

# Crosswalking IPL Metadata to Dublin Core

Michael Galloway, Michael Khoo, Xia Lin, Jung-ran Park  
Drexel University  
3141 Chestnut St.,  
Philadelphia PA 19128  
+1.215.895.2474

{mike.galloway, michael.khoo, xia.lin, jung-ran.park}@ischool.drexel.edu

## ABSTRACT

The Internet Public Library (IPL) is crosswalking its metadata to Dublin Core compliant metadata. This involves several research activities, including: quantitative and qualitative analyses of the existing IPL metadata; the creation of a new IPL metadata schema as an application profile of Dublin core; the development of a new database structure; and the development and testing of a new metadata creation and maintenance interface. This poster describes and provides illustrations of these activities.

## 1. THE INTERNET PUBLIC LIBRARY

The Internet Public Library (IPL: <http://www.ipl.org>) was created in 1995 in a library and information science class taught by Professor Joe Janes at the University of Michigan-Ann Arbor, to see what “librarianship had to say to the networked environment and vice-versa” [9]. Its success was such that the class decided to keep the project going. More than a decade later, the IPL supports and enhances library services through the provision of authoritative collections, information assistance, and information instruction for the public [8]. It is available throughout the United States as a training tool for library and information science (LIS) programs, and has trained over 5,000 students in 17 LIS programs in tasks such as the creation and editing of metadata [7].

The IPL has subject-categorized collections of more than 40,000 online resources. These collections are the cumulative result of work done by various students, volunteers, and staff members. Most of the links on the IPL website are stored, along with metadata describing them, in a MySQL database known as Hypatia. Students, volunteers, and staff make maintain the database through the Hypatia web interface, written in the Mason and Perl programming languages. Most of the IPL’s own webpages are generated by computer programs (or scripts) written in Mason and Perl. These scripts assemble various pieces of data and HTML code, such as links to external websites stored in the Hypatia database long, along with the title and abstract metadata describing those links, together into a complete webpage on the IPL.

## 2. CROSSWALKING IPL METADATA

In January 2007, the IPL servers were moved to *X University*. Subsequently, it was decided to crosswalk the existing metadata to Dublin Core. Standards compliant metadata allow for metadata exposure and harvesting and open up access to the implementation of DC-compliant tools for metadata maintenance and generation. Migration of IPL data to and from other systems will be easier, and the integration of metadata from new IPL partner collections (such as the Librarian’s Internet Index: <http://www.lii.org/>) will be supported.

A number of factors will affect the quality of the crosswalk, including the completeness, accuracy, and consistency [17] of existing IPL metadata. A survey of these dimensions is currently underway. With regard to completeness, a quantitative analysis of the existing Hypatia database using SQL queries shows that a few data fields (e.g. *Title*, *URL*, *Abstract*) are heavily used while other fields are lightly used, and several metadata fields are frequently neglected [c.f. 12, 16]. With regard to accuracy and consistency, a qualitative analysis of IPL records by volunteers and students at IPL consortium universities has begun, using an online form to assess quality of the metadata in randomly selected records from the Hypatia database. The first version of the tool is implemented using Google Docs Forms (URL), which can be presented easily in various online formats, provide built-in reporting features, and can exported easily into an Excel spreadsheet or other commonly used data formats. The form is available online [6]. Finally, an analytical comparison of existing IPL and Dublin Core shows that there is no direct field-to-field mapping between the two: fields are labeled differently, defined differently, and may have the same data recorded in different ways. A number of metadata fields are particular to Hypatia only, and some of these are no longer in use.

## 3. NEW IPL METADATA SCHEMA

Concurrent with these activities, a new IPL metadata schema was created by applying the concept of application profile. The new IPL metadata schema is designed to facilitate resource description and discovery of networked Web resources. The IPL metadata schema exploits the existing Dublin Core metadata element set maintained by the Dublin Core Metadata Initiative [4], and also contains IPL domain specific metadata elements and qualifiers. It consists of four namespaces:

- Dublin Core Metadata Element Set (DCMES Version 1.1)
- Dublin Core Metadata Element Set (DCMES) Qualifiers (2000-07-11)
- IPL-defined Metadata Element Set (IPLMES)
- IPL-defined Metadata Element Set (IPLMES) Qualifiers

IPL defined elements and qualifiers mostly concern administrative and technical aspects of metadata (e.g., service provider information, record information, user comments). These elements and qualifiers will result in null mapping to Dublin Core.

The Dublin Core metadata scheme offers flexibility built directly into the framework. Different labels for DC element names, specified element status across local guidelines (e.g., mandatory, optional, repeatable) and best practices evidences the flexibility of the Dublin Core metadata scheme [13]. The new IPL metadata schema specifies element status and repeatability by taking the

IPL context into account. However, it should be noted that there were many challenges in reaching consensus on metadata labels and element status within the IPL Dublin Core compliance group. Content designation rules and semantic aspects of IPL metadata schema need to be further developed.

#### 4. METADATA INTERFACE

As outlined above, the quality of the crosswalked IPL metadata will only be approximate, and will need to be reviewed and if necessary edited. Work has started on a metadata creation and maintenance interface to support this task. Three browser tabs or windows are currently needed to accomplish this task: one for the record, one for the external online resource, and one for the online evaluation form. Early user testing suggests that the new interface has to integrate the qualitative and quantitative analyses of the crosswalked metadata within a single window [c.f. 10]. The interface development will be a long term project that will require careful design to result in a metadata tool that is useful and avoids usability problems [3, 5, 11]. We will follow an iterative user-centered design process to develop and test the new metadata interface. Initial prototypes and mock-ups will be tested in HCI and metadata classes taught at the iSchool. Further development of the interface will continue in 2009, supported by a grant from the OCLC-ALISE Library and Information Science Research Grant Program (LISGRP).

#### 5. REFERENCES

- [1] Barton, J., Currier, S., & Hey, J. (2003). Building Quality Assurance into Metadata Creation. Dublin Core Conference 2003, 28th Sept - 2nd Oct, Seattle, WA, USA.
- [2] Beall, J. (2004). Metadata and data quality problems in the digital library. *Journal of Digital Information* 6(3).
- [3] Crystal, A., & Greenberg, J. (2005). Usability of a metadata creation application for resource authors. *Library & Information Science Research* 27 (2005), 177-189.
- [4] Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, Version 1.1:  
<http://dublincore.org/documents/dces/>
- [5] Greenberg, J., Crystal, A., Robertson, W., & Leadem, E. (2003) Iterative design of metadata creation tools for resource authors. Dublin Core Conference 2003, 28th Sept - 2nd Oct, Seattle, WA, USA.
- [6] IPL (2008). Metadata Evaluation Form.  
<http://training.ipl.org/div/backroom/col/module2/metadatevalform.html>
- [7] IPL (2008). The IPL Class(es) Of...  
<http://www.ipl.org/div/class/>
- [8] IPL (2008). The Mission and Vision Statements of the Internet Public Library.  
[http://www.ipl.org/div/about/mission\\_and\\_vision.html](http://www.ipl.org/div/about/mission_and_vision.html)
- [9] Janes, J. (1998). The Internet Public Library: An Intellectual History. *Library Hi Tech*. 16,2 (Issue 62) (1998), 56.
- [10] Khoo, M., Devaul, H., and Sumner, T. (2002). Functional Requirements for Groupware to Support Community-Led Collections Building. 6th European Conference on Digital Libraries, Rome, September 2002, pp. 190-203.
- [11] Khoo, M. (2005). The Tacit Dimensions of User Behavior: The Case of the Digital Water Education Library. JCDL '05, Denver, CO, June 7-11, 2005. Pp. 213-222.
- [12] Moen, W. E., Miksa, S. D., Eklund, A., Polyakov, S., & Snyder, G. (2006). Learning from artifacts: metadata utilization analysis. JCDL '06, Chapel Hill, NC, USA, June 11 - 15, 2006. Pp. 270-271.
- [13] Park, J. (in press). Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging and Classification Quarterly* Vo. 47 (January, 2009)
- [14] Park, J., and Lu, C. (2008). A Comparative Analysis of Metadata Creation Guidelines: Issues and Implications, 2008 ER&L Conference, Atlanta, GA, March 18-21, 2008.
- [15] Park, J. (2006). Semantic Interoperability and Metadata Quality: An Analysis of Metadata Item Records of Digital Image Collections. *Knowledge Organization* 33(1), 20-34.
- [16] Ward, J. (2003). A quantitative analysis of unqualified Dublin Core metadata element set usage within data providers registered with the open archives initiative. JCDL 2003, May 27-31, 2003, Houston, Texas. Pp. 315-317.
- [17] Wilson, A. (2007). Toward releasing the metadata bottleneck: a baseline evaluation of contributor-supplied metadata. *Library Resources and Technical Services* 51(1), 16-28.