

This is a preprint of an article whose final and definitive form has been published in the Journal of Library Metadata © 2011 Taylor & Francis.

Journal of Library Metadata is available online at: www.tandfonline.com; the open URL of this article is <http://www.tandfonline.com/openurl?genre=article&issn=1937-5034&volume=11&issue=3-4&spage=104>

Contextual Metadata in Digital Aggregations: Application of Collection-Level Subject Metadata and its Role in User Interactions and Information Retrieval

OKSANA L. ZAVALINA¹

College of Information, University of North Texas, Denton, Texas, USA

ABSTRACT

A number of digital libraries that aggregate multiple digital collections are now generating subject metadata to describe intellectual content of entire digital collections as integrated wholes and to provide context for individual digital objects within them. However, the utility of this important contextual metadata has not been empirically evaluated. Exploratory study reported in this paper examined and compared collection-level subject metadata in three large-scale aggregations of cultural heritage digital collections in the United States and the European Union and analyzed the role of collection-level metadata in information retrieval in digital aggregations based on user search queries derived from transaction logs. A small-scale targeted user study, which combined interviews and observations of users interacting with an aggregation, was undertaken to complement evidence-based content analysis data. The study revealed considerable variability in two indicators: consistency of applying controlled-vocabulary collection-level subject metadata elements beyond topical and value length of metadata elements. Both free-text and controlled-vocabulary subject metadata were found vital in answering search queries of aggregation users. Users also expressed preference for viewing complete structured collection-level metadata records, which include subject metadata. Results of this study prove importance of provision of collection-level metadata in general and subject metadata in particular to enhance user experiences and information retrieval in digital libraries.

KEYWORDS: collection-level metadata; contextual metadata; subject metadata; descriptive metadata; metadata application; digital aggregations; digital libraries; information users; information search; subject access; content analysis (communication); interview; observation.

INTRODUCTION

Since 1990s, a number of aggregations that bring together hundreds of digital collections have been built in the United States and abroad: at the national level (e.g., American Memory), international level (e.g., The European Library, etc.), state level (e.g., Texas Heritage Online), or regional level (e.g., Mountain West Digital Library). Digital collections preserve a common context and order and have been considered a valuable information resource as a whole (Lourdi, Papatheodorou, & Doerr, 2009). Collections are commonly defined as aggregations of individual items (Johnston & Robinson, 2002) that are “assembled and maintained ... for a specific purpose

¹ Oksana L. Zavalina, Ph.D., is an Assistant Professor, Department of Library and Information Sciences, College of Information, University of North Texas, 1155 Union Circle #311068, Denton, TX 76203-5017 (Email: Oksana.Zavalina@unt.edu).

and audience and according to a particular collection development plan” (International Council of Museums/CIDOC, 2007, p. 35). In addition, digital collections are thematically cohesive by topic area, holding institution, or type of materials, searchable as a distinct collection, and have a unique point of entry such as URL (Cole & Shreeves, 2004). Collections coexist in multiple layers: some collections are sub-collections or super-collections of others (Lee, 2000).

Collections play an important role in facilitating access to information (Allen & Sutton, 1993) and serve as a filtering system that helps reduce users’ data overload and enables effective information discovery (Lourdi, Papatheodorou, & Doerr, 2009). Valuable functions provided by collections include collocating sources, narrowing the search scope to increase precision and ease of use, presenting choices, and assisting in information need clarification (Lee, 2003, 2005). Such functionality becomes especially important in a federated digital resource environment. However, no research has yet analyzed how collections are being described to facilitate access to information in digital libraries that aggregate digital collections.

Subject access has been one of the central topics of research in library and information science for decades, particularly in regard to information seeking and information retrieval (Hjørland, 1997). As defined by Cochrane (1985), subject access means systematic (e.g., classification system), topical (e.g., subject headings), and natural (e.g., title, abstract words) approaches to the subject matter and encompasses both processes of subject cataloging and retrieval. Subject access in information systems is provided through metadata — “structured data about an object that supports functions associated with the designated object” (Greenberg, 2005). To paraphrase Cochrane (1985), subject metadata (i.e., metadata describing subject matter of an information resource), that is created in the process of subject cataloging, is the crucial component in subject access. In the numerous studies of subject access in library catalogs and bibliographic databases (e.g., Jackson, 1958; Lipetz, 1970; Tagliacozzo & Kochen, 1970; Matthews, Lawrence, & Ferguson, 1983; Larson, 1991a, 1991b; Borgman, 1986, 1996; Cochrane, 1986, 2000) researchers found that users experience problems with subject access, which are often caused by how the subject metadata is applied.

Metadata that describes entire collections of information objects as opposed to single information objects has long been created and used in archival community. It has been recognized as being central to facilitating access to documents contained in archival collections (e.g., Bearman, 1992). Collection-level metadata, defined as a structured, standardized and machine-readable form of metadata providing a high-level description of collection of individual objects (Macgregor, 2003), is now often applied to describing digital collections in aggregations. Collection-level metadata facilitates cross-collection searching and collection discovery across multiple digital libraries and thus has a vital role to play in facilitating subject access in aggregations of digital collections. It provides an added level of descriptive granularity: important contextual information for individual information objects (Miller, 2000). The importance of expressing and sharing descriptions of collections in order to retain context when aggregating item-level metadata is emphasized by best practice recommendations for shareable metadata (e.g., Shreeves, Riley, & Milewicz, 2006). Linking collection-level and item-level metadata helps produce higher retrieval rates for item-level descriptions, re-contextualize orphaned items by including key access points that might be lacking in item-level metadata into collection-level metadata, and facilitate browsing behavior (Foulonneau, Cole, Habing, & Shreeves, 2005). Relational attributes specifying relations between a given digital collection and its various sub- and super-collections are essential in collection-level metadata, as these

attributes have a capacity to greatly improve the navigability (Geisler, Giersch, McArthur, & McClland, 2002).

As information resources and user communities served by these resources vary widely, no universal metadata scheme exists. However, Dublin Core metadata scheme is the most widely used for describing items in digital collections (e.g., Palmer, Zavalina, & Mustafoff, 2007). While some aggregators (e.g., OAIster or Europeana) harvest digital items from distributed collections without providing collection-level metadata, many others use collection-level metadata. However, with few exceptions (e.g., Opening History, National Science Digital Library), aggregations of digital collections generally use this metadata behind the scenes to support information retrieval but display to end-users only *Title* and free-text *Description* collection-level metadata elements.

To guide collection-level metadata creation, several collection metadata schemes were developed, including Research Support Libraries Programme (RSLP) scheme,² Encoded Archival Description (EAD),³ NISOZ39.91-200x,⁴ and the most widely used in aggregations Dublin Core Collections Application Profile (DCAP).⁵ Aggregations of digital collections often use DCAP as one of their collection metadata schemes or base their local collection metadata schemes on DCAP.

Collection-level metadata falls into two distinct kinds based on how metadata values are encoded: controlled-vocabulary metadata, which draws values from formally-maintained list of terms, and free-text metadata, which relies on natural language. While few studies (Zavalina, Palmer, Jackson, & Han, 2008a; 2008b) analyzed application of free-text collection-level subject metadata (namely, *Description* field) in aggregations of digital collections, no research to date has examined application of controlled-vocabulary collection-level subject metadata. Moreover, available studies focused on a single aggregation, while for the broader understanding of the use of collection-level subject metadata in various contexts, comparative analysis is preferable.

As demonstrated by previous research (e.g., Crystal & Greenberg, 2006; Drori, 2003; Wang & Soergel, 1998), users benefit from search result interfaces that display subject metadata and perceive the free-text abstract and subject keywords to be among the most useful metadata elements to judge the relevance of retrieved documents. However, these studies examined user interactions with item-level metadata describing individual information objects in IR systems and search engines. To the end of author's knowledge, user interactions with collection-level metadata and the value of collection-level subject metadata to the users have not been previously researched. This paper presents results of a multi-method exploratory study that is beginning to bridge the research gap identified above.

METHODS

Systematic, manual content analysis – both quantitative and qualitative – of collection-level metadata records in aggregations of digital collections was the major research method employed in this study. Three large-scale aggregations of cultural heritage digital collections in the United States and Europe were selected for analysis: American Memory aggregation⁶

² <http://www.ukoln.ac.uk/metadata/rslp/schema>.

³ <http://www.loc.gov/ead>.

⁴ <http://www.niso.org/workrooms/mi/Z39-91-DSFTU.pdf>.

⁵ <http://dublincore.org/groups/collections/collection-application-profile>.

⁶ <http://memory.loc.gov>.

developed by the Library of Congress, Opening History aggregation⁷ developed by the University of Illinois at Urbana-Champaign, and The European Library aggregation⁸ of digital collections created by the national libraries in the European Union. Among these three aggregations, only the Opening History displays its entire collection-level metadata records to the end user. American Memory and The European Library aggregations keep most of their collection-level metadata (except for the *Title* and free-text *Description* elements) behind the scenes to support search and faceted browse functions. XML files with complete collection records were obtained from developers of The European Library and American Memory.

The systematic sample of collection-level metadata records in three aggregations was analyzed: 39 records from American Memory, 33 records from Opening History, and 27 records from The European Library. The resulting 99 collection-level metadata records and – whenever needed – additional documentation were closely examined to determine:

- What collection-level metadata elements are intended in an aggregation for providing subject information about digital collections?
- How consistently are collection-level subject metadata elements applied within and across aggregations?
- How long are the metadata values in collection-level subject metadata elements?

This study also analyzed a systematic sample (25%) of all collection-level user search queries initiated in the Opening History aggregation between February 2008 and January 2009. A total of 501 unique collection-level search queries were replicated in Opening History and search retrieval results were analyzed with the purpose to answer the following questions:

- What collection-level metadata elements provide matches to user keyword queries?
- How often is the user search query satisfied only through the free-text collection metadata, or only through any of the controlled-vocabulary subject metadata elements?
 - In what proportion of searches?
 - In what proportion of retrieved collection metadata records?

The technique that had been used by Gross and Taylor's (2005) for analysis of the item-level metadata role in information retrieval in online catalogs was adopted to be used in this study to analyze the role of collection-level metadata.

Observations and interviews of a small targeted sample of cultural heritage aggregations' end-users were conducted in order to answer the following additional research questions:

- What is the value of collection-level subject metadata to the users?
- How does availability and display of collection-level subject metadata affect subject access?
- Which collection-level metadata elements are important for cultural heritage aggregation users?

The interview and observation data was collected and analyzed mainly to complement the analysis of collection-level metadata and search results. The sample for interview and observation component of the study was small but the direct match of the participants' research and teaching interests with the content of the cultural heritage aggregations studied and their knowledge of aggregations like American Memory ensured richness of collected data. This rich data allowed to triangulate results and to fill out the real-life picture of the application and use of collection-level metadata.

⁷ <http://imlsdcc.grainger.uiuc.edu/history>.

⁸ <http://www.theeuropeanlibrary.org>.

Sampling of the end-users was based on the subject strengths of American Memory and Opening History aggregations, which focus on various aspects of the United States history. Department of History faculty and Ph.D. Candidates at the University of Illinois at Urbana-Champaign were targeted. Six active United States history researchers with areas of research that closely correlated with subject strengths of American Memory and Opening History were invited to participate in the study. Three of them participated in interview and observation sessions in February-April 2010 (response rate of 50%). Two participants (a full Professor and a Ph.D. Candidate) studied Native American history, while another participant (a Ph.D. Candidate) studied Japanese American history during the Second World War.

Participants of the study were observed while exploring two aggregations – Opening History that displays entire collection metadata records to its end-users and American Memory that only displays *Title* and *Description* collection metadata elements – for content relating to a topic of their research and/or teaching. Respondents were also asked to compare their experiences in these two different environments, with data collected through a semi-structured interview protocol. The interview and observation session duration ranged from 45 to 60 minutes. Each interview and observation session was audiotaped and fully transcribed. User interactions with Opening History and American Memory aggregations were recorded with the help of *Camtasia Studio Screen Recorder and Presentation Software*.

FINDINGS

Collection-Level Subject Metadata

The collection metadata elements intended for providing subject information about digital collections, including genre and object type information, in three aggregations are listed in Table 1. This table also demonstrates how mapping of these collection metadata elements (which are often named differently in three aggregations) for comparative analysis was achieved. The “common-denominator” element names in the far-left column — *Description*, *Subjects*, *Objects*, *Temporal Coverage*, and *Geographic Coverage* — are used to report results of this study.

“Common-denominator” elements	DCAP	American Memory (MODS)	Opening History (DCAP-based DCC)	European Library (DCAP-based ELAPCD)
<i>Description</i>	<dcterms:abstract>	<abstract>	<description>	<description>
<i>Subjects</i>	<dc:subject>	<subject><topic> <subject><name> <subject><occupation>	<subject>	<subject>
<i>Temporal Coverage</i>	<dcterms:temporal>	<subject><temporal>	<coverage>	< temporal>
<i>Geographic Coverage</i>	<dcterms:spatial>	<subject><geographic>	<coverage>	< spatial>
<i>Objects</i>	<cld:itemType>	<typeOfResource> <genre>	<type>	<theme>

TABLE 1. Collection-level subject metadata elements and mapping

While the American Memory uses several metadata schemes for describing its digital collections “behind-the-scenes,” only the Metadata Object Description Schema (MODS) collection metadata⁹ set was obtained for this analysis. In the MODS collection metadata set, *Subject* metadata element is further subdivided into several facets. For comparative analysis of collection metadata in three aggregations *topic*, *name*, and *occupation* facets were mapped to

⁹ <http://www.loc.gov/standards/mods/v3/mods-collection-description.html>.

Subjects “common denominator” metadata field, while *geographic* and *temporal* facets were mapped to “common-denominator” *Geographic Coverage*, and *Temporal Coverage* respectively (Table 1). *Type of Resource* collection-level metadata element is used in the American Memory to encode types and genres of objects in a digital collection; in a small proportion of collection metadata records, *Genre* metadata element was also included. Both *Type of Resource* and *Genre* were mapped to “common-denominator” *Objects* field (Table 1). American Memory uses Library of Congress Subject Headings (LCSH) for topical subject headings, Library of Congress Name Authority Files (LC NAF) for personal, corporate, and geographic names, and Index Terms for Occupations in Archival and Manuscript Collections (ITOAMC) for *occupation* subjects. It is unclear if TGM II: Genre and Physical Characteristic Terms¹⁰ controlled vocabulary is used for *Type of Resource* and *Genre* elements.

Opening History uses four subject metadata elements in its Digital Collections and Content (DCC) collection metadata scheme that is based on Research Support Libraries Programme (RSLP) Collection Description scheme¹¹ and Dublin Core Collections Application Profile (DCAP): *Subject*, *Coverage*, *Type*, and free-text *Description*. *Subject* and *Description* elements are required in Opening History metadata scheme; the use of *Type* field to encode types and genres of objects in a digital collection and *Coverage* subject metadata element to encode temporal and geographic coverage information is optional. For comparative analysis of collection metadata in three aggregations, DCC *Subject* metadata element was mapped to the “common-denominator” *Subjects* element, DCC *Coverage* element was mapped to two “common-denominator” elements (*Geographic Coverage* and *Temporal Coverage*), and DCC *Type* was mapped to *Objects* (Table 1). With exception of *Description* element, subject metadata elements in DCC collection-level metadata scheme are controlled-vocabulary elements, which draw values from formally maintained lists of terms. Terms from more than one controlled vocabulary can be used as values in *Subjects* collection-level metadata element (Figure 1). The use of at least one Gateway to Educational Materials (GEM)¹² subject heading in the *Subjects* is mandatory in Opening History. Although not formally required, Library of Congress Subject Headings (LCSH) terms are often used in *Subjects*. Controlled-vocabulary terms from Art and Architecture Thesaurus¹³ and Library of Congress Thesaurus for Graphic Materials (TGM I: Subject Terms¹⁴ and TGM II: Genre and Physical Characteristic Terms) are also allowed for use in *Subjects*. The use of terms from Getty’s Thesaurus for Geographic Names (TGN)¹⁵ is strongly recommended for values in *Geographic Coverage* element. For describing *Temporal Coverage*, the online Collection Registry entry form (Figure 1) provides a checklist of date ranges suggested by Opening History developers (e.g., “1850-1899,” “1930-1949”), as well as a free-text field which allows expressing temporal coverage differently and often more specifically (e.g., “Civil War,” “1945”). Similarly, the required *Objects* element allows using both controlled-vocabulary values suggested by Opening History developers (e.g., “Photographs/Slides/Negatives,” “Oral histories (audio files)”) and alternative free-text values (e.g., “diaries,” “letters”), while the use of the TGM II terms (e.g., “aerial views,” “bank notes”) is encouraged.

¹⁰ <http://www.loc.gov/rr/print/tgm2>.

¹¹ <http://www.ukoln.ac.uk/metadata/rslp/schema>.

¹² <http://www.thegateway.org/about/documentation/gem-controlled-vocabularies/vocabulary-subject>.

¹³ <http://www.getty.edu/research/tools/vocabularies/aat/index.html>.

¹⁴ <http://www.loc.gov/rr/print/tgm1>.

¹⁵ http://www.getty.edu/research/conducting_research/vocabularies/tgn.

GEM Subject Headings

Please assign at least one Gateway to Educational Material (GEM) subject heading to your digital collection. On the next page you will be able to add keywords or additional subject or controlled vocabulary to describe your digital collection.

The top level GEM subject headings are listed here. Each of these has several sub-headings. You can add as many GEM subject headings as you like.

Arts Educational Technology Foreign Languages Health Language Arts Mathematics	Philosophy Physical Education Religion Science Social Studies Vocational Education
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

The browse functions of the IMLS and Opening History Collection Registries will be based on the GEM subject headings. The IMLS DCC project chose GEM subject headings because broad and not overly complicated. We also felt that the GEM subject headings could be used by a wide variety of communities. To find out more information about GEM, see <http://www.thegateway.org/>.

If you need to end the session before finishing your work, click the "Save File and Come Back Later" button at the [bottom](#) of the screen.

If all information below is correct, you can go to the next screen by clicking the "Describe Topical Content of Digital Collections" button at the [bottom](#) or top of the screen.

Arts (Main Subject Heading)

Architecture

Art therapy

Careers in art

Philosophy (Main Subject Heading)

Aesthetics

Careers in philosophy

Educational Philosophy

[List of GEM Subject Headings is truncated]

Topic of digital collection:
(Please indicate controlled vocabulary, if any)

Topical phrase:

Controlled vocabulary:

Geographic coverage of digital collection:
(Please check all that apply)

<input type="checkbox"/> World	<input type="checkbox"/> Oceania (continent)	<input type="checkbox"/> Middle Atlantic U.S. (general region)
<input type="checkbox"/> Africa (continent)	<input type="checkbox"/> Pacific Islands (general region)	<input type="checkbox"/> Midwest U.S. (general region)
<input type="checkbox"/> Antarctica (continent)	<input type="checkbox"/> South America (continent)	<input type="checkbox"/> Mountain Region U.S. (general region)
<input type="checkbox"/> Asia (continent)	<input type="checkbox"/> South Asian (general region)	<input type="checkbox"/> Northeastern U.S. (general region)
<input type="checkbox"/> West Indies (archipelago)	<input type="checkbox"/> Central America (general region)	<input type="checkbox"/> Pacific Coast U.S. (general region)
<input type="checkbox"/> Europe (continent)	<input type="checkbox"/> Canada (nation)	<input type="checkbox"/> Southern U.S. (general region)
<input type="checkbox"/> Middle East (general region)	<input type="checkbox"/> Mexico (nation)	<input type="checkbox"/> Southwest U.S. (general region)
<input type="checkbox"/> North and Central America (continent)	<input type="checkbox"/> United States (nation)	

(Please separate distinct locations using a semi-colon. Use the [Getty Thesaurus for Geographic Names](#), if possible.)

Time period associated with items in digital collection:
(Please check all that apply)

<input type="checkbox"/> Pre-1400	<input type="checkbox"/> 1400s-1699	<input type="checkbox"/> 1700-1799
<input type="checkbox"/> 1800-1849	<input type="checkbox"/> 1850-1899	<input type="checkbox"/> 1900-1929
<input type="checkbox"/> 1930-1949	<input type="checkbox"/> 1950-1969	<input type="checkbox"/> 1970-1999
<input type="checkbox"/> 2000 to present		

(Please list specific time periods here. Use years or spans of years (e.g. 1863-1865 for the Civil War), and separate time periods with semi-colons.)

***Object(s) represented in digital collection:**

<input type="checkbox"/> Books and pamphlets	<input type="checkbox"/> Photographs / slides / negatives	<input type="checkbox"/> Music (audio files)
<input type="checkbox"/> Newspapers	<input type="checkbox"/> Posters and broadsides	<input type="checkbox"/> Sheet music and scores
<input type="checkbox"/> Periodicals	<input type="checkbox"/> Prints and drawings	<input type="checkbox"/> Physical artifacts
<input type="checkbox"/> Interactive learning objects	<input type="checkbox"/> Oral histories (audio files)	<input type="checkbox"/> Physical specimens (plants / animals / etc)
<input type="checkbox"/> Moving images		

Other objects:

(Separate additional terms with semi-colons. If possible, use the [LC TDM 2.0](#))

FIGURE 1. Subject metadata in Opening History collection metadata entry form

The European Library relies on The European Library Application Profile for Collection Descriptions (ELAPCD)¹⁶ metadata scheme for describing its digital collections “behind-the-

¹⁶ http://www.theeuropeanlibrary.org/handbook/Metadata/tel_ap_cld.html.

scenes.” Similarly to collection metadata scheme used in Opening History, ELAPCD is based on the Research Support Libraries Programme (RSLP) Collection Descriptions scheme and has been updated to be largely compatible with the NISOZ39.91-200x Collection Description Specification and the Dublin Core Collections Application Profile (DCAP). The ELAPCD scheme includes a total of five collection-level subject metadata elements: *Description*, *Subject*, *Spatial Coverage*, *Temporal Coverage*, and *Theme* (an element that is used for encoding genre information). Only *Subject* and *Description* collection-level subject metadata elements are required in The European Library, while *Spatial Coverage*, *Temporal Coverage*, and *Theme* are optional elements. For comparative analysis of collection metadata in three aggregations, these ELAPCD collection-level subject metadata elements were mapped to “common-denominator” elements: *Description*, *Subjects*, *Geographic Coverage*, *Temporal Coverage*, and *Objects* (Table 1). Additional required metadata element — *Collection Item Type* — uses only two values (“digital” and “not digital”)¹⁷ to allow for distinguishing between digital collections and bibliographic databases aggregated by The European Library. Since all collection metadata records in this study’s sample described digital collections, the *Collection Item Type* metadata element was excluded from the analysis of subject metadata. Universal Decimal Classification (UDC) subject terms and numeric notations (e.g., “760 Graphic arts,” “940 General history of Europe”) are used in The European Library as values in *Subjects* elements. In comprehensive digital collections (e.g., Memory of Slovakia)¹⁸ with broad topical coverage the value “all subjects” is used in *Subjects* field instead. The European Library enforces the use of controlled vocabulary in *Subjects* and *Objects* collection metadata elements by providing dropdown menus with suggested values for these two elements in its Collection Description Editor entry form (Figure 2). There is no indication that any controlled vocabulary is required for use in *Geographic Coverage* and *Temporal Coverage* collection metadata elements.

¹⁷ The European Library aggregates both digital collections and catalogs/finding aids to physical collections.

¹⁸ <http://www.memoria.sk>.

The European Library
Collection Description Editor

>> Handbook >> Add new collection

Step 1. Submit new collection description

Add more elements by using the elements table below this one

Enter a value and then choose language it is in.
Add elements through the element list below. Delete items by pressing the bin of that row.

Element	Value	Language	
update_date: *	2008-11-25		
Title: *	<input type="text"/>	en	
Title:	<input type="text"/>	en	✖
Description: *	<input type="text"/>	en	
Description:	<input type="text"/>	en	✖
Publisher: *	<input type="text"/>		
isAccessedVia: *	SRU		
Language_ISO639-2: *	<input type="text"/>		
Spatial_coverage:	<input type="text"/>	en	✖
Temporal_coverage:	<input type="text"/>	en	✖
RecordSchema: *	<input type="text"/>		
Size: *	<input type="text"/>		
Subject: *	(make selection)		
Subject:	(make selection)		✖
Subject:	(make selection)		✖
Subject:	(make selection)		✖
Subject:	(make selection)		✖
URL:	<input type="text"/>	en	✖
ThumbnailURL:	<input type="text"/>		✖
CollectionType:	<input type="text"/>		✖
theme:	-- No theme --		✖
Collection ItemType: *	Not Digital		

Add more elements by using the elements table below

Proceed / Preview XML >>

FIGURE 2. Subject metadata in The European Library collection metadata entry form

In The European Library, two free-text collection metadata elements — *Title* and *Description* — are presented in 28 different European languages, while remaining “behind-the-scenes” collection metadata elements are currently represented only in English. However, the work is ongoing to translate entire metadata records in The European Library into 27 languages other than English. This added level of complexity in collection metadata generation calls for more concise values in collection metadata elements to simplify translation efforts.

Consistency of Application

As a required field in collection metadata schemes in all three aggregations, free-text *Description* was consistently applied in 100% of collection metadata records in the sample. Figure 3 shows the frequency with which controlled-vocabulary subject metadata elements were applied in collection records across the three aggregations. Opening History consistently used all four collection-level controlled-vocabulary subject metadata elements in 100% of its collection

records,¹⁹ while both American Memory and The European Library only consistently applied collection *Subjects* and/or *Objects* elements. The collection-level subject metadata element that was applied the least consistently in American Memory was *Temporal Coverage*, while The European Library paid the least attention to the application of the *Objects* element. This much higher consistency in application of collection-level subject metadata elements in Opening History can be explained by the fact that creation of collection metadata records in this aggregation is centralized. While the content for collection metadata in Opening History is drawn directly from documentation provided by the local developers of the individual collections (Zavalina, Palmer, Jackson, & Han, 2008a), collection records are created manually by the Opening History’s staff members who follow well-developed guidelines, and special attention is paid to the completeness of collection metadata records.

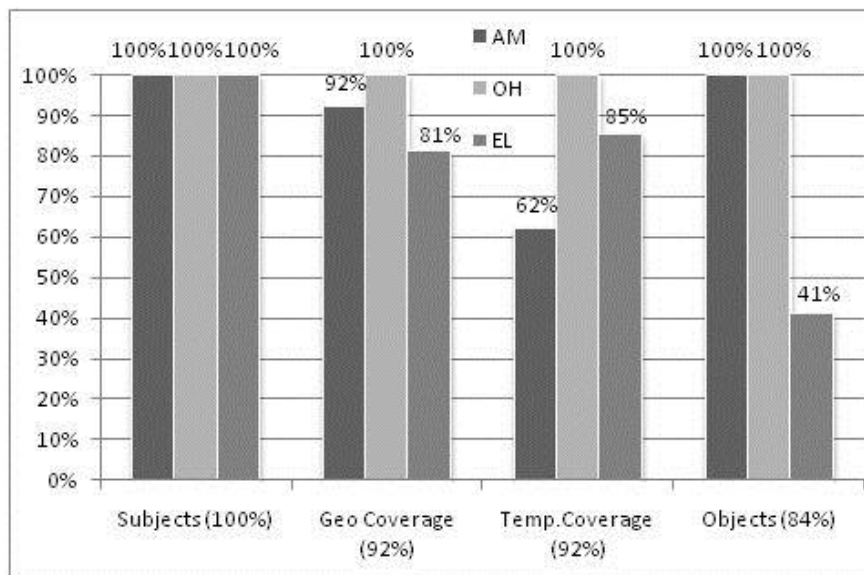


FIGURE 3. Application of controlled-vocabulary collection-level subject metadata elements

Value Lengths

Considerable variation in the length of values in free-text *Description* collection-level metadata element was observed both within each aggregation and across the three aggregations. As seen in Table 2, American Memory *Description* metadata values were the longest, with 132 words on average. Opening History *Description* metadata values contained 98 words on average, while The European Library *Description* metadata values contained on average 45 words. In Opening History, *Description* metadata values exhibited the widest range in length: from 5 to 429 words. American Memory *Description* metadata values ranged in length from 32 to 260 words, while The European Library *Description* metadata values varied the least among the three aggregations: from 13 to 114 words.

¹⁹ Only *Subjects* and *Objects* controlled-vocabulary subject metadata fields are required in Opening History collection metadata scheme. However, optional *Temporal Coverage* and *Geographic Coverage* fields are also used in 100% of the records.

Field	Aggregation	Mean	Median	Variance	Standard deviation
<i>Description</i>	American Memory	132	120	3105	55.73
	Opening History	98.2	83	4861	69.72
	The European Library	45.1	43	725.2	26.23
<i>Subjects</i>	American Memory	21.4	8	1533.1	39.15
	Opening History	19.94	11	459.5	21.44
	The European Library	8.81	7	52.62	7.25
<i>Temporal Coverage</i>	American Memory	10.05	3	448.52	21.18
	Opening History	6.88	6	18.61	4.31
	The European Library	6.17	13	60.06	7.75
<i>Geographic Coverage</i>	American Memory	3.29	2	5.06	2.25
	Opening History	13.82	15	38.15	6.18
	The European Library	1.41	1	0.54	0.73
<i>Objects</i>	American Memory	6.11	2	160.65	12.67
	Opening History	5.12	3	16.297	4.04
	The European Library	1	1	0	0
<i>Cumulative Length</i>	American Memory	171.75	144	8158.25	90.32
	Opening History	144	131	6846.44	82.74
	The European Library	61.15	50	799.52	28.28

TABLE 2. Lengths of values in collection-level subject metadata elements in three aggregations

The longest overall collection-level subject metadata values were observed in American Memory (Table 2). Both average and median indicators of the *Description* metadata value length were the highest in American Memory, with higher variability (variance and standard deviation) than in The European Library. Opening History exhibited the highest variability of *Description* metadata value length, with average and mean indicators higher than in The European Library and lower than in American Memory. While the mean and median lengths of the *Description* element values were the lowest in The European Library, the standard deviation was also the lowest among the three aggregations, which indicates higher consistency in the length of *Description* metadata values in this aggregation. In addition to *Description* elements, American Memory also exhibited the highest average length of values in *Subjects*, *Objects*, and *Temporal Coverage* elements (Table 2). However, the longest average *Geographic Coverage* element values were observed in Opening History. The values in two metadata elements — *Subjects* and *Objects* — had the highest median length in Opening History, although American Memory had the highest average length of values for these elements. Similarly, The European Library had the highest median length of values in the *Temporal Coverage*, although this element had the longest average value length in American Memory. As illustrated by Table 2, cumulative length of values in collection-level subject metadata elements (measured as sum of value length for all subject metadata elements per metadata record) was the highest in American Memory. Opening History had somewhat more concise cumulative collection-level subject metadata values, while The European Library exhibited the most concise cumulative collection metadata values overall.

Role of Collection-Level Subject Metadata in Information Retrieval in Aggregations

Collection-level subject metadata elements provided a match to a high proportion of collection-level user search queries documented by the Opening History transaction log, with the free-text *Description* playing the most important role. As seen in Table 3, in 93% of collection-level searches, at least one of the retrieved collection records had a search query match in *Description* element, while 74% of collection searches retrieved one or more collection records with a match exclusively in this element. Another collection-level metadata element that provided a significant source of matches to user search queries was *Subjects*. Half of the searches

retrieved at least one collection record with a search query match in *Subjects*; 27% of the searches retrieved one or more records with a match exclusively in *Subjects* field (unique match). In addition to *Description* and *Subjects*, two other subject metadata elements – *Objects* and *Geographic Coverage* – were among the elements that often contained search query matches (in 12% of collection searches each), including unique matches in 8% and 7% of collection searches respectively. Unlike other collection-level subject metadata elements, *Temporal Coverage* element contained search query matches in only 3% of collection searches and unique matches in only 2% of collection searches. In 10% or more of searches, search query matches were also contained in several non-subject collection metadata fields: *Title* (48% of searches), *URL* (19% of searches), *Copyright and IP Right* (15% of searches), and *Alternative Title* (11% of searches). However, occurrence of unique matches to user search queries in these metadata elements was quite low: 5%, 2%, 3%, and 2% respectively.

Collection metadata field	% of <u>collection searches</u> retrieving results with a match in a field	% of <u>collection searches</u> retrieving results with unique match (i.e., match only in this field)	% of <u>collection records</u> retrieved with a match in a field	% of <u>collection records</u> retrieved with unique match (i.e., match only in this field)
Description	93.0%	74.0%	36.4%	20.6%
Subjects	50.0%	27.0%	25.0%	10.6%
Title	48.4%	4.9%	6.4%	.3%
URL	18.5%	2.2%	6.0%	2.6%
Copyright and IP Rights	15.2%	2.7%	2.4%	.2%
Objects	12.0%	8.0%	9.2%	4.5%
Geographic Coverage	12.0%	7.0%	24.4%	12.5%
Alternative Title	11.4%	1.6%	1.5%	.2%
Notes	8.7%	3.8%	.8%	.4%
Alternative Access	6.5%	1.1%	6.3%	4.5%
Size	6.0%	1.1%	.4%	--
Contributing Institution	6.0%	--	.3%	--
Access Rights	4.9%	.5%	.3%	--
Temporal Coverage	3.0%	2.0%	14.6%	13.3%
Audience	2.2%	1.6%	.1%	--
Interaction	1.6%	1.1%	.3%	.2%
Supplementary Materials	1.6%	.5%	.2%	--
Hosting Institution	1.6%	--	.1%	--
Format	.5%	.5%	1.4%	.1%
Provenance	.5%	.5%	--	--
Collection Development Policy	.5%	.5%	.2%	--
Creator of Collection	.5%	--	--	--
Metadata Scheme Used	.5%	.5%	--	--

TABLE 3. Collection metadata elements that contained matches to user search queries

Only seven of thirty-three collection-level metadata elements had a sizeable impact on retrieval of digital collections in the search process (Table 3). Matches in *Description*, *Subjects*, and *Geographic Coverage* elements were observed in retrieved collection metadata records the most often: 36%, 25%, and 24% of collection records respectively. Almost 15% of retrieved collection metadata records had a match in *Temporal Coverage*. Moreover, 21% percent of

collection records contained the match exclusively in *Description* element and therefore would not be retrieved if *Description* element were omitted in the collection metadata records. Similarly, thirteen percent of collection metadata records would not be retrieved if *Temporal Coverage* element were omitted, 12% would not be retrieved if *Geographic Coverage* element were omitted, 11% would not be retrieved if *Subjects* element were omitted, and 5% would not be retrieved if *Objects* element were omitted. It is important to note that almost a half (41%) of digital collections would not be retrieved in collection searches if aggregators failed to apply controlled-vocabulary subject metadata elements. In addition to subject metadata elements, only two other collection-level metadata elements substantially impacted retrieval of digital collections: 8% of collection metadata records were retrieved solely due to the presence of unique resource identifiers in the records, either in *URL* collection metadata elements (3%) or in *Alternative Access* elements (5%).

User Interactions with Collection-Level Subject Metadata: Interview and Observation Findings

Observations and interviews conducted as part of this study demonstrated that collection-level subject metadata plays an important role in user interactions with aggregations.

For example, participants often engaged in collection-level subject browsing by topical or geographical facet – the kinds of browsing that are made possible through application of collection-level subject metadata. Two out of three historians participating in the observations started exploring aggregations with collection-level browse. One researcher started with the topic browse; another explored two different subject browse options before moving to search. After conducting searches, two participants browsed through the search results at collection level before selecting a collection record of interest. Yet another researcher explored subject browse option in the middle of the session, after performing item search and viewing collection and item metadata records. Overall, collection-level browse by topic was used at least once by each of the three participants during the observation sessions. Geographic browse option was selected by one observation participant, however two participants were also observed clicking on hyperlinked values in *Geographic Coverage* collection metadata field — an action which results in opening a geographic browse result window. Although all three respondents mentioned that knowing the types of materials in a collection is important for them, none used the object browse option during the observation session. However, while not knowingly initiating object browse, participants were also observed clicking on hyperlinked values in *Objects* collection metadata elements — an action which results in opening an object browse results window.

All of the observation participants engaged in subject searching (i.e., searching for entire digital collections or individual items on a specific subject), using either basic keyword search or advanced fielded search. Users reported formulating broader search queries for teaching-related searches (e.g., “Civil War”, “slavery”) and more specific search queries for research-related searches: relatively narrow topics (e.g., “Japanese American internment”), corporate names (e.g., “War Relocation Authority”), family names (e.g., “Laublin,” “Pueblo”), or tribal names (e.g., “Ogibwe” spelled a couple of different ways, “Paiutes,” “Illinois Indians,” “Crow Indians”).

Participants also reported that the search terms they use in aggregations or in search engines such as Google are not derived from any controlled vocabulary, however one of the participants commented that he is familiar with the Library of Congress Subject Headings in his area and often uses them as search terms in library online catalogs.

Interviews and observations also demonstrated that aggregation users value collection-level metadata. Each of the participants of observation study looked at the full versions of some of the collection-level metadata records that were retrieved as a result of their searches. All three participants reported that collection metadata records are very helpful in providing information important for their research and teaching. One of the respondents reported prioritizing collection metadata over item-level metadata, and commented, “I don’t necessarily need to see it [item metadata record]; generally by that time [of looking at individual item] I know how it’s been described as a collection. I know what I am looking for in general.” While looking at 244 results retrieved by his item-level search, this historian commented on the value of collection metadata in organizing the item results and providing context for the items: “if I am searching for something initially, this is too much information. I’d rather see it grouped by collection and have good metadata about the collection as a whole.”

Participants reported that several collection-level subject metadata elements are especially important in their information discovery in aggregations of digital collections. Three of the controlled-vocabulary collection-level subject metadata elements — *Subjects*, *Objects*, and *Geographic Coverage* — were named by all three participants, while *Temporal Coverage* was also named by one participant. One of the two Native American history researchers also pointed out that it would be helpful to include tribal names in the *Subjects* elements, wherever applicable. The free-text *Description* was named as important by all three participants, although one participant commented that in the case of the digital collection familiar to him, the *Description* was “a little more flowery than it needs to be for my purposes.”

Figures 4 and 5 include the snapshots of two different collection metadata record displays in Opening History and American Memory for the same digital collection — “Ansel Adams Photographs of Japanese American Internment at Manzanar.” Opening History displays to the end user all collection metadata elements in a record, while in American Memory only the free-text *Description* is displayed. When asked to compare the collection metadata displays in Opening History and American Memory, all three participants commented that the more structured approach taken by Opening History works better for them than the approach taken by American Memory. The more structured collection metadata records were preferred as more useful “because they do not require a lot of reading, while presenting information of interest (subject headings, object types, geographic coverage, etc.)” All participants commented that they are more interested in structured and hyperlinked collection metadata, which supports browsing, for their research (e.g., “the thing that I like about this is that allows you to go use these links to search for similar types of materials”), while just a long free-text *Description* might be more useful for their teaching (e.g., “if I was looking for something to teach with, this would be very helpful, because I can read through it and get more sense of the background of this collection”). It was noted by one participant that overall, Opening History is better organized for subject discovery (“you can find a lot more things”), while American Memory is more useful for known-item collection discovery (“if you know a particular collection”).

Title: Ansel Adams's Photographs of Japanese-American Internment at Manzanar

Collection Home: <http://memory.loc.gov/ammem/collections/anseladams/>

Items: Browse 244 Items in Collection

Description: In 1943, Ansel Adams (1902-1984), America's best-known photographer, documented the Manzanar War Relocation Center in California and the Japanese Americans interned there during World War II. In "Suffering under a Great Injustice", Ansel Adams's Photographs of Japanese-American Internment at Manzanar, the Prints and Photographs Division at the Library of Congress presents for the first time side-by-side digital scans of both Adams's 242 original negatives and his 209 photographic prints (with the print on the left and the negative on the right), allowing viewers to see his darkroom technique and in particular how he cropped his prints. Adams's Manzanar work is a departure from his signature style of landscape photography. Although a majority of the photographs are portraits, the images also include views of daily life, agricultural scenes, and sports and leisure activities. When he offered the collection to the Library in 1965, Adams wrote, "The purpose of my work was to show how these people, suffering under a great injustice, and loss of property, businesses and professions, had overcome the sense of defeat and despair [sic] by building for themselves a vital community in an arid (but magnificent) environment. All in all, I think this Manzanar Collection is an important historical document, and I trust it can be put to good use."

Library of Congress Subject Headings: World War, 1939-1945--Japanese Americans--California--Manzanar. Manzanar War Relocation Center--Facilities--1940-1950. Japanese Americans--Evacuation and relocation, 1942-1945.

Subjects: Asian Americans
World War II
American Culture
Adams, Ansel, 1902-, photographer.

GEM Subjects: Arts
Photography
Social Studies
United States history

Geographic Coverage: North and Central America (continent)
United States (nation)
Pacific Coast U.S. (general region)
California (state)

Time Period: 1930-1949

Objects Represented: Photographs / slides / negatives

Format: image/jpeg
image/tiff

Language: eng

Interaction with Collection: Search
Browse

Access Rights: There are no known restrictions on Ansel Adams's Manzanar photographs. Privacy and publicity rights may apply.

Copyright & IP rights: Access is permitted, subject to Prints & Photographs Division policy on serving originals which requires service of surrogate prints and copy negatives in lieu of the originals.

Size: 244

Frequency of additions: Irregularly

Alternative Access: http://memory.loc.gov/cgi-bin/oaai2_0?verb=ListRecords&set=manz&metadataPrefix=oaai_dc, http://memory.loc.gov/cgi-bin/oaai2_0?verb=ListRecords&set=manz&metadataPrefix=mods

Alternative Title: ListSets title: Records for "Suffering Under a Great Injustice" Ansel Adams's Photographs of Japanese-American Internment at Manzanar

Hosting Institution: Library of Congress

FIGURE 4. Collection metadata display example: Opening History

More search options

- Collection Home
- About This Collection

Features:

- Gallery
 - Collection Highlights
- Timeline
 - 1902-2007
- Essay
 - Born Free and Equal

Browse Collection by:

- Subject

View more collections from the [Prints and Photographs Division](#)

Collection Connection
Classroom resources for teachers

Ansel Adams's Photographs of Japanese-American Internment at Manzanar

Tom Kobayashi, Landscape, Manzanar Relocation Center, California / photograph by Ansel Adams.
[About this image](#)

Overview

In 1943, Ansel Adams (1902-1984), America's best-known photographer, documented the Manzanar War Relocation Center in California and the Japanese Americans interned there during World War II. In "Suffering under a Great Injustice": *Ansel Adams's Photographs of Japanese-American Internment at Manzanar*, the Prints and Photographs Division at the Library of Congress presents for the first time side-by-side digital scans of both Adams's 242 original negatives and his 209 photographic prints (with the print on the left and the negative on the right), allowing viewers to see his darkroom technique and in particular how he cropped his prints.

Adams's Manzanar work is a departure from his signature style of landscape photography. Although a majority of the photographs are portraits, the images also include views of daily life, agricultural scenes, and sports and leisure activities. When he offered the collection to the Library in 1965, Adams wrote, "The purpose of my work was to show how these people, suffering under a great injustice, and loss of property, businesses and professions, had overcome the sense of defeat and despair [sic] by building for themselves a vital community in an arid (but magnificent) environment. All in all, I think this Manzanar Collection is an important historical document, and I trust it can be put to good use."

[Gallery](#)
Collection Highlights

[Essay](#)
Born Free and Equal

FIGURE 5. Collection metadata display example: American Memory

DISCUSSION AND CONCLUSION

The report "Subject Data in the Metadata Record," from the ALCTS/CCS/SAC/Subcommittee (1999) recommends (5.1) inclusion of *Subject* element in item-level metadata records in digital libraries, to help achieve optimal recall and precision. Subcommittee also recommends (5.1.4) the use of other elements — namely *Type* and *Coverage* — in addition to the *Subject* to accommodate different subject-related facets such as topic, place, time period, language, etc.

This study found that American Memory, Opening History, and The European Library aggregations of digital collections follow these recommendations in their collection-level metadata. Five subject metadata elements were applied in collection-level metadata records in these three aggregations: free-text *Description* and controlled-vocabulary *Subjects*, *Objects*, *Geographic Coverage* and *Temporal Coverage*.

Consistency of application of collection-level subject metadata other than *Description* and *Subjects* varied across the three aggregations. Opening History was found to be the most consistent in applying *Temporal Coverage*, *Geographic Coverage*, and *Objects*, while the collection metadata records in the other two aggregations often lacked one or more of these metadata elements. Nevertheless, overall level of consistency of application was considerably higher for most of the collection-level subject metadata elements than in the recent studies that examined item-level subject metadata. For example, this study found that *Geographic Coverage* and *Temporal Coverage* elements were each applied in 92% of collection metadata records in the

sample, which is much higher than 7% rate of application of Dublin Core *Coverage* item-level metadata field observed by Kurtz (2010) in three university repositories, 21% of item-level metadata records in six digital video repositories in Weagley, Gelches, and Park (2010) study, or 51% of item-level metadata records harvested through OAI-PMH in Jackson, Han, Groetsch, Mustafoff, and Cole (2008) study. Similarly, *Subjects* metadata element was applied in 100% of collection-level metadata records in this study and in only 65% of item-level metadata records in Weagley et al. (2010) and Kurtz (2010) studies and 94% of item-level metadata records in Jackson et al. study. The *Objects* metadata element was found to be applied in collection-level metadata records in this study overall (84%) somewhat more often than in Weagley et al. (2010) study of item-level metadata (83%) but less often than in Jackson et al. (2010) study (91%) and Kurtz (2010) study of item-level metadata (100%).

Length of collection-level subject metadata values was found to vary significantly both within and across the three aggregations. The European Library had the highest consistency in the length of values in collection-level subject metadata elements. The highest cumulative length of values in collection-level subject metadata was observed in the American Memory aggregation. Although aggregation users expressed no specific expectations for collection-level metadata elements' value lengths, they indicated interest in both more detailed and more structured collection metadata records. It seems that as long as collection-level metadata records are structured, which allows for more time-efficient interaction with metadata, longer and therefore more detailed metadata values are preferable. However, it is important to note that, as found by Fraser and Gluck (1999) in their analysis of user interactions with item-level metadata, both extremely long metadata records and metadata element values can cause information overload.

Evidence-based data collected in this study revealed an important role played by collection-level subject metadata in matching user search queries and information retrieval of digital collections in aggregations. While collection search was most often satisfied by free-text *Description* collection-level metadata element, it would not retrieve a significant proportion of collection records without controlled-vocabulary collection-level subject metadata (*Temporal Coverage*, *Geographic Coverage*, *Subjects*, and *Objects*).

Both evidence-based data collected in content analysis, transaction log analysis, and observation, and perception-based data collected in interviews proved the importance of collection-level subject metadata to the users of cultural heritage aggregations of digital collections. All the user interactions with aggregations that were observed by the author included viewing of collection-level metadata records. Respondents stated their preference for how information is organized at the collection level and found collection record information sufficient for gaining an understanding of the content of interest. Aggregation users reported that they expect to see subject metadata such as types of objects, subjects, geographic and temporal coverage in collection-level metadata records.

This study also revealed that cultural heritage aggregation users prefer — especially for their research needs — a display of collection metadata where entire collection-level metadata records are shown to the end-user to an alternative approach taken by many aggregations, where end-user is only able to see the free-text *Description* element. Moreover, structured collection metadata display, which includes both free-text and controlled-vocabulary subject metadata elements, works more effectively for broad subject discovery, while a display limited to free-text

metadata alone offers narrower functionality and is mainly suitable for known-item information re-discovery in the more familiar and static aggregations.

Overall, this study supplies evidence for the added value of developing thorough collection-level metadata records in aggregations of digital collections and including subject metadata in these records. More specifically, structured display of collection metadata records in their entirety to the end-users is found to be beneficial for the subject access, and this benefit should be considered by creators of aggregations. More consistent application of controlled-vocabulary collection-level subject metadata elements (*Subjects*, *Geographic Coverage*, *Temporal Coverage*, and *Objects*) would improve the collection search retrieval results and support browse functionality important for cultural heritage aggregation users.

The findings of this exploratory study could become the first step in developing best practice guidelines for creating collection-level subject metadata, which can be incorporated in the future editions of the *Framework of Guidance for Building Good Digital Collections* NISO Recommended Practice document (NISO Framework Working Group, 2007) or *IFLA Guidelines for Digital Libraries*. Development of such best practice recommendations warrants large-scale analysis of collection-level subject metadata application in a number of digital libraries of varying scales (e.g., national, state, regional) and with different subject foci and target audiences (e.g., aggregations of digital collections aimed at researchers, educators, and enthusiasts in the areas of social scientists, humanities, science, technology, engineering, and medicine). Additional qualitative and quantitative characteristics of collection-level subject metadata, beyond application consistency and length, should be explored to inform these best practice guidelines. An in-depth transaction log analysis of user interactions with aggregations of digital collections will provide rich data on evidence-based value of collection-level subject metadata to the users and the role played by collection-level metadata in subject access to digital aggregations.

ACKNOWLEDGMENT

The author wishes to thank members of the research and implementation team of the IMLS-funded Digital Collections and Content project at the University of Illinois at Urbana-Champaign: especially Drs. Carole L. Palmer and Allen Renear for valuable advice and feedback on preliminary drafts of this paper, and Amy S. Jackson for explaining Opening History collection metadata practices. The author also wishes to acknowledge the support of Sally Chambers and Christa Maher who provided The European Library and American Memory collection metadata records for analysis and answered my questions.

REFERENCES

- ALCTS/CCS/SAC/Subcommittee on Metadata and Subject Analysis (1999). *Subject Data in the Metadata Record: A Report from the ALCTS/CCS/SAC/Subcommittee on Metadata and Subject Analysis Working Draft*. Retrieved November 30, 2010, from <http://archive.ifla.org/VII/s12/mom/appendx3.htm>.
- Allen, B., & Sutton, S. (1993). Exploring the intellectual organization of an interdisciplinary research institute. *College & Research Libraries*, 54, 499-515.
- Bearman, D. (1992). Contexts of creation and dissemination as approaches to documents that move and speak. In *Documents that Move and Speak: Audiovisual Archives in the New Information Age*:

- Proceedings of a Symposium held 30 April to 3 May 1990 at the National Archives of Canada*, (pp. 140-149).
- Borgman, C. (1986). Why are online catalogs hard to use? Lessons learned from information retrieval studies. *Journal of the American Society for Information Science*, 37 (6), 387-400.
- Borgman, C. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science*, 47(7), 493-503.
- Cochrane, P. (1985). *Redesign of Catalogs and Indexes for Improved Online Subject Access: selected papers of Pauline A. Cochrane*. Phoenix, AZ: Oryx.
- Cochrane, P. (1986). *Improving LCSH for Use in Online Catalogs*. Colorado Springs, CO: Libraries Unlimited.
- Cochrane, P. (2000). Improving LCSH for use in online catalogs revisited: What progress has been made? What issues still remain? *Cataloging & Classification Quarterly*, 29(1/2), 73-89.
- Cole, T., & Shreeves, S. (2004). Search and discovery across collections: The IMLS Digital Collections and Content project. *Library Hi Tech*, 22(3), 307-322.
- Crystal, A., & Greenberg, J. (2006). Relevance criteria identified by health information users during Web searches. *Journal of the American Society for Information Science and Technology*, 57(10), 1368-1382.
- Drori, O. How to display search results in digital libraries: user study. In *Proceedings of the New Developments in Digital Library, NDDL 2003, 3rd International Workshop*. Retrieved from: www.global-report.com//drori/?l=he&a=3330.
- Foulonneau, M., Cole, T., Habing, T., & Shreeves, S. (2005). Using collection descriptions to enhance an aggregation of harvested item-level metadata. In T. Sumner, & F. Shipman (Eds.), *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '05)* (pp. 32-41). New York: ACM.
- Fraser, B., & Gluck, M. (1999). Usability of geospatial metadata or space-time matters. *Bulletin of the American Society for Information Science*, 25(6). Retrieved February 12, 2011, from: http://www.asis.org/Bulletin/Aug-99/fraser_gluck.html.
- Geisler, G., Giersch, S., McArthur, D., & McClland, M. (2002). Creating virtual collections in digital libraries: benefits and implementation issues. In G. Marchionini (Ed.), *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '02)* (pp. 210-218). New York: ACM.
- Greenberg, J. (2005). Metadata and the World Wide Web. In M. Drake (Ed.), *Encyclopedia of Library and Information Science* (pp. 1876-1888). New York: Marcel Dekker.
- Gross, T., & Taylor, R. (2005). What have we got to lose? The effect of controlled vocabulary on keyword searching results. *College and Research Libraries*, 66(3), 212-230.
- Hjørland, B. (1997). The concept of subject or subject matter and basic epistemological positions. In *Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science* (pp. 55-103). Westport, CT: Greenwood Press.
- International Council of Museums/CIDOC (2007). *Definition of the CIDOC Conceptual Reference Model: version 4.2.3*. Retrieved November 30, 2010, from http://www.cidoc-crm.org/docs/cidoc_crm_version_4.2.3.pdf.
- Jackson, A.S., Han, M.-J., Groetsch, K., Mustafoff, M., & Cole, T.W. (2008). Dublin Core metadata harvested through OAI-PMH. *Journal of Library Metadata*, 8(1), 5-21.

- Jackson, S. (1958). *Catalog Use Study: Director's Report*. Chicago: American Library Association.
- Johnston, P., & Robinson, B. (2002). Collections and collection description. *Collection Description Focus Briefing Paper, 1*. Retrieved November 30, 2010, from <http://www.ukoln.ac.uk/cd-focus/briefings/bp1/bp1.pdf>.
- Kurtz, M. (2010). Dublin Core, DSpace, and a brief analysis of three university repositories. *Information Technology & Libraries, 29*(1), 40-46.
- Larson, R. (1991a). Between Scylla and Charybdis: Subject searching in online catalogs. *Advances in Librarianship, 15*, 175-236.
- Larson, R. (1991b). The decline of subject searching: long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science, 42*(3), 197-215.
- Lee, H. (2000). What is a collection? *Journal of the American Society for Information Science, 51*(12), 1106-1113.
- Lee, H. (2003). Information spaces and collections: Implications for organization. *Library & Information Science Research, 25*(4), 419-436.
- Lee, H. (2005). The concept of collection from the user's perspective. *Library Quarterly, 75*(1), 67-85.
- Lipetz, B. (1970). *User Requirements in Identifying Desired Works in a Large Library: final report*. New Haven, CT: Yale University Library. Retrieved November 30, 2010, from <http://www.eric.ed.gov/PDFS/ED042479.pdf>.
- Lourdi, I., Papatheodorou, C., & Doerr, M. (2009). Semantic integration of collection description: Combining CIDOC/CRM and Dublin Core Collections Application Profile. *D-Lib Magazine, 15*(7/8). Retrieved November 30, 2010 from <http://www.dlib.org/dlib/july09/papatheodorou/07papatheodorou.html>.
- Macgregor, G. (2003). Collection-level descriptions: metadata of the future? *Library Review, 52*(6), 247-
- Matthews, J., Lawrence, G., & Ferguson, D. (Eds.), (1983). *Using Online Catalogs: A Nationwide Survey: A Report of a Study Sponsored by the Council on Library Resources*. New York, NY: Neal-Schumann.
- Miller P. (2000). Collected wisdom: some cross-domain issues of collection-level description. *D-Lib Magazine, 6*(Sept.) Retrieved November 30, 2010 from <http://www.dlib.org/dlib/september00/miller/09miller.html>.
- NISO Framework Working Group (2007). *A Framework of Guidance for Building Good Digital Collections*. (3rd ed.). Bethesda, MD: National Information Standards Organization. Retrieved November 29, 2010, from <http://www.niso.org/publications/rp/framework3.pdf>.
- Palmer, C., Zavalina, O., & Mustafoff, M. (2007). Trends in metadata practices: a longitudinal study of collection federation metadata. In R. Larson, E. Toms, & S. Sugimoto (Eds.), *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '07)* (pp. 386-395). New York: ACM.
- Tagliacozzo, R., & Kochen, M. (1970). Information-seeking behavior of catalog users. *Information Storage and Retrieval, 6*, 363-381.
- Wang, P., & Soergel, D. (1998). A cognitive model of document use during a research project. Study 1. Document selection. *Journal of the American Society for Information Science and Technology, 49*(2), 115-133.
- Weagley, J., Gelches, E., & Park, J. (2010). Interoperability and metadata quality in digital video repositories: a study of Dublin Core. *Journal of Library Metadata, 10*(1), 37-57.

Zavalina, O. L., Palmer, C. L., Jackson, A. S., & Han, M.-J. (2008a). Assessing descriptive substance in free-text collection-level metadata. *Proceedings of the International Conference on Dublin Core and Metadata Applications*. Berlin, Germany, Sept. 22-26, 2008.

Zavalina, O. L., Palmer, C. L., Jackson, A. S., & Han, M.-J. (2008b). Evaluating descriptive richness in collection-level metadata. *Journal of Library Metadata*, 8(4), 263-292.