This a pre-print (dated January 29, 2004) for a paper submitted to the 2004 American Society for Information Science and Technology Annual Meeting. The citation for the published article follows:

Palmer, Carole L. and Ellen M. Knutson. 2004. Metadata practices and implications for federated collections. In *Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology,* Edited by Linda Schamber & Carol L. Barry. Medford, NJ: Information Today, Inc: 456-462.

Metadata Practices and Implications for Federated Collections

Carole L. Palmer
Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, 501 W Daniel, Champaign, IL 61820. Email: clpalmer@uiuc.edu.

Ellen M. Knutson
Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, 501 W Daniel, Champaign, IL 61820. Email: eknutson@uiuc.edu.

Abstract

As digital library development begins to focus on interoperability and collection federation, resource developers need to be concerned with contributing to national and international collections, while not losing sight of the needs of institutions and user communities. The Digital Collections and Content (DCC) project aims to provide integrated access to IMLS National Leadership Grant (NLG) digital collections through a centralized collection registry and metadata repository. While technical development proceeds on the repository, our research team is investigating how collections and items can best be represented to meet the needs of both service providers and diverse user communities. This paper presents results on metadata and collection representation practices based on survey data, interviews, and content analysis. Despite Dublin Core's prevalence and ease of use, problems with field richness and consistency of application persist, in part because of the distinct cultures of description that have evolved in different kinds of cultural heritage institutions. Moreover, the concept of a digital collection is widely unsettled among resource developers. This has important implications for central repositories, if, as we hypothesize, the strategic foregrounding and backgrounding of collection-level metadata proves critical for navigation and interpretation of information in large-scale federated collections.

Introduction

Having passed through about a decade of digital library development devoted largely to the production of content, we are now in a phase where efforts are turning toward interoperability and federation. Digital resource developers in libraries, archives, and museums, and the agencies that fund much of their work, recognize that the power of digital collections and the fruition of their substantial investments cannot be fully realized until there are better means for providing

access across large arrays of distributed digital material. Advances are being made on numerous fronts, including the Open Archives Initiative, NSDL core integration efforts, DLF plans for a Distributed Online Digital Library (DODL), UKOLN projects on distributed systems and interoperability, and the IMLS Digital Collections and Content project,[1] on which this paper is based. To date, research and development in interoperability have largely been concerned with technical implementation and the problems associated with integrating large sets of heterogeneous digital objects.[2] Empirical studies of the practices of digital library creation and federation, especially as they relate to the needs of user communities, are few. Yet, resource developers need a better understanding of what is at stake as they make metadata and interoperability decisions and begin to instantiate what are likely to become common practices in the future.

This paper examines current metadata and collection description practices and their implications for collection federation. The report is based on year-one results from our research with the Digital Collections and Content (DCC) project, a three-year study which aims to provide integrated access to the digital content developed through the IMLS National Leadership Grant (NLG) program. The primary objective of DCC is the creation of a centralized collection registry and metadata repository[3] based on the Open Archives Initiative Metadata Harvesting Protocol.[4] While technical development proceeds on the repository, our research team is concurrently conducting surveys and case studies to investigate how collections and items can best be represented in the repository to meet the needs of both service providers and divergent user communities.

Our research questions are multifaceted[5] and require study of metadata applications, interoperability challenges, needs of user communities, and roles of federated collections. The intent is to understand the range and evolution of metadata and interoperability issues encountered by NLG projects over time and how problems can be resolved through assistance to content providers and the development of repository tools. The results presented here coincide with the first phase of repository development which has focused on background research and data collection to 1) establish a baseline on the range of institutions, collections, and metadata represented in NLG projects and 2) inform the design of a collection level metadata schema.

Background

A long-term goal of this research is to determine how to optimize metadata in federated systems to support users' practices and needs. As we gain in interoperability, we do not want to lose advances that have been made in adaptation and access for communities of users. As Lagoze and Fielding (1998) suggest, traditional library collection functions that attend to user-based criteria are key to the success of distributed digital collection services. With federated systems, metadata is the link between collections and the requirements of users. User criteria can be relatively straightforward where the range of content is narrow and the user base is homogeneous. However, the content and users of most digital collections developed by research libraries, museums, and other cultural heritage institutions are multifarious. It has always been expensive and difficult to build heterogeneous collections that support the interests of diverse user communities, and this legacy problem stands as one of the greatest challenges in the digital environment where collections are distributed and idiosyncratic.

Differences in metadata standards reflect the various aims and practices of resource developers and their constituent user communities. In the library profession where digital metadata has regularly been applied to both digital and non-digital works, MARC and Dublin Core have been

widely adopted (Besser, 2002). Those working in particular subject domains are also building sophisticated infrastructures, schema, and guidelines to support their metadata requirements. Initiatives such as OLAC in the field of linguistics, IMS in e-learning, and FGDC in geographic sciences are noteworthy examples, and, for many standards, user communities have informed or been participants in the development of metadata specifications. For example, GEM represents a broad consortium of educators, and TEI was painstakingly developed by humanities scholars.

The digital library community could now benefit from broad assessment of how widely schemes have been applied and their advantages and disadvantages for supporting the actual work of users. Some comparative studies have been done, such as Greenberg (2001) who has shown that for visual resources administration functions are well supported by EAD, but REACH elements are superior for discovery. But, we have a very limited understanding of how schemes compare for communities of users. In general terms, we know that users have little awareness of metadata and that they often do not understand how best to interact with a given system. Our studies of scholars have shown that they experience a kind of digital blur, a lack of control and understanding of navigation and retrieval, and an inability to exploit digital sources for basic scholarly activities (Brockman et al., 2001; Palmer & Neumann, 2002).

With the advent of federated collections, there are increasing expectations that existing standards will be adopted and propagated by resource developers and the generation of reaggregated collections and new services will follow. But, as individual institutions are coming to understand the tradeoffs in metadata choices and applications for themselves and their users, they now must also be concerned with contributing to national and international digital federation efforts. Unfortunately, we have yet to determine the compatibility of local and global requirements. As the DCC project progresses we expect that our research will increase the field's understanding of the value of different metadata schemes and the levels of granularity needed to support functions important to both service providers and user communities.

Methods

Data are being collected from the entire group of NLG awardees in the first and third years of the project to monitor progress and change in metadata practices and perceptions. At the same time, we are conducting a series of case studies of selected projects based at 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 users. This multi-method approach allows us to perform analysis across a large sample of projects to address general research questions while studying a smaller, representative sample in more depth for fuller analysis of specific research questions.

Our first goal was to gather baseline information from the 95 NLG digital collection projects awarded between 1998 and 2002. Twenty-seven additional projects from the recent 2003 NLG awards are also being integrated into the analysis, resulting in a working sample of 122 projects. The first stage of the baseline was produced through a content analysis of the original grant proposals to identify the kinds of institutions, range of collections, metadata schemes and standards, aims of the projects, and personnel involved. In cases where information was not specified in the proposal, data were collected from project web sites. In addition, a two-part survey was administered to project managers of the original 95 projects. (The same survey will be sent to the new projects about six months into the grant cycle.) The first part of the survey 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. The second part was more subjective in nature, soliciting information about participants' experiences applying metadata, the content and form of their collections, and opinions about the role of a repository of NLG collections.

The other major objective for year-one was to begin building case studies that could be followed during the course of the three-year project. Based on the information gathered from the content analysis, we selected a purposive sample of 20 projects that represent the diversity of the NLG program in institution type and size, metadata use, and the type of materials included in their digital collection. Case development began with phone interviews with project managers and metadata specialists to discuss their experiences with collection building and metadata application. Thirteen interview sessions, ranging in length from 40 to 80 minutes, have been completed and fully transcribed. The group includes four university libraries, two public libraries, two historical societies, one university archive, one state library, one botanical garden, one non-profit organization, and one zoo.

Data collection and analysis is ongoing. Over the next few months the remaining seven first-round interviews will be completed and a short e-mail follow-up survey will be administered to all respondents to the survey. The follow-up survey probes for further information about collection description, differentiation of subcollections, and efficacy of metadata schemes. A focus group is also scheduled for March to begin the line of investigation related to functionality and use of the digital collections by end users. Results from the content analysis, surveys, and the completed interviews are presented here.[6] The first section provides a profile of the types of institutions and their metadata practices and experiences based on the content analysis and the survey, drawing from the case study interviews for further explication and discussion of results. The second part is an analysis of survey and interview data related to collection representation practices.

Institutions and Metadata Applications

A total of 326 institutions were listed in the NLG proposals, either as the primary institution or a partner in the 122 projects. Figure 1 shows the breakdown of the number of the institutions by type. One hundred and six academic libraries participated, greatly outnumbering the other types of institutions. In fact, only 29 projects did not involve an academic library, academic department, or a museum based in a university. The next largest category was museums (66), followed by historical societies (28), and public libraries (19). The "Other" category includes six government agencies, four special libraries, five companies, three Native American tribes, two herbaria, a zoo, a state park, and a national historic site. These frequencies do not take into account all contributing institutions. Some projects are statewide initiatives, like the Maine Memory project, which was awarded to Maine Historical Society and currently has over 100 contributors with as many more planning to contribute in the future. Further, each library consortium represents many libraries, one of which is a group of 300 public libraries.

Because of the abundance of university libraries represented in the sample, it has not been meaningful to cross-tabulate data by type of institution. However, there do not appear to be important differences in the survey responses between university and non-university institutions, and the interviews suggest that all institutions are experiencing similar successes and challenges. Moreover, due to the large number of collaborations, many project managers based in academic

libraries are highly aware of museum and archives perspectives and have represented those interests well in their responses.[7]

[Insert Figure 1]

The breakdown of metadata scheme application is presented in Figure 2. Thirty five percent of the projects proposed to use or are using multiple metadata schemes. This tendency needs to be taken into consideration when interpreting the other frequencies presented in the figure. Most multi-scheme projects reported use of Dublin Core in combination with one or more additional schemes. A total of 67 out of the 122 projects, or 54%, reported use of Dublin Core either alone or in combination with another scheme. Only four projects that used multiple schemes did not incorporate MARC or Dublin Core as one of the schemes. Fifteen, or 34% of projects using multiple schemes, chose to use three or more schemes. Although TEI and EAD were rarely used alone, they were applied in combination with other schemes. Slightly more than one third of the projects with multiple schemes used or proposed to use TEI in some form; just under a quarter used or proposed to use EAD. It is interesting to note that use of multiple schemes was not associated with collaboration but was fairly equally divided among collaborative and non-collaborative projects. This was contrary to our initial hypothesis that collaborative projects would be more likely to apply multiple schemes to accommodate the needs of the different participating institutions.

[Insert Figure 2]

Less than one third of all the projects did not use MARC or Dublin Core to describe digital objects in their collections. In our interviews, participants expressed a preference for MARC's field richness, but valued Dublin Core for its perceived ease of application. Nonstandard use of fields seemed to be more prevalent with Dublin Core. For example, in one case the source field was appropriated to provide information about the original object that had been digitized, and in other projects the data placed in the description field had been extended to compensate for the lack of appropriate fields in Dublin Core.

Eleven percent of the projects applied a locally developed scheme exclusively, several of which were derived from Dublin Core. Another 8% used a local scheme in addition to Dublin Core or MARC. Projects chose to apply a local scheme (or in some cases forego descriptive metadata altogether) for a number of reasons: customization was needed to capture information unique to the materials, information already recorded in a database or some other local information source was to be imported, or existing standards did not allow projects to adhere to their goals. For instance, to support specific K-12 education objectives, it was important for one project to define a field for learning standards. In particular, interactivity with the materials was a goal that participants felt was inhibited by the available metadata schemes.

Metadata scheme choice was influenced by factors that might be expected. The degree to which a standard had been previously implemented and tested was of central importance, as was use by peer institutions. Compatibility with local systems was also a driving force. Several librarians reported that their choice of MARC was due to their OPAC's inability to handle Dublin Core records. The three most common problems with description were: consistent application of the chosen metadata scheme within a project, identification and application of controlled vocabularies, and integration of sets of data, schemes, and vocabularies either within an institution or among collaborators.

Revisions of standards and newly developed schemes introduced an ongoing layer of work that

was rarely anticipated in the project plan. In one case, a new DTD for EAD dictated extensive revision of existing records and parsing applications, and the introduction of METS had lead to the assessment of its potential for coverage of image copyright and productions. In addition, there were clear tensions between local practices and what is perceived as the best route for interoperability. One project that began with Dublin Core decided against using it part way into the grant, favoring MARC and TEI for representing the texts in their collection. Later they ended up mapping their metadata back to Dublin Core for OAI interoperability. As projects progress, the unstable standards environment has made it difficult to advance without shifts, reconsiderations, and adaptations in original metadata plans. Naturally, this has implications for staffing and management of metadata operations, another key area of concern for project managers.

Many respondents had difficulty finding qualified staff or discovered they needed a higher level of metadata expertise than expected. We received pointed comments about the need for formal metadata education in LIS schools, but there was not full agreement on this matter. There were managers who reported a high level of success employing graduate students who had no applicable technical background. The metadata expertise challenges did not seem to have a negative impact on the satisfaction level of those working on digital projects, however. Managers reported high morale among project personnel and a general sense of pride in project accomplishments.

The technical problems applying item level metadata were moderate in number but frustrating nonetheless. Early adopters faced hurtles with schemes and software. One contributor to a statewide initiative discussed crashing the first system they implemented during a training session, which lead them to invest considerable effort in programming their own system. Even systems that had been in use for some time had disappointing limitations. One metadata librarian gave an account of basic information retrieval features not supported by their software, including phrase searching and stopwording. As a result, some hotlinked headings in the database would retrieve almost the entire collection.

The importance and utility of collaborative initiatives was strongly represented in the interviews,[8] however collaboration did pose additional metadata challenges. Choosing a metadata scheme that works well across varied cultural heritage institutions is the first challenge, and there was no evidence of any scheme meeting the expectations or needs of all the institutions. Consistent application of the selected scheme was another ongoing difficulty, in part because of the distinct cultures of description that have evolved in different kinds of institutions. One respondent described the problems inherent in their collaboration with twelve partner institutions.

> We're encountering difficulties in choosing controlled vocabularies and data fields that would work across all 12 datasets; reconciling the diverse element sets; dealing with library-style preferences (making it easy to find is of the utmost importance) vs. museum-style preferences (giving information about it once you get there). (Survey 418)

Museums often do not have subject categories or titles associated with their objects but tend to have rich contextual descriptions based on research and interpretation. Integrating these data across institutions requires understanding and negotiation of both professional practices and principles. But, there is a sense that the different descriptive approaches can blend into more comprehensive representations of collections.

> Everybody knows what a title is. Well, if you've got a rock in your museum, what's the title of that rock? Well, no, it's not as simple and straightforward as in the published environment. . . . So, I don't think at the collection level. I tend to think at the item level and to think at the collection level, personally I think that is one of the strengths that our archivists and museum people can bring to the discussion because that is their level of experience and for catalogers you know, the little pieces, parts are what I'm interested in. (Interview NC030814).

The collaborative projects encompass a mix of institutions with different objects and aims. To some degree, they serve as micro models of what DCC is trying to accomplish. This is especially true of the large, statewide initiatives and the projects that are concentrating on metadata production and OAI interoperability. For this reason, we have selected a number of these projects for expanded case study. We will be following their progress closely and collecting additional data from the lead institution, and partners when possible.

Content and Structure of Collections

The objects represented in the NLG collections are highly diverse. They come from a range of cultural heritage institutions and include digital reproductions of sheet music, transcripts of oral histories as well as the corresponding digitized audio files, and digitally produced artwork. The vast majority of collections contain images of artifacts: maps, photographs, museum objects, and different kinds of texts. While the image format limits search and analysis capabilities of text, it does allow users to see the documents in their original form and condition. A small percentage of the collections are exclusively texts, and a few projects are producing encoded texts, with about 20 using or investigating the use of TEI in their projects. The content analysis suggests that for text collections in particular there is likely to be inconsistency in format and type descriptions, indicating a need for us to monitor this closely as the repository develops.

The survey responses to a question about elements that should be designated for collection description proved to be well aligned with the DCC schema under development at the time.[9] Suggestions for elements, beyond the basics of title, subject, description, type, format, etc., often reflected traditional modes for identifying collections in museums and archives, such as by donor or correspondent. Fields for user oriented data were also suggested, such as audience and lesson plans. Participants emphasized institution as a primary element for the collection description scheme, and some suggested all contributing institutions should be identified (which for some projects could be in the hundreds). Many also mentioned the desire to connect or situate the digital collection in the context of the physical collection.

Many questions remain about how projects will actually describe their digital collections with the DCC schema. For instance, there is not always a one-to-one relationship between the project and the collection. A project may produce or coordinate production of multiple collections or a collection may consist of distinct subcollections (Shreeves & Cole, 2003). Moreover, the content of some projects may not be considered collections, per se, by their developers, but rather exhibits, learning modules, or multimedia compilations. In the interviews, respondents frequently did not have a firm idea of how many collections they were creating, suggesting that they may not have yet thought about how their collections should be represented in a federated repository. A few avoided answering the question altogether. This excerpt is a good example of what we encountered: "We have a problem with that word collection. We fought about that word, so when you use it what do you mean?" (Interview RD031217).

A slight majority of participants considered their digital product to be multiple collections. For collaborative projects, the delineation of collections was generally made by institution, while for non-collaborative projects format was the most common differentiation. Some described the digital product as a new integrated whole, but during the course of the discussion, most went back and forth on the matter.

> I guess coming from an archivist point of view I would consider those multiple collections, but … I'm familiar with OAI and those types of activities. I guess that it could be considered a single collection as well in the larger scheme of what type of materials are being collected and how these digital collections are being created, because what we're creating is actually something very new and I guess I would take a wider viewpoint than strict archival provenance. (Interview MS030905)

It was strongly suggested by some that, from the user's perspective, collections were not a useful construct for describing or organizing material in the digital environment.

The survey data also suggested some uncertainty about the internal structure of collections. More than 75% of respondents reported that their collections are divided into subcollections, with most listing between 3-10 units. However, several specified much larger numbers between 20 and 50, and above. We suspect that some of these responses may reflect the number subject areas or contributing institutions represented in the collection rather than differentiated subcollections.

The most common divisions were by topic and type of material, with geographic categorization emphasized by many, as well as time period, and to a lesser degree, audience. Almost an equal number indicated the need to differentiate administrative units, with categorizations including both owning institution and subunits within institutions. In the coming year, we will be able to compare the responses on collections and subcollections to the actual collection description data submitted by the projects to the repository.

Considering that conceptions of individual collections are still being formulated as projects progress, respondents' ideas about the NLG federated collection were understandably amorphous. In both the surveys and interviews, many respondents were unsure of the role of the DCC repository. As might be expected from this group of respondents, there was considerable interest in the resource for information on up-to-date practices for digital projects and IMLS funding trends. But, clearly this type of current awareness could be achieved with a project directory and would not require building a repository. Forty percent of the respondents recognized how the resource could benefit reference and research services at their institutions, but few perceived it as a helpful tool for end users. There were scarcely any comments about the repository's potential for supporting programmatic resource sharing or the creation of new configurations of collections.

Conclusion

The creation of digital collections requires significant resources, not all of which are fully covered by grant funding. The projects are forcing institutions to assess their priorities and make hard decisions about their primary objectives. Public libraries are having the most difficulty making room for new initiatives.

> Our intent was to make the department a part of the regular staffing structure, but over the last two years we've been hit by fairly severe budget cuts. We lost, I would say between 25 and 30 percent of our fulltime staff at the central library. So continuing to finance something

> like this when you're faced with keeping the doors open is, ah, very difficult—a very difficult situation to be in. (Interview CJ030908).

Another public library indicated that they would need to reconsider their commitment to digital projects. None of the new 2003 projects were based in a public library, and only two public libraries were listed as partners.

The basic work of creating, reconciling, and updating item metadata is a huge undertaking even for resource-rich institutions. That not withstanding, production will greatly improve as we are able to support these tasks as routine work and focus on melding interoperability practices with professional, institutional, and user based requirements. For the repository, both item and collection level metadata will be essential for providing different types of discrimination amidst the aggregation. For instance, item description supports retrieval of objects with the same attributes, but richness and consistency of metadata have emerged as key quality issues, and we have just launched a line of investigation in this area.

Collection and subcollection description can help the user ascertain features like uniqueness, authority, and context of the objects retrieved. Collection information provides background for the digital items, a trace to their roots and the work performed by collection developers, curators, and their institutions more generally. In the collection registry, the collection description will be in the foreground, identifying aggregations of objects as meaningful wholes, the associated institutions, and perhaps a rationale for the assemblage and information on related collections. In the item level metadata repository, the collection data will be in the background, providing added context as needed. As federated collections grow and evolve, the strategic foregrounding and backgrounding of collection metadata may greatly improve the digital blur that users experience searching large-scale information resources.

We expect that some of the confusion about collection definition will get sorted out, at least in a pragmatic way, as projects go through the process of creating their collection level metadata records for the DCC repository. At minimum the process should raise awareness of the need for collection description and questions about the impact of differentiating and linking collections and subcollections within and among collection records. We expect to see a range of approaches to description that reflect the variety of guidelines found in the literature. Collections may be represented in very broad terms, neutral to size and transience (Johnston & Robinson, 2002), or with an emphasis on permanence and the actors in the process (CIDOC, 2002), or as information seeking contexts (Lee, 2000). As Hill et al. (1999) argue, collection metadata is critical *because* it is unlikely that one approach to collection definition will be sufficient in the digital environment.

NOTES

[1] This is a collaborative project between the University of Illinois Library and the Graduate School of Library and Information Science, supported by a 2002 IMLS NLG Research & Demonstration grant. Project documentation is available at http://imlsdcc.grainger.uiuc.edu/.

[2] See Hunter (2003) for a recent review of metadata research, with a section devoted to interoperability and coverage of technologies for integration, sharing, and exchange.

[3] Hereafter, the dual collection registry and metadata repository will be referred to simply as the repository.

[4] For background on the OAI approach see Shreeves, Kaczmarek, and Cole (2003).

[5] See http://imlsdcc.grainger.uiuc.edu/researchplan.htm for an outline of the research questions guiding the project.

[6] These results update a preliminary report presented at the 2003 Dublin Core conference (Knutson, Palmer, & Twidale, 2003) and will be further updated with incoming data and analysis for the ASIST 2004 conference proceedings and final presentation.

[7] For an introduction to these perspectives as they relate to collection description see Dunn (2000) on museums and Sweet and Thomas (2000) on archives.

[8] Fifty-nine percent of the projects are collaborative, supported by a special NLG funding track encouraging library-museum collaboration (see http://www.imls.gov /grants/l-m/l-m_lead.asp)

[9] See Shreeves and Cole (2003) on development of the DCC schema, which is an adaptation of the Research Support Library Programme (RSLP) collection schema (Powell, Heaney, & Dempsey, 2000).

REFERENCES

Besser, H. (2002). The next stage: Moving from isolated digital collections to interoperable digital libraries. First Monday, 7(6). Available: http://firstmonday.org/issues/issue7_6/besser /index.html

Brockman, W.S., Neumann, L., Palmer, C. L., & Tidline, T.J. (2001). Scholarly work in the humanities and the evolving information environment. Washington, DC: Council on Library and Information Resources.

CIDOC. (2002). Definition of the CIDOC object-oriented Conceptual Reference Model version 3.4. Available: http://cidoc.ics.forth.gr/docs /cidoc_crm_version_3.4.rtf.

Dunn, H. (2000). Collection level description – the museum perspective. D-Lib Magazine, 6(9). Available: http://www.dlib.org/dlib /september00/dunn/09dunn.html

Greenberg, J. (2001). A quantitative categorical analysis of metadata elements in image-applicable metadata schemas. Journal of the American Society for Information Science and Technology, 52(11): 917-924.

Hill L., Janee G., Dolin R., Free J. & Larsgaard M. (1999). Collection metadata solutions for digital library applications. Journal of the American Society for Information Science, 50(13): 1169-1181.

Hunter, J.L. (2003). A survey of metadata research for organizing the web. Library Trends, 52(2): 318-344.

Johnston, P. & Robinson B. (2002). Collections and collection description. Collection description focus briefing paper. No. 1. Retrieved Online May 12 2003. http://www.ukoln.ac.uk/cdfocus/briefings /bp1/bp1.pdf

Knutson, E., Palmer, C.L. & Twidale, M. (2003). Tracking metadata use for digital collections. In DC-2003: Proceedings of the international DCMI metadata conference and workshop, pp. 243-244.

Lagoze, C. & Fielding, D. (1998). Defining collections in distributed digital libraries, D-Lib Magazine (November). Available: http://www.dlib.org/dlib/november98/lagoze /11lagoze.html

Lee, H. (2000). What is a collection? Journal of the American Society for Information Science. 51(12): 1106-1113.

Palmer, C.L., & Neumann, L.J. (2002). The information work of interdisciplinary humanities scholars: Exploration and translation. Library Quarterly, 72(1): 85-117.

Powell, A., Heaney, M., & Dempsey, L. (2000). RSLP collection description. D-Lib Magazine, 6(9). Available: http://www.dlib.org/dlib /september00/powell/09powell.html

Shreeves, S.L. & Cole, T.W. (2003). Developing a collection registry for IMLS NLG digital collections. In DC-2003: Proceedings of the international DCMI metadata conference and workshop, pp. 241-242.

Shreeves, S.L., Kaczmarek J.S., & Cole, T.W. (2003). Harvesting cultural heritage metadata using the OAI protocol. Library Hi Tech, 21: 159-169.

Sweet, M. & Thomas, D. (2000). Archives described at collection level. D-Lib Magazine, 6(9). Available: http://www.dlib.org/dlib /september00/sweet/09sweet.html

FIGURES

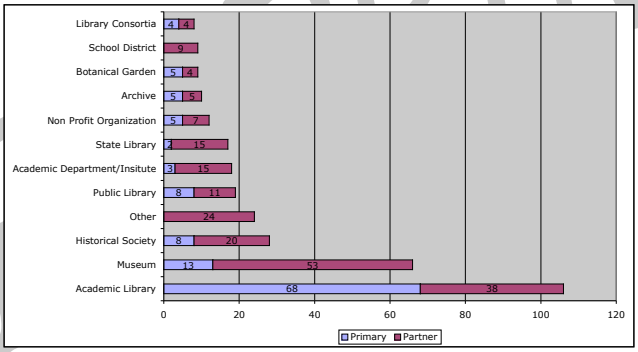

Figure 1: Number of Participating Institutions by Type

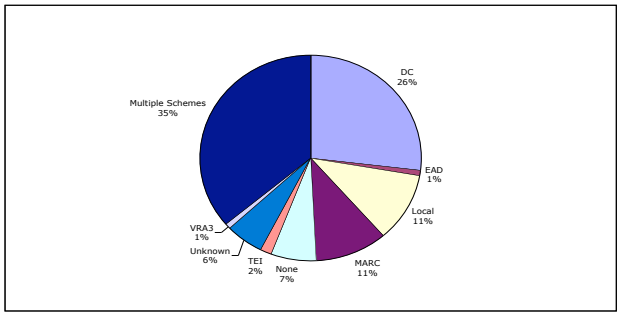

Figure 2: Percentage of Projects by Metadata Scheme