

Collection Definition in Federated Digital Resource Development

Carole L. Palmer, Ellen M. Knutson, Michael Twidale, Oksana Zavalina

Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign,
501 E. Daniel, Champaign, IL 61820

email: Carole L. Palmer (clpalmer@uiuc.edu), Ellen M. Knutson (eknutson@uiuc.edu), Michael Twidale (twidale@uiuc.edu), Oksana Zavalina (zavalina@uiuc.edu)

As part of a federation project providing integrated access to over 170 digital collections, we are studying how collections can best be represented to meet the needs of service providers and diverse user communities. This paper reports on recent results from that project on how digital resource developers conceive of and define their collections. Based on content analysis of collection registry records, survey and interview data, and focus groups, we identify collection definition trends including a broadening of target audiences, elaboration of subject representation, and a lack of clearly defined selection criteria. Our findings reveal high variability and ambiguity in the collection construct. We discuss how the concept of collection is being continuously defined through the processes of digital resource development and federation and how rapidly changing conceptualizations are likely to impact adoption, tailoring, and development of digital collections and their use.

Introduction

While it has been noted that our field has never had a widely accepted definition of the term “collection” (Lee, 2000; Hill et al., 1999), a common understanding of the construct seems more elusive than ever. On a practical level our collection practices are changing in the digital era, and the principles that guide collection building are evolving. Conceptually, the fluidity of digital content seems to have made the idea of the collection more abstract than in the past (Manoff, 2000), even though the dynamic nature of collections has always been accepted and managed in practice, as repositories of all kinds commonly add and delete objects (Currall, Moss, & Stuart, 2004).

Characterizations of digital collections vary widely in the literature (e.g., International Council of Museums/CIDOC, 2002; Johnston & Robinson, 2002). Traditional user-based collection criteria are still being considered in the design of selected digital collection services (Lagoze & Fielding, 1998). Permanence is emphasized by some, while others stress transience or are neutral on the subject. Collection developers, curators, and users may be highly visible or largely unacknowledged. Collections have been conceptualized as contexts for information seeking (Lee, 2000) and bodies of raw materials for interpretation and presentation (Lynch, 2002). Moreover, as we will discuss in this paper, among digital content developers there is little agreement on what constitutes a collection.

Surely we could continue to build and improve the digital holdings of our individual institutions without a shared understanding of what a collection is. But, as distributed collections are federated into diverse, large-scale repositories, and that content is further repurposed into new specialized aggregations, the concept of “the collection” takes on renewed importance. It may be of use to institutions that wish to have their assets effectively represented in the vast mass of web-based content and for users who need contextual information to evaluate content and understand relationships among digital sources. The concept may also take on new

connotations or prove to be insufficient as a basis for understanding the units that make up shared repositories.

Our recent work building an IMLS Digital Collection Registry and an item level repository (hereafter referred to simply as “the Registry”) for 176 cultural heritage digital collections led us to questions about the role of the collection as a defining or organizing unit in the digital environment. As we developed and implemented a collection level metadata schema and studied metadata practices more generally, we learned about the various ways that resource developers conceive of collections and what attributes they find most important in describing their collections. Because of the many types of participating institutions in the Registry – including academic and public libraries, museums, archives, botanical gardens, and historical societies – the different approaches, or what we refer to in an earlier paper as “cultures of description,” quickly became apparent (Palmer & Knutson, 2004).

In this paper our aim is to represent how our participants conceive of their collections and their ideas about the role of the collection construct more generally in digital resource development, and to consider the implications for collection description. The variety of conceptualizations reveals, we believe, an evolving understanding of the potential of digitization to support different kinds of uses in different ways. We also consider what related concepts, such as project, resource, and exhibit, contribute to our understanding of how to represent digital content. The analysis sets a foundation for our continuing work on the value of collection description and will hopefully stimulate further discussion in the field on how description decisions impact the design and use of federated digital resources.

Methods

During the course of the Registry project we have assembled a rich body of data using several techniques, including a content analysis, questionnaires, semi-structured interviews, and focus groups with digital project managers and metadata librarians, who we refer to hereafter as “resource developers.” This paper is based on a subset of these data related directly to the problem of collection definition. To expand our data on collection issues, we also performed content analysis of changes made by resource developers to collection level metadata records in the Registry and conducted eleven interviews that focused specifically on thoughts and decisions about how collections are represented in those records. Below we give a more detailed overview of the methods design for the study.

Stage 1, background: The first stage of data collection was a content analysis of original grant proposals to identify the kinds of institutions, range of collections, metadata schemes and standards, aims of the projects, and personnel involved across Registry participants. In cases where information was not specified in the proposal, data were collected from project web sites. These data were gathered for ninety-five digital collection projects launched between 1998 and 2002. Later, twenty-seven additional projects launched in 2003 were integrated into the analysis, resulting in a working sample of 122 projects.

Stage 2, baseline: In 2003 two surveys were administered to the 122 Registry participants to gather data on metadata practices and perceptions. The response rates were 76% for Survey 1 and 72% for Survey 2. Survey 1 was designed to verify, update, and augment key administrative, content, and technical information gathered in the content analysis, and to identify changes in these factors since the project was first proposed. This survey asked for information about the type of material in the digital collection, what metadata schemes were used, the intended audience, and other specifics about the digital collection and its technical implementation. Survey 2 was more subjective in nature, soliciting information about how the

institution would use a Registry of digital collections, what elements should be included in a collection description, and issues or problems encountered applying, using, and trying to share metadata. With that baseline in place, we are currently conducting a follow-up survey to monitor progress and change. Supplemental baseline data were collected during this stage through an e-mail follow-up survey and a focus group. These questions probed for further information about differentiation of subcollections and efficacy of metadata schemes. A focus group was conducted in March of 2004 with 11 participants representing 9 institutions to further investigate current functionality and use of the digital collections by end users.

Stage 3, case development: During the interim between the two survey rounds we studied selected projects in more depth, drawing a purposeful sample of twenty projects from the academic, public, and state libraries, museums, historical societies, and other cultural heritage institutions, to capture the full range of operations and requirements of various services and user groups represented across the Registry participants. This stage of data collection began with 23 phone interviews with resource developers to discuss their experiences with collection building and metadata application. Interviews probed for details on: 1) the history and background of the project, 2) views on collection level descriptive metadata and what is important in representing particular collections, 3) roles of a Registry, 4) staffing and technical issues or problems, and 5) sense of the usage of the digital collection. In February of 2005 a second focus group was conducted with nine resource developers to collect data on functionality and interoperability of metadata schemas used, and expectations for the project.

Stage 4, collection focus: Data collection in this stage began once the Registry went live in February 2005. The Registry was populated with collection records developed by the registry development team. Resource developers were responsible for editing the entry for their collection and for adding child or subcollection records for the main collection record. We have been tracking changes made to collection records by resource developers and conducting phone interviews to validate our observations and learn more about collection description perceptions and practices. Eleven interviews were conducted in Fall 2005. The sample was selected to represent diversity in terms of institutional affiliation, metadata schemes used, and the types of materials included in digital collections. The open-ended interview questions probed for information about the reasons for making changes or additions to collection level description, level of satisfaction with the controlled vocabulary used for subject description, ability to capture collection uniqueness in collection description, and adequacy of harvested item-level metadata for repository end-users. Additionally, the Registry transaction log has been analyzed for types of searches conducted and semantic similarity between user keyword queries and terms in controlled vocabularies.

This staged and multi-method approach has allowed us to perform analysis across a large sample of projects to address general questions on metadata applications while studying a smaller, representative sample in more depth for fuller analysis of our specific research questions related to collections. The approach gave us the ability to assess views and actions prior to implementation of the schema for the Collection Registry and responses to the functioning prototype of the Registry and metadata repository. In the most recent round of data collection, participants had not only reviewed their collection descriptions but could also assess how their collections and items “behaved” after being federated with over 170 other collections in the Registry. Data collection and analysis is ongoing through 2007 as we continue to study the implications of collection definition for federated repositories. This paper presents results from the content analysis of collection level metadata records, transaction log analysis, surveys, focus groups, and the completed stage 3 and 4 interviews related to collection definition.

Results

Explicating collections

"We have a problem with that word collection. We fought about that word, so when you use it what do you mean?" (Interview AK031217).

Our respondents working in the day-to-day world of digital resource development were grappling with the "collection" construct in terms of its role and relevance in the digital environment. As the excerpt above illustrates, they were also debating the meaning of the term itself. Some placed a high degree of importance on the uniqueness, history, and name of their collection(s). They often stressed the physical collections from which their digital resources were derived and the need to recognize the individual institutions that own the materials. Moreover, they identified a problem with not knowing where "the collections begin and end." However, a number of respondents raised questions about the need for such structures and differentiations, especially when surmising the user's experience with digital materials. There was a general sense that users do not care about collections and thus are not assisted by collection description or a collection orientation in the design of resources.

When asked how they decided what to include in their digitization project, many discussed the scope and parameters in very general ways, outlining the basic content areas. Several gave brief, loose criteria such as "humorous," "interesting," "poignant," or indicated their decisions were based on what would be "useful for users." Few had articulated formal selection or inclusion policies or guidelines, beyond what was originally outlined in their grant proposals.

In fact, only eight Registry records contained information in the collection development policy field. Half of these briefly stated their collection development focus, which tended to be defined by the collection's audience, geographical and temporal scope, or physical collection(s) from which the digital was derived. For one collection the need for preservation was listed as one of the factors influencing selection decision-making. Only one statement stressed selection practices based on the value and significance of the documents. Another four Registry records linked from the collection development policy field to some kind of a document located on the project's website. Little consistency was observed across these documents. Only one of them was actually entitled "Collection Policy" and provided approximately a page-long detailed outline. The rest were just general "About" or "Collection Strengths" sections of the website that provided brief information on sources the digital collection was derived from, basic geographical and temporal scope, or types of materials in the collections, without outlining any actual selection criteria.

Differentiating collections

"I would consider those multiple collections, but ... I guess that it could be considered a single collection as well." (Interview IF030905)

A slight majority of respondents considered their digital resource to be multiple collections. When they made clear differentiations, the breakdown was usually made by institution for collaborative projects, while format was the most common distinction for non-collaborative projects. However, many resource developers clearly stated that they were creating one new thematic or "intellectually meaningful" virtual collection, which is "more curated" or has a "unifying purpose." A surprising number of respondents did not have a firm idea of how many collections they were creating or vacillated on the topic. For example, an

archivist took the view that their resource was an integrated whole, which is somewhat typical for an archival perspectives on collections, but in the course of discussing the matter fell back into talking in terms of identifiable multiple collections. Others just neutrally presented both points of view, asserting it could be thought of as one or several.

The survey data also revealed some uncertainty about subdivisions within collections. More than 75% of respondents reported that their collections are divided into subcollections, with most listing between 3-10 units. Several specified much larger numbers between 20 and 50, and above. The most common subcollection divisions were by topic and type of material, with geographic categorization also emphasized by many, as well as time period, and to a lesser degree, audience. Some differentiated administrative units, with categorizations for both owning institution and subunits within institutions. But while contributing institution and associated collections were considered viable subcollection differentiations by many at the time of the survey, in practice most resource developers just filled in the related attributes in the collection record and did not specify child collections in those areas. We suspect that some of the high subcollection numbers reported in the survey were actually reports of the number of subject areas or contributing institutions represented in their collection, since this level of subdivision was never actually represented in the collection records or clearly identified in the resources themselves. The content analysis of the Registry records showed that only four collections listed subcollections (from 4 to 26 of them) in their collection records, and differentiations were mostly by contributing institution, topic or time period, and sometimes by genre of objects in collection (e.g. portrait vs. poster), authors, or provenance. The discussion among the focus group attendees on subcollections emphasized the importance of education themes, exhibits, institutions, material type, and topics of scholarly or historical interest as possible subcollection delineations. And, interestingly, the user perspective emerged again in this session, with the group suggesting that subcollections can make materials appear “too chosen” for users’ purposes.

Describing collections

“I think [collection description] is useful for users but it might also be confusing ... I cannot say how to improve it, how not to make it confusing.” (Interview DS051114)

While many projects represented in the Registry were completed in recent years, few developers had given consideration to collection description issues, beyond perhaps a response in their grant proposal to a standard question on sharing collection-level descriptive records. However, once the Registry was created, they were responsible for reviewing and editing the Registry records. Between January 2005 and April 2006, 47 resource developers made a change or addition to their collection record, five came back a second time to make additional modifications. Nine people reviewed their entry but made no additions or changes. Twenty complete collections were added. It is important to note that in our analysis we do not differentiate between parent and child records because they are equally detailed. In other words, when we indicate that a collection record was changed it could be a record for a parent or a child collection.

The vast majority of the changes were additions to the collection descriptions, not a change of what was filled in by the Registry development team based on survey and website information. That is, most resource developers enriched their records rather than refining them. If changes or additions were made, between 1 and 16 fields were modified. More than half modified more than 5 fields. Figure 1 shows the number of records for

which a field was modified. Here we discuss the two most frequently changed fields focusing mainly on the changes to the subject field.

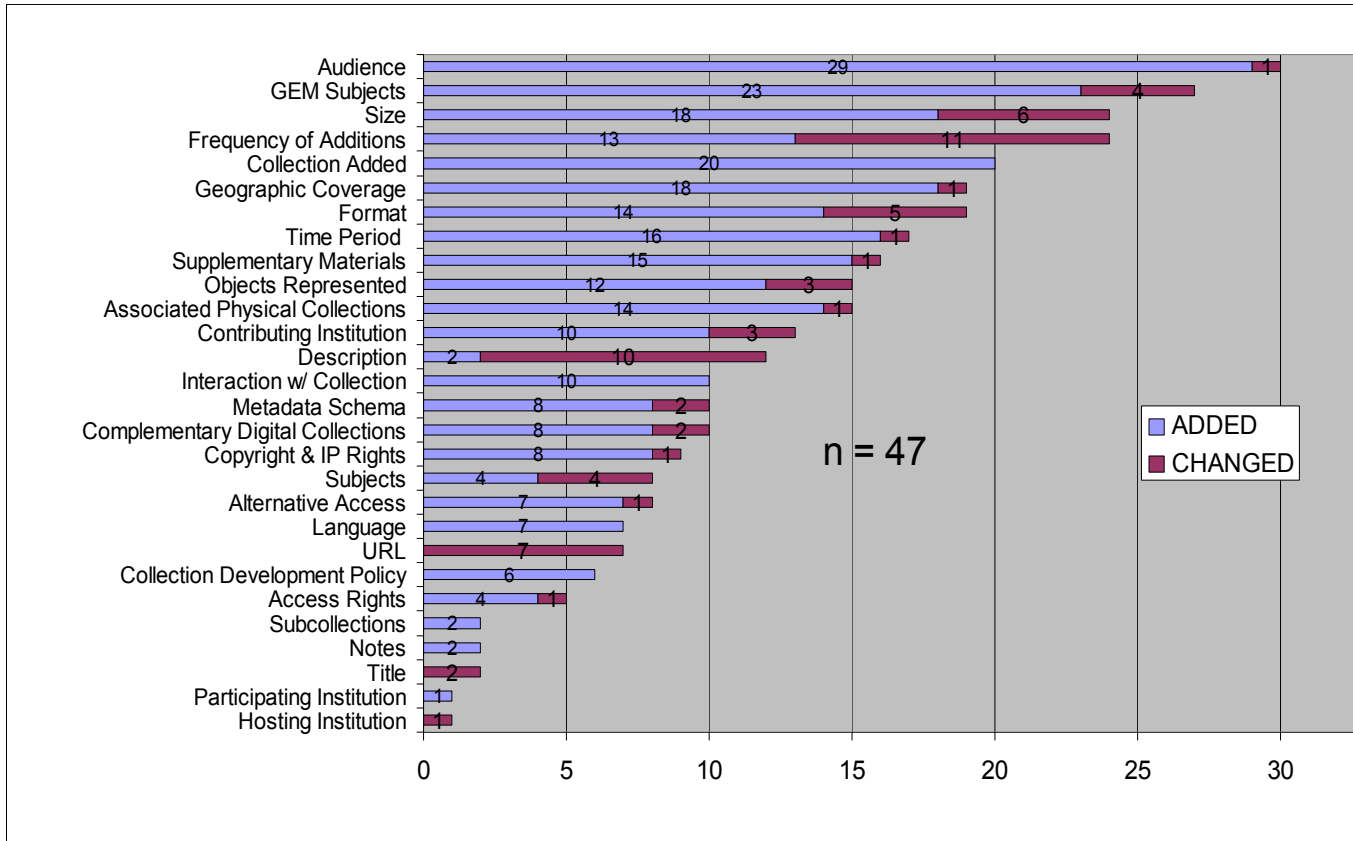


Figure 1. Changes to collection description records

The field that participants modified most frequently was audience, with 30 developers adding anywhere from 1 to 12 new audiences. Most added 1 or 2 new audiences; it was only when there were no audiences listed originally that larger numbers were added.

GEM Subjects was the next most frequently modified field, with twenty-seven modifications. Again the vast majority were adding headings, both at the top level (1-3 headings added for 17 collections) and the 2nd level (1-54 headings added for 25 collections). There were four changes made: two collections removed one 1st level GEM subject heading each, while one collection removed one 2nd level and another removed two 2nd level subject headings; one collection replaced one 2nd level subject heading.

Preliminary content analysis of the subset of 22 modifications to the GEM subject field shows little conceptual match between the GEM subject headings used in collection description and the subject terms used at the item-level. The item-level description subject terms tend to have a much higher level of specificity than GEM subject headings. Nineteen of these twenty-two collections used alternative Subjects field in their collection descriptions in addition to GEM Subjects, and there is only minor conceptual overlap between the two groups of subject terms exhibited in only 7 collections.

Transaction log analysis also demonstrates little match between the subject query terms entered by the Registry searchers and the GEM subject terms (Zavalina, 2006). A sample

of more than 60% of all user keyword queries made in the collection Registry between February and September 2005 was compared to the terms in three controlled vocabularies – GEM, Library of Congress Subject Headings, and Art and Architecture Thesaurus – in search for semantic match. Both authorized and alternative forms of the subject terms were incorporated into analysis whenever available. The analysis reveals a low level – 11 % – of semantic match between the actual user keywords and GEM subject terms. Art and Architecture Thesaurus, although much more detailed than GEM subject scheme, matched only 23% of user keywords, while terms from Library of Congress Subject Headings matched 71% of user queries. Such a situation might be explained by another finding: while the four most widely applied types of user keyword queries in the Registry were found to be concept (24%), person (22%), object (20%), and place (16%), GEM subject scheme deals with only one of them – concept – and Art and Architecture thesaurus does not cover place and person.

It is clear from the above analysis that GEM subject scheme does not fully meet the needs of collection description. Phone interviews revealed that although resource developers are not fully satisfied with the subject breadth and scope of the GEM subject scheme, they are not aware of any good, single alternative for describing collections in such a diverse repository.

Federating collections

“[The Registry] could be really important in contacting the movers and shakers of the ...community, ...but I would never think of that as a primary source.” (Interview NT030820)

Resource developers’ views on the role and value of federated collections were largely amorphous. This is not surprising since we found that for many their ideas about their own collections were still being formulated as their projects progressed. Although all respondents were aware of our collection registry initiative, the surveys and the interviews indicated that developers were unsure of the value of such federated resources for their institutions. Only 40% of survey respondents recognized how a federated resource could benefit reference and research services at their institutions, and few perceived it as a helpful tool for end users.

As might be expected from a group of digital resource developers, there was considerable interest in the Registry as a source for information on up-to-date practices for digital projects and grant funding trends. But, clearly this type of current awareness could be achieved with a simple project directory and would not require building a formal metadata registry and repository. There were scarcely any comments about the repository’s potential for supporting programmatic resource sharing or the creation of new configurations of collections. Those who did recognize these kinds of contributions tended to be involved in large collaborative projects that were, in themselves, a kind of collection aggregation or federation.

Users of collections

“The user doesn’t care; the user doesn’t even have to know what the word collection means.” (Interview OD030821)

As discussed in the Describing Collections section of this paper, the majority of the modifications made in collection Registry records by project managers dealt with specifying the user audience, mostly widening it. In-depth discussion on users in the most recent set of interviews afforded insights into how user groups are understood by resource developers, particularly in relation to the audience attribute in the collection schema. Many developers were hesitant to give a concrete description of their audience because they only had anecdotal evidence of who was using their collections. But, it was common for audience to be thought about in two ways, as the audience(s) that resource developers hoped for and the audience(s) that actually seemed to be using the resource. Several respondents noted that they were currently focusing on studying who their audience is and how to better serve them. Some of these respondents also predicted that they would alter and probably limit their audience field in their collection description based on results of such studies.

In terms of selection decisions there was a split on whether specific audiences were considered when developing a collection. For instance one collection geared sections of their resource to K-12 teachers and included auxiliary content on how the material could be used in the classroom. The interviews revealed that some resource developers believe it is impossible for a digital collection to be all things to all people and that customization might be a good solution to the problem. Yet, many developers did not feel that they had a distinctive audience. Essentially any audience could use the collection with none having a particular advantage for interacting with the material. A number of projects prided themselves on being able to serve the needs of many diverse audiences.

User input was being considered in the creation of digital collections in academic library settings. For example, in one case an advisory board of researchers came up with ideas for new digital collections to support their research and teaching, and suggested specific materials they wanted to see in these new collections. Similar responses to users were documented for museum and other cultural institutions settings where user perspectives are less commonly consulted in collection building. In one case, teachers requested that developers of museum multimedia software – a historical game designed to be played solely on museum workstations – recode their product into a take-home version that teachers now use to prepare students for museum visits. In another case, user focus group results pushed museum collection developers to work with the owner of a specific geological slide collection. In a third case, selection practices of a historical society were affected by teachers' involvement and the need to meet learning standards.

Alternatives to “collections”

“It’s some sort of other thing that we don’t have a shared term for really.”
(Interview MC051024)

For most developers it is quite easy to describe and discuss the nature of what they think of as their “project,” while the collection notion often remains ambiguous. Some collection ambiguity is of course spawned by the grant writing process that is essential for getting external support for digital collection building. Crafting interesting new projects in the form of a competitive grant proposal, and the actual work of managing those projects once they are funded, blur the notions of project and collection. Often there is not a one-to-one relationship between a project and a collection. A project may produce or coordinate production of multiple collections or a collection may consist of distinct subcollections based in different projects. And, of course, in the digital environment, materials digitized for a particular project can be added to more than one collection.

Some project developers think about their resources as something other than a project or a collection. The terms applied often related to the background of the developer and the culture of description for their profession. For example, some archivists applied the term archive, and some discussed their archive in relation to “artificial” and “organic” collections, differentiations traditionally used in archival theory and practice (e.g., *A Manual for Small Archives*, 1999). Museum curators, not surprisingly, frequently spoke of their resources as “exhibits,” but other developers have also adopted this term, which is evidence of its wider acceptance in the digital library community (e.g., Tennant, 1999). The related notions of “display” or “tour” were also applied. To a lesser degree, resources were also discussed as “tools” and “lessons.”

Discussion

The lack of explicit collection criteria for most resources was unexpected, especially since 64% of survey respondents were associated with academic libraries where policy based selection for collection development is common practice. Considering the solid body of literature on selection practices for digitization (e.g., Brancolini, 2000; De Stefano, 2001; Hazen, Horrell, & Merrill-Oldham, 1998; Gertz, 2001), it is unclear why so many projects would proceed with a limited plan or vague criteria. While it is true that for many projects there might be times when formal guidelines would need to be altered due to technical, procedural, or funding constraints, principles of criteria based collection development have served libraries well in practice, even as they have been applied to digitization efforts.

As yet, the tradeoffs among creating records for child collections, differentiating subcollections, and embedding distinctions in parent metadata records are not fully understood. While many developers indicated an intention to break down their collections into subcollections, especially by contributing institutions or subject areas, in practice many opted to represent these aspects as attributes within one collection level record. This tendency may be related to the description cultures. Archivists are quite accustomed to describing collection level units, while many librarians are not. Our research team will be working to try to understand if there are functional benefits to having more systematic child descriptions or at least more uniform articulation of parent-child relationships. The notes field is another encoding aspect that needs further investigation. There has been minimal use of this field thus far, which could be interpreted as an indicator that the schema developed by our project team was sufficient for describing collections without applying such an open-ended field. On the other hand, this is the only field where the special character of a collection, as an “exhibit” or a “tour” could be articulated.

Resource developers had concerted ideas about the limited value of explicating or differentiating collections for users of digital resources. But, the notion of the user in relation to these issues was usually discussed very generically by respondents, as if there is one prototypical user envisioned. And, the tendency in editing collection records was to enlarge the identified user base to include a wider range of the general public. At the same time, user testing at some sites has identified some very distinct user groups, such as professional architects and genealogists, and teachers were an initial focus of many projects. Moreover, if our collections are being created as “lessons” or “exhibits” this suggests that some customization is likely to be in order in relation to these purposes, if not user groups associated with these aims.

Respondents’ unfamiliarity with federated repositories was surprising considering that this population of digital resource developers represents professionals that are very likely to have a

higher degree of awareness of trends in digital libraries than the typical library, museum, or archive professional. Their inability to imagine applications for their institutions or their users suggests a need for wider professional development on the potential and future role of federated digital libraries. But, our perspective is that the potential of federation for providing new information resources and services is an area that needs to be taken up by practitioners to help guide priorities for research and development. For example, quite a few projects added information about associated physical collections, which suggests that traditional notions of collection are still important in articulating the context for items in a collection. But, there is still much to learn about how collections might function as “contexts” (Lee, 2000) from the user perspective. While many resource developers seemed convinced that the collection unit or description held no value for users, previous research has shown that humanities scholars identify strongly with library collections at their institutions considering them critical capital for producing quality research (Brockman et al, 2001). It is yet to be determined if the same value can or should be transferred into digital federations or if we should be striving for new types of value across larger communities of users.

Diverse uses of the term collection

"When I use a word," Humpty Dumpty said, in a rather scornful tone, "it means just what I choose it to mean, neither more nor less." (Carroll, 2005)

In exploring the potential forms and uses of digital collections one may well feel that at times they have moved into a virtual, indeed a looking glass world, one that at times appears very like the physical world, and at other times very different. In undertaking this project, we have been surprised at the variety of ways that the term “collection” has been used, and indeed that the term is noted as fuzzy or problematic by some of our respondents. We believe that it is interesting to consider this issue for the light it can shed on the evolving understanding of the potential and practice of creating and using digital resources and in relation to the use of physical resources.

Given this observed vagueness, varieties of use, and concern voiced about the term, it could be suggested that our respondents might be ignorant of how the term is used in different ways in the fields of Library and Information Science and Museum Studies. That is highly unlikely, given the skills, expertise, and professionalism exhibited in their work and their discussions with us. It could also be that since many projects involve multidisciplinary collaborations, participants have discovered that different disciplines use the word “collection” to mean subtly different things. The problems with misunderstandings arising from these different shades of meaning may have led them to be cautious of using the term. There is some evidence for this reason applying in some cases, but it may not be the sole cause. Another reason might be that the word “collection” as it is used for activities in the digital world is being forced to serve many different purposes – sometimes several purposes at once, not only by different members of a project, but even by an individual.

The potential of digitization to allow new kinds of uses means that the conceptualization of collection may be put under considerable strain to encapsulate all of them. While some uses may be completely novel, others are best viewed as new combinations of pre-existing practices. This leads to different themes playing out with different levels of emphasis in different projects at different times. For example, a digital collection might be viewed as the same as the originating physical collection in just another manifestation, as a selection from it, as a selection from many collections, as an accumulation of different collections into a super collection, as adding informational value to the central physical

collection, as becoming something else in its own right, and yet more. Note that any one project can have aspects of several of these themes at once. Similarly, individuals may consider a single theme as a central aspect of their concern, or take several into account as they conceptualize and refine their project. In the light of these themes (and there may well be many others) it becomes much less surprising that the act of digitization has led to a variety of creative uses and something of an overloading of the meaning of the word “collection.”

Conclusion

“I think it [digitization] is allowing us to start thinking about it [the collection] differently.”
(Interview EW051102)

It seems that in the digital world, the concept of collection has a rather protean nature. This accounts, we believe, for the difficulties that our highly articulate and thoughtful respondents had when we questioned them about this particular issue. It is possible to construct many different kinds of digital collections with many kinds of features and uses. A single project may incorporate several of these features and uses, and add or invent yet more as it evolves. It is this evolving use and meaning that is of greatest interest. The terminological issues are best seen as symptoms that can help us gain greater insight into the underlying causal issues.

Given such diversity of uses and meanings of digital resources, it is important to consider the implications for education, project planning, research, and systems design. The most immediate need, and fortunately maybe the easiest to put into action, is to raise awareness of this diversity of meaning and of potential uses. If different individuals, teams, organizations, and potential funding bodies believe that there is a single meaning and that all members of the community can or should share that meaning, we are in danger of much costly talking at cross purposes. Merely acknowledging the variety of interpretations and the strengths and weaknesses of each can help in undertaking a more productive dialogue.

By studying a large group of different digitization projects we have found that there is considerable diversity in the understanding of the concept of collection within and across projects. We ascribe this to digitization opening up a large possibility space which different projects are exploring in different ways in accordance with their own goals, opportunities, and constraints. For some this has meant that the very term “collection” has become problematic, while others use the term in a range of ways. It is important to consider how to help individuals and projects address the challenge of many possible uses and evolving uses of both physical and digital collections, particularly in the light of federated collections offering yet more potential uses and interpretations.

Acknowledgements

This research was supported by a 2002 IMLS NLG Research & Demonstration grant. Project documentation is available at <http://imlsdcc.grainger.uiuc.edu/>.

References

Brancolini, K. (2000). Selecting research collections for digitization: Applying the Harvard model. *Library Trends*, 48(4), 783-798.

Brockman, W. et al. (2001). *Scholarly work in the humanities and the evolving information environment*. Washington, DC: Digital Library Federation/Council on Library and Information Resources.

- Carroll, L. (2005). *Alice through the looking glass*. Cambridge, Mass.: Candlewick press.
- Currall, J., Moss, M., & Stuart, S. (2004). What is a collection? *Archivaria*, (58), 131-146.
- De Stefano, P. (2001). Selection for digital conversion in academic libraries. *College and Research Libraries*, 62(1).
- Gertz, J. (2001). Selection for preservation in the digital age: An overview. *Microform and Imaging Review*, 30(2), 48-56.
- Hazen, D., Horrell, J., & Merrill-Oldham, J. (1998). Selecting research collections for digitization. *Microform and Imaging Review*, 27(3), 82-93.
- Hill, L. et al. (1999). Collection metadata solutions for digital library applications. *Journal of the American Society for Information Science*, 50(13), 1169-1181.
- International Council of Museums/CIDOC. (2002). Definition of the CIDOC object-oriented Conceptual Reference Model version 3.4. Retrieved June 20, 2006 from http://cidoc.ics.forth.gr/docs/cidoc_crm_version_3.4.rtf.
- Johnston, P., & Robinson, B. (2002). Collections and collection description. Collection description focus briefing paper. No. 1. Retrieved June 20, 2006 from <http://www.ukoln.ac.uk/cd-focus/briefings/bp1/bp1.pdf>
- Lagoze, C., & Fielding, D. (1998). Defining collections in distributed digital libraries. *D-Lib Magazine*, (November). Retrieved June 20, 2006 from <http://webdoc.gwdg.de/edoc/aw/d-lib/dlib/november98/lagoze/11lagoze.html>
- Lee, H. (2000). What is a collection? *Journal of the American Society for Information Science*, 51(12), 1106-1113.
- Lynch, C. (2002). Digital collections, digital libraries, and the digitization of cultural heritage information. *First Monday*, 7(5).
- Manoff, M. (2000). Hybridity, mutability, multiplicity: Theorizing electronic library collections. *Library Trends*, 48(4), 857-876.
- A Manual for Small Archives* (1999). Archives Association of British Columbia. Retrieved February 9, 2006 from http://aabc.bc.ca/aabc/msa/4_organizing_archival_material.htm
- Palmer, C., & Knutson, E. (2004). Metadata practices and implications for federated collections. In L. Schamber & C. Barry (ED). *Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology*, 456-462.
- Tennant, R. (1999). Access: Where the rubber meets the road. *Library Journal*, 124(7).
- Zavalina, O. (2006). User searches in IMLS DCC Collection Registry: Transaction log analysis. Technical Report UIUCLIS--2006/3+IMLS. Champaign, IL: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.