DCPAPERS

Tracking Metadata Use for Digital Collections

Ellen Knutson, Carole Palmer, Michael Twidale
Graduate School of Library and Information Science, University of Illinois, USA
{eknutson, clpalmer, twidale}@uiuc.edu

Keywords: *Interoperability, Metadata, Digital Collections*

1. Introduction

We are investigating how resource developers can best represent digital collections and items to meet the requirements of divergent service providers and user communities. Our first goal is to establish a baseline that describes the institutions, collections, and initial metadata schemes of Institute of Museum and Library Services (IMLS) National Leadership Grant (NLG) digital collection projects. This research is a component of the larger Digital Collections and Content (DCC) Project, funded by IMLS to build a collection registry and metadata repository for the NLG collections. Collection registries organize large aggregations of digital content from multiple institutions to make relevant resources easier to find and more visible to end-users [1]. As we gain in interoperability, we do not want to lose advances that have been made in adaptation and access for communities of users. Library collection functions that attend to user-based criteria are key to the success of distributed digital collection services [2]. Variations in metadata standards reflect the variant roles and use of digital objects and the different aims and practices of resource developers and their constituent user communities.

In this poster we outline the types of institutions, metadata schemes, and materials contained in the NLG digital collection projects. It should come as no surprise that MARC and Dublin Core are the most often used metadata schemes in these projects, and that academic libraries make up the majority of the institutions involved in creating digital collections. However, we found it interesting that whether or not a project was collaborative, and not the type of material contained in the collection, seemed to influence what metadata scheme was chosen. From the baseline information gathered for this poster we can begin to understand the evolution of metadata issues within and between projects over time to inform the development of useful and usable collection aggregations.

2. Methods

We performed a content analysis of 94 NLG proposals funded from 1998 to 2002. Project web sites were also consulted in cases where the information was not specified in the proposal. The following factors were recorded for each proposed project: type of the institution, type of content, metadata scheme(s) proposed for testing or use, collection description, subject matter, number of items in collection, standard vocabularies proposed, and project personnel. Here we report on the baseline analysis directly related to the metadata schemes specified in the proposals. The results provide the foundation for our ongoing study of NLG metadata use in relation to the needs of user communities, repository applications, and interoperability.

3. Institutions

Of the 94 NLG projects, 53% were collaborative efforts. A total of 227 institutions participated, either as the main institution or a partner. Figure 1 shows the break down of the number of participating institutions by type. Eighty-one academic libraries participated, greatly outnumbering the other types of institutions. The next largest category was museums (44), followed by historical societies (21), and public libraries (17). The "Other" category includes three government agencies, two special libraries, two companies, two herbaria, a zoo, a Native American tribe, a state park, and a national historic site.

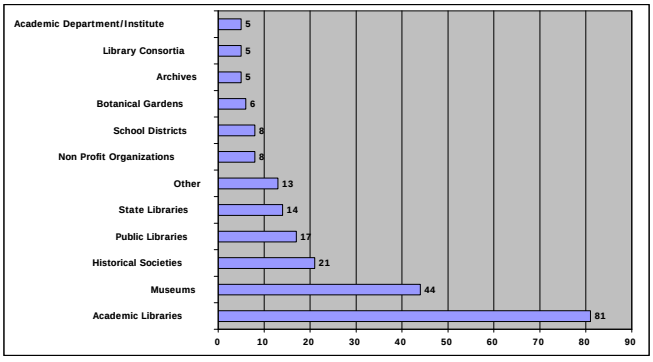

Figure 1. Number of Participating Institutions by Type

4. Metadata Schemes

We were not always able to ascertain what metadata scheme a project proposed to use, if any. Some simply did not include this information in their proposal or on their web site. The break down of schemes is displayed in Figure 2. About 26% of the projects (25) proposed to use multiple schemes. Dublin Core (DC) and MARC were the most common (on their own or in combination), 37% and 39% respectively. The "Other" category includes 2 EAD (Encoded Archival Description) and 3 TEI Header (Text Encoding Initiative). These two schemes were most often used in conjunction with another scheme.

It is interesting to note that institutions working alone were much more likely to select the MARC standard, while

https://doi.org/10.23106/dcmi.952107566

institutions working on a collaborative project were more likely to select DC. Collaboration was not associated with whether or not multiple schemes were used (Table 1).

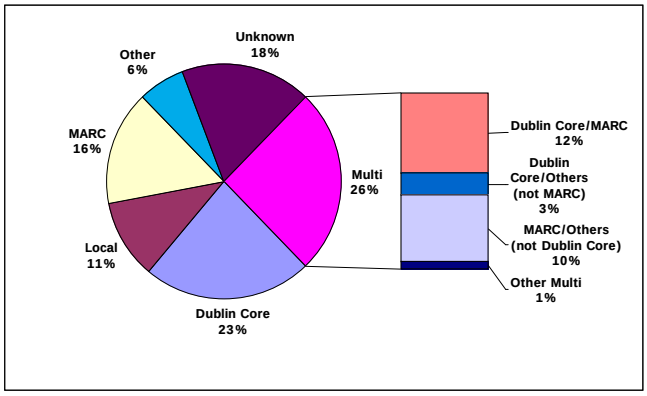

Figure 2. Percentage of Institutions by Metadata Scheme (Note: included in the percentage of institutions using MARC and Dublin Core are two institutions that also specified use of TEI.)

Table 1. Metadata by Collaboration

	Non Collaborative	Collaborative
Dublin Core	3	19
Local	5	5
MARC	13	2
Other	3	3
Unknown	8	9
Multiple Schemes	12	12

5. Collections

As Figure 3 shows, the vast majority of collections contain images. The image category includes reproductions of maps and artifacts.

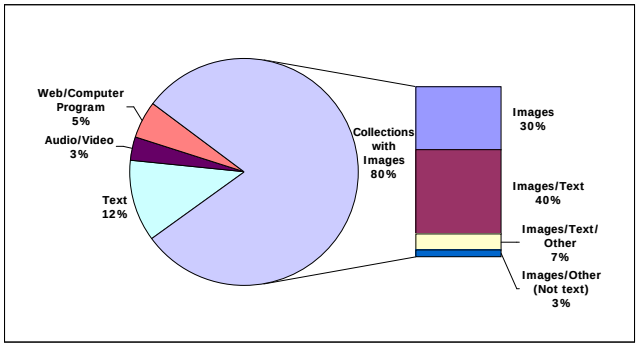

Figure 3. Percentage of Collections by Type of Material

For collections including images, DC is the most popular (Figure 4). MARC is also used frequently, either on its own, combined with DC, or with CIMI (Computer Interchange of Museum Information), VRA Core (Visual Resource Association Core), EAD or TEI Header. Research has shown that DC ranks high in its support for discovery of images [3], but we know less about the effectiveness of MARC for digital images and will be examining this as the study proceeds.

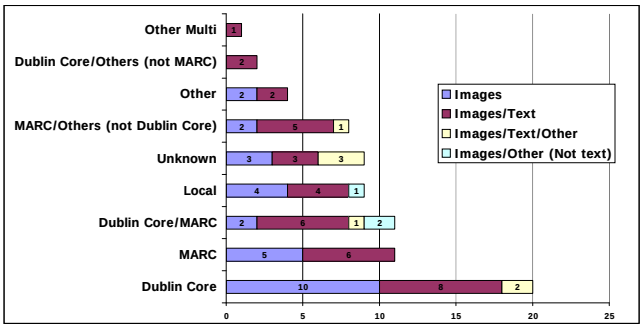

Figure 4. Metadata for Collections Containing Images

Immediate questions that arise from the preliminary analysis of this baseline data include: What are the reasons for the popularity of DC and MARC? What are the factors that influence this choice? Obvious reasons include perceived appropriateness and compatibility with existing collections and prior staff experience and requisite skills with the standard. The DCC project is currently in the process of conducting a survey with the 94 institutions to further develop the baseline analysis. Early results will be presented at the poster session.

6. Further Research

As our research continues, we will be investigating other factors are at play in metadata applications and how they evolve as projects progress and collections are used. Over the next two years, we will conduct interviews and focus groups with a representative group of NLG grant awardees, and will also administer a second large-scale survey in the final year of the project. Over the three-year period we will be tracking resource developers' metadata decisions and applications to answer the following questions: What factors play a role in selection of a metadata scheme? Do these change with experience or over time? If so, why? How does collection complexity—such as multi-type, multi-institution, re-purposing, evolving goals, and federation—impact application of metadata and usability of collections?

References

[1] Miller, P. (2000). Collected wisdom: Some cross-domain issues of collection level description. *D-Lib Magazine*, *6*(9). Retrieved July 21, 2003 from http://www.dlib.org/dlib/september00/miller/09miller .html

[2] Lagoze, C. & Fielding, D. (1998). Defining collections in distributed digital libraries. *D-Lib Magazine*, *4*(11). Retrieved July 21, 2003 from: http://www.dlib.org/dlib/november98/lagoze/11lagoze.html

[3] 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.

https://doi.org/10.23106/dcmi.952107566