Jung-ran Park (Drexel University)
Yuji Tosaka (The College of New Jersey, Ewing)
Caimei Lu (Drexel University)

Locally added homegrown metadata semantics
Issues and implications

Abstract
Using the data collected from a nationwide study drawn from the community of cataloging and metadata professionals, this study aims to assess the current state of locally added metadata elements used across digital repositories. Addition of locally created metadata elements is a fairly common practice adopted by over one-third of the survey participants. Homegrown elements for descriptive metadata comprising local notes and description, local personal and place names, and local subjects are added most frequently, followed by administrative, technical, and preservation metadata. The major reason for extending metadata is to accommodate the perceived needs of local collections and their users. Yet, there are currently few open mechanisms for finding and sharing documented information about such localized metadata practices.

1: Aim and scope of the study
The rapid proliferation of digital repositories by libraries and other organizations has led to a growing awareness of metadata as an essential mechanism for describing and organizing information objects and facilitating effective resource discovery and access in a networked environment. However, there is a lack of a common data model that cataloging and metadata professionals can readily reference as a mediation mechanism during the process of descriptive metadata creation and application of controlled vocabulary schemes for subject description. The development of such a mediation mechanism calls for in-depth research on various issues surrounding the current status of metadata creation practices.

One of the most critical yet relatively unexplored issues that hinder the creation of a common data model is locally added metadata elements that are found in many digital repositories. Heery (2004) points out the increasing number of local extensions and variants to metadata schemes used to accommodate domain-specific requirements. While such local additions and extensions may be useful in providing rich, detailed descriptions for local communities (Perkins 2001; Han et al. in prep.), they may present a roadblock to metadata interoperability, especially when locally defined fields are not mapped to appropriate metadata elements in the aggregated environment. Given the propagation of digital library and metadata projects, it is therefore important to gain a better understanding of the extent of the current use of locally added metadata elements and their potential effects on resource discovery and sharing across digital repositories. Such efforts will make a potential contribution to digital library research at a time when best practices are being actively sought after in the creation and management of quality digital collections.

The goal of this study is to assess the local addition and variation of elements from metadata schemes and examine the grounds and implications of such local addition in digital repositories. Using nationwide survey data mostly drawn from the community of cataloging and metadata professionals, we seek to explore the following study questions:

• What homegrown metadata elements are added in individual digital repositories?
• What are the major criteria used to add locally created homegrown metadata elements?
• How are local metadata practices documented and shared?

In what follows, we will first present an overview of recent studies relating to the use of locally added metadata elements across digital repositories. We will next present the survey method employed to conduct this study and the general characteristics of survey participants. Then, we will report on the way that homegrown metadata elements are utilized across institutions and present data analysis in terms of the prevailing current state of metadata creation practices concerning the use of locally added homegrown metadata elements in digital repositories and collections. The final section will summarize the study and present some suggestions for future studies.

2: Literature review

Despite the critical importance of metadata in a distributed networked environment, many studies have identified an underlying tension between the need to provide good access in the local environment and to create sharable metadata records that work well in aggregated collections. According to the National Information Standards Organization guideline (NISO 2007), “good metadata conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection”. The document then highlights the importance of developing local application profiles that provide specifications of metadata elements and content rules and add local elements to the selected metadata scheme to meet the requirements of local contexts. The guideline also stipulates, however, that “good metadata supports interoperability” based on the use of “standard metadata schemes” that facilitate resource discovery and exchange “in global contexts beyond those in which it was created.” Shreeves et al. (2006) also emphasize that “metadata may be of high quality within its local context, but may be compromised when taken out of this context for various reasons”.

Although many publications, including the NISO document, have offered prescriptive advice and guidelines on how to create quality metadata that supports interoperability, there have been limited investigations into how local metadata elements are added in practice during the metadata creation processes across distributed digital repositories. Park (2005; 2006) conducts an analysis of 659 metadata item records in three digital image collections using Dublin Core (DC). In addition to a variety of inaccurate, incomplete, and inconsistent metadata element uses, she identifies widespread use of locally added elements. Some of the most frequently identified locally added metadata elements are local information such as “contact information”, “ordering information”, and “acquisition”. In addition, Park finds null mapping local elements such as “full text”, “note”, and “scan date”, although they can be mapped onto pertinent DC metadata elements.

A recent study by Han et al. (in prep.) presents locally-defined unique fields created by 21 CONTENTdm-based collections which utilize the Dublin Core metadata scheme. They found that out of 491 fields from 21 collections, 171 (34.8%) fields are locally defined unique fields that are not in Simple or Qualified Dublin Core. The unique fields are categorized into descriptive (74.3%), administrative (19.9%) and technical metadata (5.8%). However, the study shows that 107 (84.3%) out of 127 unique descriptive metadata fields can be mapped onto pertinent DC metadata elements. This indicates that
DC metadata semantics affect the correct application of the DC metadata scheme in local contexts. The results of the study bring forth some strategies for enhancing interoperability across local collections.

While several recent surveys have been conducted to gain a baseline understanding of current metadata practices in digital repositories (University of Houston 2006; Ma 2007; Smith-Yoshimura 2007; Smith-Yoshimura & Cellentani 2007; Markey et al. 2007), none address the question of locally added homegrown metadata elements. This is a rather unfortunate omission considering the potential impacts of locally added metadata elements on resource discovery and access in a networked world. Given the lack of empirical research on the use of locally added metadata elements and their implications in the wider network context, we will examine, in the following sections, how widespread such local metadata practices are and what motivates cataloging and metadata professionals to add locally created homegrown metadata elements during the metadata creation processes.

3: Method

For the study, we conducted a Web survey using including both structured and open-ended questions. The survey was extensively reviewed by members of the advisory board, a group of three experts in the field, and was pilot-tested in advance. The survey included many multiple-response questions that called for respondents to check all answers that applied. Participants were recruited through ten major electronic mailing lists of metadata and cataloging professionals. We also sent out individual invitations and distributed flyers to metadata/cataloguing communities.

During the 62-day period from August 6 to October 6, 2008, a total of 303 completed responses were received by the survey management system. The respondents’ profiles regarding job titles and job responsibilities clearly show that most of the individuals who completed the survey questionnaire engage professionally in activities directly relevant to the research objectives, such as descriptive and subject cataloging, metadata creation and management, authority control, non-print and special material cataloging, electronic resource/digital project management and integrated library system management. Although the largest proportion of participants (44.6%) chose the “Other” category to the job title question, most of their job titles are found to be associated with one of the cataloging and metadata activities. Thus, it is reasonable to assume that the respondents are in an appropriate position to provide first-hand, accurate information about the current state of metadata creation in their institutions.

Concerning the institutional background of participants, less than half of the survey respondents (39.9%) provided institutional information. We believe that this is mostly due to the fact that the question was optional, following a suggestion from the Institutional Review Board at Drexel University. Of those who provided institutional information, 75.2% are from academic libraries, followed by public libraries (17.4%) and other institutions (7.4%). In terms of work experiences, more than half of the respondents (58%) reported over 5 years of experience in cataloging and metadata creation: 6 to 15 years (31.6%) and over 16 years (26.4%). Approximately one-third of the respondents (34.5%) reported 1 to 5 years of experience, while the rest (8.1%) reported less than a year of professional experience.
4: Results
In this section, we present our findings in the following three areas: (1) the extent of digital repositories adding locally created homegrown metadata elements and the type of local extensions and variants to the selected metadata scheme(s); (2) the reasons for adding homegrown data elements and values in the local workflow for metadata creation; and (3) the current documentation status to accommodate such homegrown metadata practices.

4.1: The current status of homegrown metadata element use
Addition of locally created metadata elements is a fairly common practice among digital repositories. Over one-third of the survey participants (n=108, 38.0%) add homegrown metadata elements to the selected metadata scheme(s), while slightly less than half (n=134, 47.2%) indicate otherwise. In terms of metadata schemes, there are a few that are widely used among the survey participants’ institutions. The most frequently mentioned scheme is MARC (Machine Readable Cataloging). The Dublin Core (DC) metadata schema is the second most widely used, followed by EAD (Encoded Archival Description), MODS (Metadata Object Description Schema), VRA (Visual Resource Association) and Core and TEI (Text Encoding Initiative).

To examine the types of homegrown metadata elements, we asked the participants to specify what homegrown metadata elements are added in their digital repositories. One-third of the survey respondents (n=91, 30%) using local extensions and variants to the selected metadata scheme(s) provide information to this question.

For the purpose of the study, we classify their descriptions into the following three categories based on the most prominent features and common characteristics described in their answers. Descriptive metadata is further subcategorized as shown below:

• Descriptive metadata
  - Local subjects
  - Local classification
  - Local names for person and location
  - Local notes and description
  - Genre and type terms

• Administrative, technical, and preservation metadata

• Structural metadata

Out of 133 locally added metadata elements, analysis of the responses reveal that descriptive metadata (n=78, 58.7%) are the most frequently added metadata elements to a given metadata scheme. Subcategories of descriptive metadata composed of local subjects (n=19, 14.3%) and classification (n=8, 6.0%) and local names (person and place) (n=22, 22.1%) are the most frequently added metadata elements to a given metadata scheme. Local notes and description (n=23, 23.1%) including thesis and dissertation related notes are also a frequently mentioned subcategory. The subcategory of Local notes and descriptions contains various types of home grown local elements encompassing information on physical description, archival information relevant to the institution, electronic thesis/dissertation (e.g., discipline, advisor), funding support and gifts. In addition, genre and type vocabularies (n=6, 4.51%), a subcategory of descriptive metadata, especially for non-textual resources, are also locally added to a given metadata scheme.
As shown, local subjects, classification, names and descriptive notes are assigned in
unique, locally created elements rather than in element(s) already provided in the
selected metadata scheme(s). However, as discussed in the related studies (Park 2006;
Han et al. in prep.), most of the above descriptive elements are questionable to consider
as truly homegrown local metadata elements in the sense that they can be replaced with
relevant metadata elements (e.g., description, subject) drawn from metadata schemes.

Analysis also shows that administrative metadata (e.g., metadata record creator, date
of record creation) \( (n=42, \ 31.57\%) \) encompassing technical metadata (e.g., format
characteristics, digitization specification) \( (n=6, \ 4.51\%) \), rights (e.g., access restriction,
copyright) and preservation metadata (e.g., conservation information, resource
Permanence guarantor) \( (n=2, \ 1.50\%) \) are the most frequently added homegrown local
metadata elements. Structural metadata (e.g., browsing and navigation of web pages)
\( (n=5, \ 3.75\%) \) are also observed in the participants’ descriptions of locally added home
grown metadata elements.

4.2: Criteria for adding homegrown metadata elements
What are the factors that have shaped the current state of homegrown metadata element
use reported thus far? In this section, we turn our attention to major reasons and
constraints that led many institutions to localize metadata in their digital collections.

The multiple responses \( (n=259) \) to the survey question (i.e., “Why do you add locally
created metadata elements? [please check all that apply]”) indicate that the primary
grounds for adding homegrown metadata elements can be found in the motivation
behind the nature of local collections and resources \( (n=83, \ 32.0\%) \) together with the
characteristics of the target audience/community of the local collection \( (n=63, \ 24.3\%) \)
and the local practices of cataloging and metadata creation \( (n=37, \ 14.3\%) \) and local
system configuration \( (n=20, \ 7.7\%) \). It is noteworthy that more than one-tenth of the
responses \( (n=37, \ 14.3\%) \) present that it is necessary to add homegrown metadata
elements owing to the lack of metadata elements/descriptions for local collections and
resources from the selected metadata scheme(s). As stated earlier, the majority of the
survey participants \( (n=170, \ 58\%) \) possess over 5-year experiences in engaging with the
core activities of cataloging and metadata creation and management. This is reflected in
the fact that only 3 responses \( (1.2\%) \) indicate that local addition is due to their lack of
full understanding of given metadata schemes.

The survey results above confirm the “inward focus” in current metadata creation
practices as reported in the OCLC RLG survey (Smith-Yoshimura 2007). It is evident
that libraries still tend to define their locally affiliated users as the primary audience for
their digital collections. As Hillmann (2008) points out, however, this inward focus of
adding locally created metadata elements may no longer make sense in a networked
world where users increasingly rely on general search engines such as Google as their
primary mechanism for resource discovery. Then, the question is if all the current costs
of customizing local metadata “within the context of a local catalog and local
collection” are really justified – especially given its potential risk to achieving an
optimum level of interoperability – and if the focus should shift in the future to
standards-based metadata creation optimized for shareability and external discovery.
4.3: The current status of homegrown metadata element documentation practices
As discussed above, over one-third of the survey participants \( (n=108, 38.0\%) \) are using locally added homegrown metadata elements largely to accommodate local, domain-specific requirements. The survey result further suggests that local extensions and variants to the selected metadata scheme(s) are allowed in local metadata guidelines of approximately two-thirds of participant institutions \( (n=98, 69.5\%) \) out of 141 responses. The balance of the respondents reported either no allowance \( (n=21, 14.9\%) \) or do not know \( (n=22, 15.6\%) \).

The high percentage of local documentation/guidelines allowing for homegrown metadata elements foregrounds some issues related to metadata interoperability and resource discovery. The large number of such metadata guidelines evidence a significant and widespread effort spent in assuring local flexibility for adding homegrown metadata elements. It is this flexibility that enables libraries to make adjustments and modifications to the selected metadata scheme(s) corresponding to local needs. Such locally added metadata has the potential to facilitate the information needs of users in local environments. This flexibility, however, may also engender hindrances in achieving interoperability and resource discovery across digital repositories.

This concern about metadata interoperability seems to be justified by the fact that these locally defined element sets typically are not documented on the Web. Heery (2004) points out that local metadata application profiles must be documented and made accessible at least in human-readable form in order to encourage the reuse of existing metadata beyond the immediate local environment. It is noteworthy that only about one-fifth of the respondents \( (n=28, 19.6\%) \) to the survey question state that their organizations make the documentation/guidelines for metadata creation available to the public on their Web sites. The current level of difficulty in finding information about homegrown metadata elements is particularly alarming because metadata sharing can be well supported when a common data model is established to map between various metadata schemes and overlapping semantics and incorporate different homegrown metadata elements in a distributed digital environment.

5: Conclusion
In this paper, we explored the current status of metadata creation practices concerning the use of locally added homegrown metadata elements through an examination of survey responses drawn mostly from cataloging and metadata professionals. The survey results indicate that the widespread use of homegrown metadata elements may present a potential challenge to the effective reuse and sharing of metadata in the networked environment.

While locally added metadata elements have the potential to hinder a maximum level of metadata consistency and interoperability, nearly 40% of the survey participants extend the metadata they create by adding a variety of localized metadata elements. This practice is adopted for several reasons, but apparently the major motivation is to accommodate the perceived needs of local collections and their users, as indicated by the two most common responses: “to reflect the nature of local collections/resources” and “to reflect the characteristics of target audience/community of local collections.” Local conditions are also cited (though less frequently) from institutional and technical
standpoints. Many institutions continue to follow existing local practices for cataloging and metadata creation, while other institutions are making local metadata additions due to constraints by their local systems.

The high occurrence of locally added administrative metadata elements encompassing technical, rights and preservation metadata indicates that there is a need for metadata schemes to provide a more robust supporting mechanism and guidance for implementing administrative metadata elements in local contexts. The study results also indicate that domain-specific controlled vocabularies such as genre and type terms for non-textual resources need to be further developed. The results of the present study also highlight the important role of application profiles that facilitate the meeting of local needs through adoption of other sources of metadata elements from various metadata schemes. Methods for facilitating the sharing of such local application profiles in turn would lead to increased metadata interoperability across digital collections. In addition, further development of a common model/mechanism for facilitating semantic mappings between metadata schemes is also needed to address the need to extend metadata and promote interoperability.

A constraining factor of this study lies in its survey population. For future studies, incorporation of other research methods such as follow-up telephone surveys and face-to-face in-depth interviews would be valuable in understanding the current status of homegrown metadata element usage. A longitudinal study also would provide needed information on changing patterns of homegrown metadata usage in a rapidly digital environment. Future research with more varied data sources will enable us to gain a fuller picture of the widespread practice of adding local extensions and variants in a networked world.

Acknowledgments
This study is supported through a research award from the Institute of Museum and Library Services. We would like to express our appreciation to the anonymous reviewers for their invaluable comments.

This study is a part of a larger project investigating issues surrounding the metadata creation process, the employment of controlled vocabulary schemes, metadata quality control measures and the new competencies and skill sets demanded of cataloging and metadata professionals in the digital era. (General metadata creation practices, semi-automatic metadata application, metadata semantics and quality control, and the professional development of cataloging and metadata professionals are reported on in-depth in separate studies.)

References


*Web documents have been accessed 22 May 2009.*