not easily or losslessly converted to a form usable by record-centric systems.[23]

This, in turn, would have some impact on possible strategies that exist for archives to expose shareable metadata for use in new and unanticipated environments. The options listed in Riley and Shepherd include collection-level descriptive metadata traditionally incorporated into bibliographic databases, item-level descriptive metadata that can be made possible for a select few archives, construction of metadata aggregators that understand multilevel description to make full use of EAD-encoded hierarchical finding aids, or creation of EAD files that could contain file- or series-level data, with links out to an external system providing more granular description, perhaps in a metadata structure standard such as MODS or Dublin Core.[24] All this indicates that archivists have a long way to go before reaching an integrated solution for generating interoperable collection-level description and item-level metadata.

In their 2009 study, Lina Bountouri and Manolis Gergatsoulis pushed the inquiry further to investigate the semantic relationships between EAD and MODS, and proposed a crosswalk between the two.[25] They recognized that EAD and MODS are "two of the most widely implemented metadata schemas for the description of (digital) material in the field of cultural heritage." However, the schemas embrace two representation traditions that "have quite different documentation logic."[26] EAD inherits the multilevel representation tradition of archival description, while MODS is "a bibliographic element set that may be used for a variety of purposes, and particularly for (digital) library applications."[27] The former encompasses the exact representation of the archival materials as

a whole and an in-depth indication of their arrangement and complexity, while the latter is less complex and gives a horizontal view of the object described.[28]

> As a result, EAD and bibliographic metadata can supplement but not replace each other. Their coexistence is adequate in various cases. One of the most common cases is the use of EAD to represent the hierarchical relationships (i.e., parent-child relationships), to provide the users with the ability to identify the content and context of collections and to navigate complex hierarchical collections using structural links. On the other hand, bibliographic metadata can be used to present full analytic descriptions of specific (digital) objects, since they provide a satisfactory number of metadata fields for the documentation of various types of material.[29]

Rather than investigating possibilities for supplementing or replacing EAD and bibliographic metadata, we took an integrated approach in this study. We examined the efforts archivists have made to bring archival description and digital object metadata together and the challenges they face in representing digital content in an archival context. The design of the study recognizes the significance of granularity in digital content representation as well as the complexity archival context brings into the digital representation world.

Research Design

We used a mixed-methods research design for this study. In May of 2011, we began to collect data using a combination of systematic random sampling and purposive sampling to identify examples of digital archives and special collections located in North America. Efforts were made to include sources with a broad coverage of digital collections containing archival and special collection materials. We selected three sources, namely, the *Scout Report* online archives, the Society of American Archivists (SAA) *Repository of Primary Sources* list maintained by Terry Abraham, and the Internet search engine Bing, where we conducted a keyword search.

The *Scout Report*[30] is the flagship publication of the Internet Scout Project, a part of the National Science Foundation's National Science Digital Library (NSDL) project. Published every Friday, it provides a fast, convenient way to inform users of valuable resources on the Internet. A team of professional librarians and subject matter experts selects, researches, and annotates each resource. Published continuously since 1994, the *Scout Report* claims to be one of the oldest and most respected publications about valuable resources on the Internet, including digital library and archival collections made available online. The SAA *Repository of Primary Sources* list[31] is an online directory of over five thousand websites describing physical holdings of manuscripts, archives, and other primary sources worldwide, with the majority of the sites from North America.

The list is organized into geographical region sublists; we consulted the western United States and Canada, and eastern United States and Canada sublists. After exhausting the sublists pertinent to North America, we moved on to our final phase of data collection using the Bing search engine.

We used purposive sampling when examining the *Scout Report* because of the arbitrary occurrence of relevant digital collection examples throughout the source. We studied newsletters published in five recent years (2007 to 2011) and identified examples of digital archives by examining article titles that included terms such as: *digital archives, digital repository, digital collection(s), electronic records archives, electronic records, online archives, online records, online repository*, and *online collection*.

We again used purposive sampling when conducting the keyword search through Bing. We searched for the terms *digital archives* and *digital collections*; we also examined every other relevant link on the results pages.

We gathered examples of digital collections from the SAA *Repository of Primary Sources* list using systematic random sampling methods. Pulling examples only from the western United States and Canada sublist and the eastern United States and Canada sublist, we examined every third list item. Organizations with broken links, with links directing to outdated pages, or with websites in French we did not investigate further.

The identification of digital collections sites serves two purposes in this study. The analysis of general technical background of the selected digital collection sites, including software used and metadata schemas adopted, helps to identify the technical environments in which archival records, reformatted by means of digitization, are described and made available for use on the Web. The selected digital collection sites also provide a pool from which we selected digital archival collections for further analysis of the typological descriptions of digital archival representation. Implied in this approach is the understanding that the archival community works on a larger platform in the digitization movement and may share technical infrastructures with other digital collection communities.

General Trend of Digital Collection Organization and Representation

What is the general trend in organizing materials and making them available for use online when primary resources in special and archival collections have been digitized? As described in the previous section, the study started with an extensive search for digital collections posted online by archival and special collection communities. The goal of the search was to analyze a representative number of archival and special digital collection sites from a relatively broad variety of sources. After data cleanup and removal of duplicates, we compiled a

total of 276 examples of digital collection sites. Digital collections identified in this study are primarily affiliated with one organization or institution. A few of them are designated as consortium-based or collaborative collections, created by more than one organization or institution. The sample breakdown by organization type, as shown in Table 1, generates four subgroups: 195 (70%) from educational institutions, 25 (9%) from public libraries, 21 (8%) from government organizations, and 35 (13%) from other types of organizations. Although college and university libraries, special collections, and archives dominate the collection source list, we collected a reasonable number of digital collections constructed by public libraries, state and municipal libraries and archives, and other organizations such as museums, historical societies, and private archives.

Table 1. Number and Percentage of Sample Digital Collections by Organization Type (Total = 276)

Organization Type	Count	Percentage of Total Digital Collections Sampled
Educational	195	71%
Public libraries	25	9%
Government	21	8%
Other	35	13%

How digital collections are organized and represented is closely associated with their technical environments, that is, what digital collection management systems are used and what metadata schemas are adopted. Many, but not all, digital collection sites provide technical background information. Of the 276 digital collections selected for this study, 209 (73%) have identifiable content management systems. However, there are 67 sites (27%) whose systems cannot be identified due to the lack of information on the sites. As shown in Table 2, among the 209 identifiable systems, the distribution of systems shows a dominant leading system (CONTENTdm for 145 sites), a distant second system shared by a few collections (Fedora for 6 sites), and a long tail of other systems for 58 sites with fewer identical occurrences, thus categorized as "other" in the table. The "other" category accounts for a variety of systems and platforms, including digital collection management software, Web-based applications, viewers, manually coded websites, or any other system for managing digital collections that is not explicitly labeled or described as software by its creator(s).[32]

Table 2. Distribution of Identifiable Digital Collection Management Systems (Total = 209)

Collection Management System	Count	% of Total Identified Systems
CONTENTdm	145	69%
Fedora	6	3%
Other	58	28%

Due to the dominance of CONTENTdm as a digital collection management system, Dublin Core understandably remains on the top of the list of metadata schemas adopted to organize and represent digital collections with identifiable systems in this study. As shown in Table 3, the total number of sites using Dublin Core (153) is actually higher than the total number of sites using CONTENTdm (145), which indicates that digital collections from other digital collection management systems may also use Dublin Core. The two runner-up metadata schemas identified from the sampled collections with identifiable systems are METS (8) and MODS (4). Forty-four digital collection sites are categorized as "other," including locally developed metadata standards and/or undetermined metadata schemas.

Table 3. Distribution of Metadata Schemas Adopted in Identifiable Digital Collection Management Systems (Total = 209)

Metadata Standard/Schema	Count	% of Total Identified Systems
Dublin Core	153	73%
METS	8	4%
MODS	4	2%
Other	44	21%

The dominance of digital collection management software systems such as CONTENTdm and metadata schemas such as Dublin Core, METS, and MODS indicates that when primary resources in special and archival collections are digitized, the general trend is to organize and describe them individually rather than collectively. The movement to digitization generates digital objects with associated metadata. As a result, traditional archival materials are supplied with item-level description and metadata when turned into digital objects, a new element in digital archival representation that archivists are challenged to incorporate into the traditional archival description.

Typology of Digital Archival Representation

What would a typical description model of a digital archival collection be like in this metadata-centric digital collection description environment? What efforts are being made by archivists to ensure that digital items in archival collections are made accessible to users in their archival context? To address these questions, we selected examples of digital archival collections from the digital collection site pool. We define *digital archival collections* in this study as provenance-based institutional records or personal papers digitized and made available on the Internet with a conscious effort to present both archival context and digital item content. Reviewing each of the 276 digital collection sites, we manually selected a representative example of digital archival collections from each site. Only collections with a built-in connection to established archival context were selected, that is, those collections providing explicit linkages to archival finding aids. Digital collections constructed out of archival context were beyond the scope of this study.

From the 276 digital collection sites, 27 digital archival collections met the aforementioned selection criteria. Typically, we took one sample collection from a site, and if we found multiple archival collections on one site, we selected a representative collection. As the total number of the sites (276) is much larger than the selected number of the archival collections (27), we can assume that the majority of digitized collections identified in this study are not archival collection-based materials, or are not organized in accordance to archival principles. The data analysis described in the previous section supports this statement. Seventy-three percent of the total sites have identifiable digital collection management software systems. Of those sites with identified systems, close to 70% use CONTENTdm. In terms of metadata schemas, Dublin Core is a dominant choice. A typical digital collection description found in this study using the Dublin Core metadata schema, especially in CONTENTdm, starts with an overview of a collection, followed by a multifield search/browse/display, typically by title, author, subject, and description.

The 27 digital collections selected from the 276 digital collection sites cover a variety of archival and special collection institutions, although most of them are affiliated with educational institutions. In spite of the relatively small and uniform data sources, the collections provide useful data to address the research questions. Our analysis of the 27 digital archival collections focused on the level of integration of description/metadata of digital items into archival descriptive systems. We carefully reviewed the online description of each collection, categorizing them to look for integration patterns. Three representation models emerged through the analysis, as described in the following three subsections.[33] The three models are presented to accommodate an imbalance of

collection numbers associated with each model. We used collective data analysis and description for the first two models and individual case analysis and description to explain the last model.

Embedded Model

Of the 27 digital archival collections, we categorized 9 as using an embedded model to represent digital content in archival context. A common characteristic of this representation model is that an archival finding aid serves as the main access interface while digital objects are embedded in the hierarchical structure at various levels as appropriate. This could be a full online archival finding aid, a partial digital collection description, or a brief inventory listing. In any event, archival context description expressed in the hierarchy of collection, series, file, and item serves as a main gateway for users to gain access to digital content. The design of this model inherits the traditional archival principle of describing archival materials from the general to the specific. The model guides users from archival context to digital content and helps them to achieve a fuller understanding of the meaning of archival records in the context of their creation.

The level and depth of representation in this model are fairly flexible. Digital content can be described as groups as well as individual items. If multiple items in a folder are digitized as one set of digital content, a summary title with or without date range may serve as the only descriptive information for all materials in the folder. Individual items within the folder can be arranged in a sequence to facilitate browsing, but no descriptive information is provided at that level. Folder-level description information can be textual, by year range, or even number range, depending on the aggregation feature of the materials. The information can be presented as free text in a linear fashion, or as tabbed text in a faceted format at the associated level.

Digital content can reach down to the item level in the embedded representation model. Similar to folder-level description, description of digital items can be linear or faceted. Depending on the level of sophistication in description, additional description, if available, may be displayed along with digital images. However, this is the exception rather than the norm. A typical embedded model, as observed in this study, represents digital content in an archival context of multilevel hierarchical structure with linear folder- and/or item-level description.

Search capacity for digital content is limited in this model. Except for a general keyword search, detailed search capacity is normally not available because the model is not built with searchable metadata to enable specific field search. The majority of the 9 digital cases in the embedded model are not associated

with identifiable digital collection management systems. In contrast, more than 75% of the total digital collection sites in this study, as described in the previous section, have identifiable software systems in which close to 70% use the CONTENTdm software with the Dublin Core metadata schema. This helps to explain why keyword searching and contextual browsing of archival description are two dominant access tools in the model rather than more in-depth content searching of digital items.

The embedded representation model is contextual, hierarchical, and mostly linear (nonfaceted). The effort made by the archival community to link digital content to archival description online greatly benefits users, especially those who would have never made their way to archives. However, the nature of traditional archival description is such that digital content has been hidden at the bottom rather than made obvious at the front. Embedding digital content in online archival description helps to shorten the distance between archival collections and archival users, but does not seem to offer a sufficient mechanism to build a more granular relationship between archival representation and digital content, which will become more crucial in a networked world increasingly dependent on linked data to achieve more effective search results.

Segregated Model

Most of the digital archival collections we identified in this study (15 out of 27) do not rely on the hierarchical structure of archival description to represent and display digital content of archival materials. In an effort to take advantage of searchable metadata assigned to individual items in the process of digitization, a new representation model has been constructed to facilitate more direct access to digitized archival items. In this model, description of digital objects, in the form of multifaceted metadata, serves as the main access point for users to search or browse the content of digital collections. Archival context, in the form of online archival finding aids, mostly exists as an external link to provide additional historical and documentary background for digital materials. Rather than burying it deep, as in the embedded model, digital content has been detached from archival context as a segregated entity in this model.

The segregated model of digital archival representation provides more flexibility to the representation and access of digital content. The model allows for more specific descriptive information about digital items, such as title, subject, description, author, creator, recipient, contributor, date, place, and material type. In some cases, name and subject indexes previously created for physical archival items have been successfully repurposed to represent digitized items. This way, digital objects have been associated with a variety of attributes that can be made searchable by field. In contrast to linear textual item-level

description in traditional archival practice, description by field provides more searchable metadata. As a result, users can browse or search digital items by various attributes.

The practice helps to bring standardization to the description of digital items. As shown in the description of the embedded model, the technical environments for digital collections represented in the embedded model are either locally based or not explicitly declared. The majority of them do not use widely accepted digital content management software or standard metadata schemas to manage and describe their digital collections. In contrast, the majority of the digital collections in the segregated model are mostly constructed using CONTENTdm, a popular digital collection management software system sponsored by OCLC for the storage, management, and delivery of their digital collections. Most of them use the standard metadata schema (Dublin Core) to describe their digital items. The standard practice increases accessibility of digital content in a digital collection, and, more importantly, enables interoperability across collections in the digital archival and library communities and beyond.

An obvious disadvantage of the segregated model is that archival description loses its position as a primary gateway for access to digital items. Digital content, empowered by newly supplied descriptive metadata, becomes the main interface for users to search for digital collections. Archival context, provided by means of an external link, has been de-emphasized into a secondary position for users to refer to if they need more information. In most cases, they are treated as two interrelated independent entities. If a portion of an archival collection has been digitized and supplied with additional metadata, digital objects are grouped together to form the digital portion of the corresponding archival collection. In other words, digital content (digital objects and associated metadata) and archival context (archival finding aids) remain separate and refer to each other when needed.

Parallel Model

Three digital archival collections identified in this study manage to use both online finding aids and searchable item-level metadata to represent and display digital objects. Implementation methods may vary among the three, but they seem to share one identical representation goal. Digital content in this model is present in two representation systems. Digitized items are directly linked to online finding aids and, at the same time, made searchable or browsable by various metadata specific fields—within a single collection or across multiple collections.

The parallel model can be implemented within one digital collection as shown in the Thomas E. Watson Papers Digital Collection, jointly digitized by the

Southern Historical Collection and the Carolina Digital Library and Archives.[34] In this implementation, the finding aid "is encoded in Encoded Archival Description (EAD) XML format," and descriptive metadata for item-level objects "is encoded in the Metadata for Object Description Schema (MODS) and the Text Encoding Initiative (TEI)." As a result, "researchers can browse facsimiles of the manuscript collection using the finding aid in much the same way they would when visiting their physical counterparts." At the same time, they can search and browse the collection using "a searchable, browsable index of letters from the Correspondence Series and images from the Pictures Series of the Thomas E. Watson Papers."[35]

The Washington State Digital Archives search interface offers multiple metadata-based search options.[36] Users can search digital items by first name and last name, and by keyword. The Digital Archives consists of collections called record series. The name search and keyword search can be conducted across all record series or within a selected records series. In addition to the name and keyword search, the Digital Archives offers a detailed search function. Users can select a record series, which will prompt a search interface based on how the series is indexed and what metadata are associated with the series. For example, users can search the Real Property Record Cards series by parcel number, house number, street name, and legal property description and the Census Records series by last, first, and middle names, by birth place, and by "year from" and "year to."

The key to the flexible search capacity at the Washington State Digital Archives is that its record series are made searchable based on available metadata. However, record series in the Washington State Digital Archives also serve as the focal point for archivists to provide contextual information for records in each collection. On the collection page that lists all of the record series at the archives, users can find a brief description for each series, expand each series title to browse the record sets contained within each record series, and finally click on each record set title for detailed information, including a search interface.[37] Records created to perform the same business function are arranged together to form record series. Records generated by the same provenance are arranged to form record sets. Each record thus has its fixed place in the collection that associates the record to its creating body, business purpose/function, and relationship with other records. Information about record series, record set, and record title can also be used to assist in a detailed search, providing more accurate results.[38] For its functionality to support contextual representation of digital content, the Washington State Digital Archives is understandably claimed to be "a first of its kind repository," created by "blending the latest technologies with traditional archival theory."[39]

The Online Archives of California (OAC) took conscious steps in its development over the past decade to experiment on how best to represent and display the digital content of archival collections. In 2002, the OAC interface underwent its first significant redesign "to integrate finding aids with associated digital content."[40] Seven years later, based on user feedback and community input, another redesign of the OAC interface was released. The redesign "grew out of several ideas."[41] One of them was the launch of Calisphere, "a free website that offers educators, students, and the public access to more than 200,000 primary sources."[42] Although the "content in Calisphere is drawn from the digital content in the Online Archive of California," Calisphere repurposes digital materials by reorganizing them in a way that "best serves general users and K–12 audiences." The redesign "freed the OAC to focus on being purely a site for researchers."[43] The idea of two sites serving two user communities is clearly articulated on the project website:

> These two websites exist because they serve two very different user needs. For research-oriented users who want to go beyond what is available online and locate the actual, physical item, the OAC is the best starting point. For users whose primary interest is to view digitized images and documents, Calisphere is a place to explore online content. In addition, Calisphere provides K–12 educators with a subset of content organized and aligned with California Content Standards.[44]

Although users go to Calisphere "for digitized primary sources," researchers using online finding aids and collection guides on the OAC also have access to selected digital items "for immediate online viewing."[45] The practice of using two representation systems as access tools for digital objects reveals the designers' efforts to provide contextual access as well as topical access to digital archival materials. However, similar to those of the Thomas E. Watson Papers Digital Collection and the Washington State Digital Archives, the two representation systems generate two separate access interfaces. Users would have to visit and conduct searches in two collection sites to fulfill their use needs.

The parallel model manages to take advantage of the two representation systems by presenting them simultaneously online. The model offers more comprehensive solutions by exploiting the representation and access capacities of the two systems. The strategy can help shorten the distance; however, it does not help much in improving the level of integration of archival context and digital content as long as the two systems continue to remain parallel rather than integrated.

Challenges of Digital Archival Representation

The archival community has gradually increased its digital presence by making more and more archival materials directly accessible on the Web. The data collected in this study show that archivists have made conscious efforts to build connections between archival description (context) and digital items (content), and, as a result, distinct representation models have emerged from digital archival practice as discussed in the previous section. However, at the level of integration of archival context and digital content in each model, archivists are challenged to achieve an ultimate goal of making digital archives more accessible and better contextualized in the networked digital world.

The embedded model closely ties digital content to its archival context. Digital objects, when embedded at their appropriate aggregate levels in archival finding aids, can be retrieved and interpreted contextually to minimize the loss of the archival meaning. Because of its close tie to archival tradition, archivists may feel most comfortable adopting this model. However, by embedding digital objects in traditional archival hierarchy with minimal discovery metadata, this model offers limited accessibility to its digital content. The lack of digital discoverability in the embedded model may cause serious concerns in the networked and linked-data world where users have become used to, and therefore expect, a higher level of granularity in search results, as a by-product of greater semantic information discovery on the Web.

The segregated model manages to make digital objects more accessible by highlighting discovery metadata. The model increases the accessibility of digital content, but decreases the centrality of archival context in digital archival representation. That digital content has to be segregated or "released" from archival hierarchical structure may make archivists feel uneasy about this model. The latter may likely be viewed as a restraint rather than as a facilitator for digital archival representation. Having lost its central role in archival information retrieval, an archival finding aid only exists as an external link to provide users with additional background information if needed, which could easily be bypassed, ignored, or simply unnoticed.

If archival context is highlighted in the embedded model, and digital content is highlighted in the segregated model, the parallel model represents efforts made by archivists to bring the two together by making digital objects both browsable in archival finding aids and searchable with item-level metadata. The parallel model aims to make digital content searchable as individual items, as well as contextualized in an archival descriptive framework. However, evidence shows that content and context may still have to remain apart as separate interfaces and, arguably, serve different purposes. How to bring them

together in an integrated interface seems essential to archivists' efforts to make digital content accessible online to meet all user needs.

Conclusions and Implications

Two representation systems are found in the three models identified in this study. Archival description provides contextual information and internal structure of digital collections; descriptive metadata provides item-level multifaceted representation of digital objects. The former is characterized by its hierarchical structure and linked to digital objects with limited descriptive metadata; the latter by its granular representation and discovery capacity. We concluded that currently the two systems only exist within separate systems or access interfaces. Are there any compelling reasons that prevent the more satisfactory integration of the two representation systems so that archival context and digital content can be better represented in digital archives?

The question may be addressed by looking into the nature of records before and after digitization, the enabling technologies of the two representation systems, and the uses of digital archival collections. When analog records are digitized, 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. The new relationships may be more than traditional archival description systems can contentedly handle. It is not clear if and how archival description systems can be remodeled to be capable of expressing more complicated relationships embedded in digital object metadata.

The documentary relationships among archival records are inherently hierarchical (e.g., from collection, to series, to file, to item) and traditionally captured in document-centric archival description systems. In contrast, multifaceted representations for digital objects are essentially categorical (e.g., title, author, date, subject) and better expressed by data-centric descriptive metadata systems. The two representation systems seem to rely on different content and encoding standards and are facilitated by different enabling technologies, for example, XML-based EAD in contrast to Dublin Core–supported CONTENTdm. To what extent can technologies and software systems affect the representation systems archivists decide to use? In other words, does the application of EAD and CONTENTdm contribute to the segregation rather than integration of two representation systems?

The findings of the study also reveal that digital archival materials represented in the two systems are provided with different forms of access and thus serve two different purposes. Archival description provides contextualized

access to archival records, which may better serve research purposes. Item-level metadata provides granular access to information in the archives, which may better meet general information inquiry/use needs. Further research is needed to test and justify some assumptions about these practices. For example, if archival materials are not all used for archival purposes, should archival representation/access systems be designed to facilitate both archival and nonarchival use of archival materials? In other words, should archival digital objects be represented archivally as well as bibliographically to meet different user needs?

In a rapidly changing digital environment, the archival profession faces challenges in locating appropriate representation tools to advance its mission of providing more accessible digital content while concurrently maintaining sufficient archival context. Several strategies have been developed, as shown in this study, to describe digitized archival collections and make digital objects accessible online. Challenges arise when two representation systems adopted to describe and provide access to digital archival materials are complementary but not compatible. The traditional minimal metadata approach that relies on archival context to retrieve archival items may lead to limited digital accessibility, but it is equally unacceptable when granular access to digital content may have to be achieved at the expense of archival context. Item-level metadata adds a new dimension of digital discoverability to digital archival records, and it is important that the archival profession continues to develop strategies to integrate this new dimension into digital archival representation to achieve its dual goal of making digital archives more accessible and better contextualized to serve the needs of the user communities in the twenty-first century.

Appendix A (All sites accessed in February and March 2012)

Embedded Model (9 Collections)

- Andover-Harvard Theological Library. Unitarian Service Committee. Chairman and President, William Emerson. Records, 1939–1957: A Finding Aid.
- University of Massachusetts Special Collections and Archives. Kenyon L. Butterfield Papers. Selected Records Related to Women's Education at Massachusetts Agricultural College, 1906–1924.
- University of Hawaii at Manoa Library. Archives and Manuscripts Department. Chaplain Higuchi Wartime Correspondence.
- Brown Archival and Manuscript Collections Online. Guide to the William James Linton Papers.
- Carnegie Mellon University Libraries. Heinz Electronic Library Interactive Online System (HELIOS).
- University of Baltimore Langsdale Library Special Collections. Collection of Robert Breck Chapman.
- Harvard Law School Library. Edmund M. Morgan Papers on the Drafting of the Uniform Code of Military Justice.
- The Jacob Rader Marcus Center of the American Jewish Archives. Issac Mayer Wise Digital Archive.
- Mount Holyoke College. Office of the President: David Bicknell Truman. Coeducation Subject Files, 1960–1978 (bulk 1969–1978).

Segregated Model (15 Collections)

- University of Maryland, Baltimore County (UMBC). Ferdinand C. Latrobe Papers (Maryland Digital Cultural Heritage).
- Columbia University Libraries. Rare Book and Manuscript Library. Lehman Special Correspondence Files.
- University of Mississippi Libraries. Archives and Special Collections. Digital Collections. James O. Eastland Collection.
- University of Vermont, UVM Libraries's Center for Digital Initiatives. Fletcher Family Papers.
- University of Delaware Library. Special Collections. George S. Messersmith Papers.
- Purdue University e-Archives. George Palmer Putnam Collection of Amelia Earhart Papers.
- Boise State University. Albertsons Library Digital Collections. Nell Shipman Collection.
- Southern Methodist University. Central University Libraries. Digital Collections. Horton Foote Photographs and Manuscripts.

- Boston College University Libraries Digital Collections. Bobbie Hanvey Photographic Archives.
- University of the Pacific. Holt-Atherton Special Collections. John Muir Papers.
- Auburn University Digital Library. Eugene B. Sledge Collection.
- Northern Kentucky University. Special Collections and Archives. Bruce Family Papers.
- American Antiquarian Society. Photographs of Seventeenth and Eighteenth Century Structures in Massachusetts taken 1887–1945 by Harriette Merrifield Forbes.
- St. Catherine University. Digital Collections. Ade Bethune Collection.
- University of Iowa Libraries. Iowa Digital Library. John P. Vander Maas Railroadiana Collection.

Parallel Model (3 collections)
- University of North Carolina Libraries. Digital Collections. The Thomas E. Watson Papers Digital Collection.
- Washington State Archives. Digital Archives.
- Online Archive of California (OAC)/Calisphere.

Notes

1 Luciana Duranti, "Concepts and Principles for the Management of Electronic Records, or Records Management Theory Is Archival Diplomatics," *Records Management Journal* 20, no. 1 (2010): 79.

2 Oliver W. Holmes, "Archival Arrangement—Five Different Operations at Five Different Levels," *The American Archivist* 27 (January 1964): 21–42.

3 *ISAD(G): General International Standard Archival Description* (Ottawa: International Council on Archives, 2000), 12.

4 Kent M. Haworth, "Archival Description: Content and Context in Search of Structure," *Journal of Internet Cataloging* 4, nos. 3–4 (2001): 22.

5 David Bearman, "Item Level Control and Electronic Recordkeeping," *Archives and Museum Informatics* 10, no. 3 (1996), http://www.archimuse.com/papers/nhprc/item-lvl.html.

6 Wendy Duff, "Will Metadata Replace Archival Description: A Commentary," *Archivaria* 39 (Spring 1995): 33–38; David Wallace, "Managing the Present: Metadata as Archival Description," *Archivaria* 39 (Spring 1995): 11–21; Heather MacNeil, "Metadata Strategies and Archival Description: Comparing Apples to Oranges," *Archivaria* 39 (Spring 1995): 22–32.

7 David Wallace, "Archiving Metadata Forum: Report from the Recordkeeping Metadata Working Meeting, June 2000," *Archival Science* 1, no. 3 (2001): 253–69.

8 Adrian Cunningham, "Digital Curation/Digital Archiving: A View from the National Archives of Australia," *The American Archivist* 71 (Fall/Winter 2008): 533.

9 Greg Bak, "Continuous Classification: Capturing Dynamic Relationships among Information Resources," *Archival Science* 12, no. 3 (September 2012): 287.

10 Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68 (Fall/Winter 2005): 248.

11 Mark A. Greene, "MPLP: It's Not Just for Processing Anymore," *The American Archivist* 73 (Spring/Summer 2010): 193.

12 Ricky Erway and Jennifer Schaffner, *Shifting Gears: Gearing Up to Get into the Flow* (OCLC, 2007), 6, http://www.oclc.org/research/publications/library/2007/2007-02.pdf.

13 Steven L. Hensen, "Archival Cataloging and the Internet: The Implications and Impact of EAD," *Journal of Internet Cataloging* 4, nos. 3–4 (2001): 75–95.

14 Jenn Riley and Kelcy Shepherd, "A Brave New World: Archivists and Shareable Descriptive Metadata," *The American Archivist* 72 (Spring/Summer 2009): 91–112.

15 Lina Bountouri and Manolis Gergatsoulis, "Interoperability between Archival and Bibliographic Metadata: An EAD to MODS Crosswalk," *Journal of Library Metadata* 9, no. 1 (2009): 98–133.

16 Hensen, "Archival Cataloging and the Internet," 75.

17 Hensen, "Archival Cataloging and the Internet," 89.

18 Hensen, "Archival Cataloging and the Internet," 89–90.

19 Hensen, "Archival Cataloging and the Internet," 90–91.

20 Riley and Shepherd, "A Brave New World," 91.

21 Riley and Shepherd, "A Brave New World," 98.

22 Riley and Shepherd, "A Brave New World," 98.

23 Riley and Shepherd, "A Brave New World," 98–99.

24 Riley and Shepherd, "A Brave New World," 100–101.

25 Bountouri and Gergatsoulis, "Interoperability between Archival and Bibliographic Metadata," 98.

26 Bountouri and Gergatsoulis, "Interoperability between Archival and Bibliographic Metadata," 99, 107.

27 Bountouri and Gergatsoulis, "Interoperability between Archival and Bibliographic Metadata," 105.

28 Bountouri and Gergatsoulis, "Interoperability between Archival and Bibliographic Metadata," 107.

29 Bountouri and Gergatsoulis, "Interoperability between Archival and Bibliographic Metadata," 107.

30 See Internet Scout, "Previous Issues of the Scout Report," http://scout.wisc.edu/Reports/ScoutReport/about.php.

[31] See *Repositories of Primary Sources*, http://www.uiweb.uidaho.edu/special-collections/Other.Repositories.html.

[32] A variety of open-source softwares or platforms are used in those sites to manage and display digital objects, including DSpace, Omeka, Django, MS Access, FilemakerPro, SobekCM, XTF platform, Luna Insight, Adobe ColdFusion Web Application, Adobe Acrobat Viewer, Acuman, CORAL, DLXS SGML/XML database, ARCHIVE-IT Web service, Greenstone, MS Silverlight, ExLiBris DigiTool, DALNET, CORSAIR, VIVA2, and INMAGIC DB/Text Web Publisher.

[33] See Appendix A for the list of the subsections of digital archival collection sites.

[34] University of North Carolina Libraries Digital Collections, The Thomas E. Watson Papers Digital Collection, http://www.lib.unc.edu/dc/watson/.

[35] Watson Papers Digital Collection, "About the Digital Collection," http://www.lib.unc.edu/dc/watson/digital_collection.php.

[36] Washington Secretary of State, Washington State Archives—Digital Archives, http://www.digitalarchives.wa.gov/Home.

[37] Washington State Archives—Digital Archives, "Available Collections," http://www.digitalarchives.wa.gov/Collections.

[38] Washington State Archives—Digital Archives, "Search," http://www.digitalarchives.wa.gov/Search.

[39] Washington State Archives—Digital Archives, "Background and History," http://www.digitalarchives.wa.gov/StaticContent/background.

[40] University of California, California Digital Library, "OAC History," http://www.cdlib.org/services/dsc/oac/history.html.

[41] California Digital Library, "OAC Redesign Project," http://www.cdlib.org/services/dsc/projects/oac_redesign.html.

[42] California Digital Library, "Calisphere," http://www.cdlib.org/services/dsc/calisphere/.

[43] California Digital Library, "OAC Redesign Project."

[44] California Digital Library, "Calisphere."

[45] California Digital Library, "About OAC," http://www.oac.cdlib.org/about/.

ABOUT THE AUTHORS

Jane Zhang is an assistant professor at the School of Library and Information Science, Catholic University of America. She holds a PhD in library and information science with an archival concentration from Simmons College and a joint master of archival studies and library and information studies from the University of British Columbia. Her teaching and research areas include records and recordkeeping, archival theory and practice, electronic records and digital archives, digital curation, and organization of information. Zhang worked at the Harvard University Archives and the University of Calgary Archives before joining the faculty at Catholic University.

Dayne Mauney is a research librarian at the Consumer Electronics Association in Arlington, Virginia. She graduated from the Catholic University of America in January 2013 with her MLS in digital libraries. Prior to graduation, she spent time interning in the Special Collections Department at Virginia Tech, and also in the Ralph Rinzler Archives at the Smithsonian Center for Folk Life and Cultural Heritage.