Finding Museum Information in the Internet Commons: a report on the CIMI Dublin Core Metadata Testbed Project

John Perkins, Consortium for the Computer Interchange of Museum Information (CIMI), Canada and Angela T. Spinazze, ATSPIN Consulting, USA

Abstract

Launched in May 1998, the CIMI Dublin Core (DC) Metadata Testbed is an initiative that is evaluating the Dublin Core Metadata standard for resource discovery for museum information. Testbed participants created a test database of over 200,000 records representing an array of collections from around the globe and drafted a guide to best practice for DC use in museums. This paper will report on the results of the Testbed's second phase which focuses on completion of the guide to best practice; the identification and recommendation of DC qualifier elements for enhanced descriptive capabilities; the examination of Resource Description Framework (RDF) as an effective method for enabling interoperability between applications that exchange metadata; and an initial effort in mapping DC elements to the CIMI Access Points as a step towards providing effective resource discovery.

Needles in a Haystack

The World Wide Web is now a wonderfully chaotic place. It has millions of Web sites and nearly half a billion pages, pictures, audio and moving images making it increasingly difficult to find and access information. Searches of Web indexes like AltaVista or Yahoo! usually turn up overwhelmingly large results or very general information in unpredictable order and often with questionable relevance. People looking for authoritative information have a hard time telling the gems from the irrelevant. The Web is crying out for a method of cataloguing the information, known as Internet resources, it contains.

Enter the metadata movement. Metadata is a broad, if not opaque, term for various types of information such as library catalogs, directories, subject indexes, and even content ratings or reviews. But metadata is also more than this. As Thomas Baker points out in his overview of current metadata work:

> ... researchers are now extending or reinventing library and information science for the technical and functional requirements of entirely new types of digital contents and services. Online publication is replacing the printed journal literature in some fields of science. Libraries and museums are using scanners to improve access to rare or brittle materials from our cultural heritage. Bookstores and mail-order outlets are taking orders directly from their online catalogs. In all such situations, metadata helps people find what they need, verify its authenticity, process it in an appropriate format, and perhaps to order or pay for it over the Web. (http://www.cs.ait.ac.th/~tbaker/Kyunghee.html)

Metadata is clearly going to be very important in the evolution of our ability to make cultural heritage information widely and easily available for education, research, scholarship, commerce and enjoyment.

Organizing the Web: Metadata for Discovery

For the past three years the *Dublin Core Metadata Initiative* (DCMI) has developed a simple set of 15 elements for describing Internet resources. These 15 elements, known as the *Dublin Core (DC)*, have been developed by a diverse international group of libraries, museums, government information programs, and commercial publishers in collaboration with the World Wide Web Consortium. The hope for the DC is that it will support the discovery, retrieval and use of the digital information resources from a vast number of sources, much in the same way an online library catalogue helps us find books and journals in the libraries.

At the same time the intellectual construction of DC has been taking place software and tools have been created and an enthusiasm for the potential utility of the DC is rapidly increasing. Software companies like Microsoft (in Internet Explorer 5.0) and organizations like The New York Times, CNN, Reuters and Amazon.com all are investing in the use of metadata and the emerging standards for expressing it (Baker, 1999).

CIMI has spent the last 18 months working on an application of the Dublin Core metadata standard specifically for use in museums to help address the problems just described. The *CIMI DC Testbed* is a multi-national, multi-institutional effort involving eighteen participants in eight countries. Testbed participants include museums, systems developers, national information organizations and cross disciplinary museum studies projects. These institutions include: ADLIB Information Systems, American Museum of Natural History, Australian Museums On-Line (AMOL), Canadian Heritage Information Network (CHIN), Crossnet Systems Ltd. Distributed Systems Technology Centre (DSTC) Pty Ltd, Gallery Systems, Integrated Arts Information Access Project Walker Art Center and The Minneapolis Institute of the Arts, International Institute for Electronic Library Research/deMontfort University, mda, The National Museum of Denmark, The Natural History Museum (London), Solomon R. Guggenheim Museum, Taiwan Digital Museum Project, UK Office for Library and Information Networking (UKOLN), Willoughby Associates Limited/Intermuse Division.

CIMI Dublin Core Testbed: sharing the experience

Our study of the Dublin Core standard started with DC version 1.0, as articulated in RFC 2413 (otherwise known as DC Simple). We reviewed each of the 15 DC elements and associated definitions, mapped those elements to existing element sets currently in use at each of the participant institutions, built a test database of Dublin Core records and then examined the placement of values within elements to determine how effective they might be in aiding resource discovery. What we found was that the element definitions and suggested usage didn't speak directly to museum needs. Therefore, clarification of the elements, from a museum perspective, was required.

The result is the *CIMI Guide to Best Practice: Dublin Core*. Designed to assist museums and cultural heritage organizations with interpreting and implementing Dublin Core, the Guide was published in July 1999. It brings together real world experiences from Testbed participants and offers a museum-centric approach to using the Dublin Core standard. The Guide is available in Word and PDF formats from the CIMI web site (http://www.cimi.org).

The Current Agenda

The conclusions drawn from our study of DC Simple provided the approach for continuing the Testbed into its second phase. Beginning with the premise that a DC Simple record is effective in cross-domain searching but, lacks the richness of information that we want to see in a metadata record, we realized that we needed to move beyond DC Simple and examine the concept of qualification. While qualifiers have been actively discussed and pursued throughout the Dublin Core community for quite some time, it wasn't until earlier this year that a structured approach was proposed for how to do it.

In addition to the question of which qualifiers for which elements, there are the questions of how to encode the metadata records as well as how they will be recognized and exchanged in a networked environment such as the World Wide Web. Looking beyond the present use of HTML as an encoding scheme, we recognized that the Resource Description Framework (RDF) (http://www.w3.org/TR/REC-rdf-syntax/), the World Wide Web Consortium approved standard for exchange of metadata on the Web, and Extensible Markup Language (XML) (http://www.w3.org/XML/), a more streamlined version of SGML, are expected to become the format and syntax of choice for encoding metadata and because of this we needed to include them in this year's agenda of work.

We began our discussion about Dublin Core qualifiers with a review of the Dublin Core Data Model Working Group's recommendation http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/) for how to express the structure of Dublin Core within RDF. What we quickly discovered was that the structured approach for expressing element and value qualifiers would be a very useful tool for our review. We set out to examine each Dublin Core element and the need to qualify it both at the element and value level. As part of this review process we compiled lists of suggested qualifiers (many taken from our previous discussions involving DC Simple) or those currently in use by other organizations within and outside of the cultural heritage community (such as the Gateway to Educational Mate-

© Archives & Museum Informatics, 1999

rials (GEM) project, Taiwan Digital Museum Project, the Art Museum Image Consorium (AMICO), etc.). The lists continue to grow as we examine each element in more detail. To date, we have worked through six of the fifteen core elements documenting the purpose for qualification, a CIMI list of suggested qualifiers, schema where appropriate, and our observations about what types of questions qualification may help to answer.

Early on in our discussions about how we would approach element qualifiers, we decided to assume that qualification was not required. Our challenge, therefore, has been to provide a rationale, or purpose, for qualification of any element. As we moved on to define what we wanted to learn from our experience exploring DC element qualifiers, we considered the set of CIMI Access Points first documented in 1995 for a previous CIMI project. The CIMI Access Points were defined as the result of an analysis of the information categories that could be extrapolated from questions users ask museums and the proposed answers museums might give to those questions. We discovered that our work with DC elements had an affinity with this set of attributes. We too are interested in knowing what categories of information users are looking for and, which fields (or in this case elements) might correspond to those categories. Also, we need to know if museums have nformation to answer the questions posed to them by their users. Knowing the categories of information being sought we are in a better position to evaluate if the existing content will adequately answer the question.

The attribute set and the sample questions outlined in the CIMI Access Points document have proved a useful resource in our discussions especially as a means for differentiating the type, quality and quantity of information we would expect to find in a DC Simple and DC Qualified record. For example, it becomes clear that only part of the question, *What Inscriptions are there on coins from the reign of Hadrian?* can be answered by a DC simple metadata record. This underscores the importance of being able to answer that question through the availability of a DC qualified record.

In addition to using the CIMI Access Points, we have reconstructed the Test database with a new set of Dublin Core records based on the Guide to Best Practice. The database too has been instrumental in our examination of element qualifiers as it provides us with a means to test our assumptions against real museum information. The database continues to evolve as we move through the core element set.

Our work continues through the end of 1999 when we will publish a recommendation for museum implementation of Dublin Core qualified and share our experiences working with the RDF and XML. For more information, visit the CIMI web site at http://www.cimi.org or contact Angela Spinazze, CIMI DC Testbed Project Manager at ats@atspin.com.

© Archives & Museum Informatics, 1999