World Library and Information Congress: 70th IFLA General Conference and Council

22-27 August 2004
Buenos Aires, Argentina

Enabling Metadata: Creating Core Records for Resource Discovery

Lynne C. Howarth
Faculty of Information Studies
University of Toronto
Toronto, CANADA
howarth@fis.utoronto.ca

Abstract: *The proliferation in electronic resources over the past decade has been accompanied by intense activity in the development and application of metadata schemas. Recognizing that growth, and understanding the need for some direction in the selection and use of metadata standards, the IFLA Cataloguing Section Working Group on the Use of Metadata Schemas drafted a report for Worldwide Review (2003-04), inviting comment on recommendations deriving from three objectives that formed the Working Group's terms of reference. This paper discusses responses to the review and, based on that feedback, offers suggestions for some possible next-steps. The need for a "core record" framework requires further exploration – as a standard (?), as a basic set of elements for legal deposit (?), as the raison d'être for a generation of software tools that can extract and/or generate an essential metadata record (?).*

1.0   Introduction

Along with the voluminous output of digital information that has accompanied the maturation of access to the Internet and Web authoring technologies, there has been vigorous activity towards developing and refining metadata schemas. Whether defining "new uses" for existing standards – for example, assigning the Dewey Decimal Classification for organizing and accessing electronic resources through a subject gateway such as Renardus – or creating entirely new

schemas, such as Dublin Core, or the Online Information Exchange (ONIX) standard for publishing, the pace of growth in metadata applications and project implementations has been remarkable. In recognition of the increasing importance of metadata – defined broadly as "data about data" – and the need to both monitor and influence the direction of schema design, interpretation, and application, the IFLA Cataloguing Section Working Group on the Use of Metadata Schemas was established at the IFLA 1998 Conference in Amsterdam, the Netherlands. The Working Group drafted its terms of reference agreeing to focus on outcomes deriving from the following three objectives:

- *Objective 1*: to create an inventory of the development and implementation/application of metadata schemas in different countries
- *Objective 2:* to provide guidance (and ultimately, as appropriate, guidelines) to libraries as to when and/or how best to use metadata records and bibliographic records (catalogue records)
- *Objective 3:* to determine a metadata "core record" – i.e., a set of most commonly occurring elements in selected metadata schemas – that could be used by authors and/or publishers of electronic records to enhance resource discovery, and to provide, where appropriate, elements for incorporation into bibliographic records (catalogue records)

In fulfilling its terms of reference, the Working Group held five sets of formal meetings, beginning in Bangkok, Thailand at the IFLA 1999 Conference, and continuing at the IFLA 2000 Conference in Jerusalem, Israel, the IFLA 2001 Conference in Boston, USA, the IFLA 2002 Conference in Glasgow, Scotland, and the IFLA 2003 Conference in Berlin, Germany. Over the five-year period, the Working Group concentrated on completing its three objectives, co-sponsored with the IFLA Information Technology Section an IFLA program (2000) and an IFLA Discussion Group (2000-2003), and engaged in discussions with the DCMI Libraries Working Group (2001-). These various efforts culminated in a draft report, *Guidance on the Structure, Content, and Application of Metadata Records for Digital Resources and Collections*. With a goal of soliciting extensive international feedback, the report was posted to the IFLA Website for worldwide review for three months across the November 2003 to February 2004 period. It should be noted that the Working Group judged the metadata applications landscape too fluid to suggest anything other than "guidance" at this particular time, noting that "guidelines" might be appropriate at a future date.

The draft report focused on two key areas, namely, (1) using metadata in libraries or other information-intensive organizations, and (2) determining a metadata core record framework. A total of sixteen responses were received from nine countries, including, Australia (1), Canada (1), Denmark (1), France (2), Germany (1), New Zealand (1), Sweden (1), the United Kingdom (2), and the United States (6). While the total included three responses from individuals, the majority came from national cataloguing agencies, professional association cataloguing committees, committees responsible for a particular metadata schema, and academic libraries. Thus, comments represent the collocation of input and expertise from many. In general, feedback was extensive and detailed; the Working Group extends its sincere gratitude to respondents for their carefully crafted and thoughtful input.

This paper broadly summarizes responses to each of the three objectives that the Working Group undertook within its terms of reference, and on which it reported in its draft for Worldwide

Review. Based on that feedback possible next-steps and points requiring further reflection are outlined. It should be noted that specific revisions, additions, or recommendations have been incorporated into the final report delivered to the IFLA Cataloguing Section at IFLA 2004 in Buenos Aires, Argentina.

2.0   Summary of Responses to Worldwide Review

2.1 Comments Regarding the Draft Report in General
Respondents for the most part praised the draft report for its pragmatism and level of detail, while also noting that, because of the dynamic nature of metadata standards development, there was information that was either out-of-date (e.g., CIMI was recently wound up) or incomplete (e.g., mention could be made of the Open URL standard, or the Metadata Encoding and Transmission Standard [METS], or the Metadata Object Description Schema [MODS]), or the IFLA Draft Statement of International Cataloguing Principles). It was suggested that the report more carefully define the term, "metadata", and rethink the scope of the document to emphasize and promote, "…the importance of using appropriate existing or emerging standards, and recognised schema, wherever possible." Several respondents noted that the role of data conversion between metadata schemas was inadequately addressed, and stressed the need to highlight the value of consistency for sharing resources across projects, communities, portals, repositories, etc. While the Working Group's intention was to solicit critical assessment of the document in order to identify required revisions, and to inform subsequent actions, several respondents commented on the desirability of having the final report widely accessible to the international library community and beyond.

2.2 Comments Regarding Each of the Three Working Group Objectives

Working Group Terms of Reference: Objective 1: *To create an inventory of the development and implementation/application of metadata schemas in different countries.*

When the Working Group was initially constituted in 1998, the number of projects employing emerging metadata schemas, such as Dublin Core, was relatively small and amenable to monitoring. By 2001, the situation had changed so substantially as to render a meaningful inventory of metadata-related projects unrealistic. Moreover, it was felt that, with the emergence of metadata registries for identifying and tracking implementations and adaptations of different metadata schemas, there were formal and well-maintained sources for the international bibliographic community to access. The Working Group therefore proposed the following recommendations for Worldwide Review.
- The Working Group encourages the IFLA web content manager to continue work on the metadata site to include and maintain links to key information sources and registries for those metadata schemas most widely used within the bibliographic community.
- The Working Group further suggests that the schemas referenced in this report and used as a basis for the common core of metadata elements, constitute the minimum set of metadata standards to which the IFLA metadata site points.

While respondents concurred with the first recommendation, and commented on the overall usefulness of the IFLA metadata site, they also noted the potential challenges of maintaining

current links while also identifying new sources in such a volatile field. This seemed a laudatory but highly problematic task for one individual to manage along with other ongoing responsibilities. An alternate, perhaps more feasible approach was proposed by an individual who suggested, "You need to use your impetus towards your first objective to encourage individual librarians and libraries to publish the results of their small projects, including their procedures and best practices. This is what is currently missing from the literature." Likewise, a national cataloguing committee proposed that the IFLA "Digital Libraries: Metadata Resources" website, "… be kept by regional IFLA contributors appointed by IFLA to monitor major projects in their respective regions." Regarding the second recommendation, there was mixed support, with some expressing reservations to the idea of a "common core" similar to those summarized under Objective 3, below.

> ***Working Group Terms of Reference: Objective 2:*** *To provide guidance (and ultimately, as appropriate, guidelines) to libraries as to when and/or how best to use metadata records and bibliographic records (catalogue records).*

Many respondents agreed with the Working Group's premise that, given the dynamic nature of the current metadata environment, it is premature to assign guidelines as to when to create either a catalogue record or a metadata record. While one cataloguing committee observed that it may *never* be appropriate to provide guidelines, per se, it did suggest that, "It is a very common scenario for metadata records to be created as surrogates for the digital objects within a digital collection, and then for a collection-level catalogue record to be created as a surrogate for the entire collection and as a means of transporting the online catalog user from the catalog to a digital collection. It appears to us that this model is well established and that it would not be premature to mention this scenario by way of illustration". Other respondents also emphasized the need for more examples to serve as possible models. As one noted, **"**The introduction to initiating a metadata project is very helpful. You should continue to publish it and gather the introductions others have written in an attempt to disseminate them to as wide an audience as possible of the growing number of information professionals considering digital library projects." Finally, a national agency suggested that Objective 2 be rephrased as, "to provide guidance to libraries on evaluating and selecting a metadata schema or schemas".

In addition to suggesting the inclusion of some "best practice" scenarios, respondents identified a number of topics that they considered essential to, but missing from, this part of the draft report. Some noted that a discussion of metadata that supports the preservation process warranted its own separate section, as does more explicit reference to control of data content and the need for disambiguation of terminology based on the context in which the database of resources is being created. A national agency suggested that, "a crucial category of metadata is totally missing: *structural metadata*". This group of respondents noted that five distinct levels should be included as follows: (1) collection level (set); (2) aggregates (objects gathered by type and by responsibility for preservation; (3) primary objects; (4) intermediate objects (a view of the object); and (5) terminal objects (digital content files).

Several respondents took issue with the distinction made between metadata records and so-called "traditional cataloguing". As one national agency offered, "… thinking in terms of 'metadata versus bibliographic records' creates a dichotomy that is unhelpful, and that it is more useful to think about metadata as a continuum. Instead of a library taking the approach that it has a digital

collection that requires metadata records, it may be more constructive to take the approach that the library has resources that require descriptions. This will help avoid the false perception that digital resources require metadata records, whereas other resources require bibliographic records."

A cataloguing committee suggested that the distinction was not useful, proposing that Objective 2 be restated in this regard as: "To provide guidance to the library community on the nature and implementation of metadata schemes, including bibliographic or catalogue records, so that a project can evaluate and select which scheme will best match the goals of the project." This was echoed in a concern that was articulated as follows:

> ...the guidelines treat all options as equal, and do not actively promote the benefits of using, or building on, standards-based metadata and existing schema where possible. Given the context of the objective 2, which was to provide guidance on when to use metadata records versus when to use catalogue records, it might be worthwhile to limit the scope of the guidelines to shareable metadata. In this context, the creation of local schema should be employed only when available schemas have been rejected as they demonstrably do not meet organisational needs.

Several respondents commented on the practicality of this section, noting, for example, that, "… the checklist of questions to ask when selecting a schemas is potentially useful". An academic library metadata specialist observed that: **"**The introduction to initiating a metadata project is very helpful. You should continue to publish it and gather the introductions others have written in an attempt to disseminate them to as wide an audience as possible of the growing number of information professionals considering digital library projects." This was echoed by a national library, adding, "We would recommend this section becomes available as a stand-alone document". Notwithstanding the revisions and additions that may be required to enhance this part of the draft document, there appears to be support for Objective 2 and what may result in the final report from the Working Group.

> **Working Group Terms of Reference: Objective 3:** *To determine a metadata "core record" – i.e., a set of most commonly occurring elements in selected metadata schemas – that could be used by authors and/or publishers of electronic records to enhance resource discovery, and to provide, where appropriate, elements for incorporation into bibliographic records (catalogue records).*

Based on responses, it is fair to conclude that the draft report failed to articulate clearly the intention behind the proposal for a metadata core record framework. In many cases respondents interpreted the "core of cores" (Chan) as a "pseudo Dublin Core", albeit with fewer elements. One national library commented that, "Defining this 'core of cores' element set appears to have entailed undertaking a similar task to that undertaken by the Dublin Core community several years ago…. Interestingly, and perhaps reassuringly, the conclusion reached is not dissimilar to that reached by DCMI; with only one exception the recommended elements are a subset of the original fifteen DC elements." A similar assessment offered the following: "It is hardly surprising to observe that the result of the search for a minimum metadata set is very close to the Dublin Core fifteen basic elements (DCMES). This conclusion matches with the fact that

DCMES is a de facto standard[1] for many metasearch interfaces that must deal with a large scale of formats. As a consequence, it seems useless to establish a minimum metadata set because it already exists and libraries are among the most active users of this reference tool." One national cataloguing committee endorsed the concept of the "core of cores", " … as a framework with the potential to bring about a common, minimal standard and extended interoperability between metadata produced in different projects and by different agents."

Other comments focused on recasting Objective 3 and its proposed "core of cores" altogether. As one national cataloguing committee noted, "Since librarians are already active participants in the ongoing development of many of the metadata schemas noted in this section, the best way to accomplish the aims of Objective 3 would be through the development of a "library application profile"[2] *specific to* any metadata schema that a library might choose to employ, rather than a general across-the-board set of elements meant to apply to any metadata schema." In a similar vein, a national agency offered the following:

> There is some confusion as to what was intended for the "core of cores"-- it seems that it was starting out to describe a core set of elements that should be part of any metadata schema that was designed (a good idea); but, it sometimes implies that these are core data elements that should be in every metadata record (i.e., every metadata record should have something from each of these core elements)-- this would be problematic in that not all elements would be appropriate for every record. We suggest the emphasis should be more clearly placed on core elements that should be a part of all metadata schemas, and remove references to a core "*record"* per se.

Finally, one group of respondents recommended, constructively, that the Working Group re-evaluate and re-formulate its third objective, "… to satisfy themselves that there is a definite and distinctive role for the metadata core framework, given that schemas which are also intended to represent 'core" elements already exist and have been widely adopted."

Similarly strong opinions were voiced in response to the Working Group's recommendation, under Objective 3, that ***national libraries and other bibliographic or cataloguing agencies consider adopting the ten element common core framework as a standard metadata record structure for organizing digital resources, including those submitted to national library collections under legal deposit.***   Several respondents felt that such a proposal was premature. Others observed that the ten-element set could only serve as a starting-point to associate with a legal deposit because of the complexity and constraints of the legal context *per se*. Finally, it was proposed that, prior to finalizing any recommendations, a separate survey be undertaken to determine how national libraries handle the legal deposit of electronic documents.

In addition to interpretations of the nature and intent of the ten-element core record framework that differed from the intent of the Working Group, there were a number of proposed revisions or changes to individual elements, their content or definition. Some respondents suggested alternate or additional elements.   As the Working Group itself experienced, the proposed core element,

---

[1] And now an ISO standard.
[2] This terminology comes from the Dublin Core Metadata initiative, where such a profile has been developed.

"Resource Type" was especially problematic with its division into two facets (type of content; type of carrier). It was noted that some core elements, such as Language, will not apply for all material types. Some respondents observed that, while the scope of "Date" is articulated ("The date(s) could describe when the work was created, published, modified, accessed, etc."), there are no specific proposals as to how this should be done. Several commented on the potential challenges of creating, qualifying, and maintaining a *standards-compliant* framework comprised of ten elements, themselves vulnerable to differences in interpretation.

3.0 Conclusions and Next-Steps

The Worldwide Review of the draft document, *Guidance on the Structure, Content, and Application of Metadata Records for Digital Resources and Collections,* has provided the members of the IFLA Cataloguing Section Working Group on the Use of Metadata Schemas with a wealth of data, and a diversity of perspectives. The sample of opinions, quoted as applicable, offers only a glimpse of the extensive feedback with which the Working Group was able to craft its final report. Giving credence to the expression, "what you intend is not always what others hear in what you say", the Working Group devoted considerable attention to articulating for itself the nature, purpose, and ultimate utility of the proposed "core of cores".

The title of this paper implicitly poses the question of what metadata is required to facilitate (or enable) resource discovery. What metadata is essential or "core" to providing access to a digital object? This was the question that the Working Group undertook to explore as part of its terms of reference; it remains an important question in the present environment where the concept of sharable metadata (and issues of interoperability) has assumed a priority for libraries and other information-intensive cultural institutions. While respondents to the Worldwide Review have seriously questioned the purpose and ultimate need for the proposed "core of cores" framework, the Working Group needed to do the same, ensuring that, if there is a clear role, it requires careful articulation and an obvious application (a set for legal deposit?). What should a "core record for resource discovery" contain? How should that data be created or derived or exchanged (shared)? Once these questions have been satisfactorily addressed, then issues related to standards development and maintenance, to automatic extraction and/or generation of appropriate metadata, and to cross-schema interchange (crosswalks; interoperability; etc.) will have a proper context, some common ground, and an obvious raison d'être.