2012

Authority Control in the Next Generation. A Report of the LITA/ALCTS Authority Control Interest Group Program, American Library Association Annual Conference, New Orleans, June 2011

Lynnette Fields
Southern Illinois University Edwardsville, lfields@siue.edu

AUTHORITY CONTROL IN THE NEXT GENERATION. A REPORT OF THE LITA/ALCTS AUTHROITY CONTROL INTEREST GROUP PROGRAM, AMERICAN LIBRARY ASSOCIATION ANNUAL CONFERENCE, JUNE 2011.

The Authority Control Interest Group (ACIG) provides a forum for discussion of a variety of issues related to authority control and the sharing of authority data. The program at ALA Annual 2011, "Authority Control in the Next Generation," consisted of routine reporting and presentations on technology implementations that will revolutionize authority control. The Library of Congress supplied an update report and three speakers discussed the future of authority control with linked data.

The Library of Congress report was presented by Janis Young, a senior cataloging specialist in the Library of Congress' Policy and Standards Division (PSD). She began with an update on RDA implementation. She reported that the U.S. national libraries will implement RDA no earlier than January 2013. LC's RDA testers will resume cataloging using RDA in November 2011. The PSD and PCC will resume work on Library of Congress Policy Statements (LCPS) later in 2011.

The Virtual International Authority File (VIAF) now has 20 members, who are contributing authority records for personal and corporate names. As of June 20, 2011, the weekly list of new and changed subject headings will be monthly instead of weekly. This will increase efficiency and streamline workflows at LC. A new subject heading proposal system will be implemented no earlier than July 18, 2011. This new proposal system will be for Library of Congress Subject Headings (LCSH), Library of Congress Genre/Form Terms (LCGFT), and Children's Subject Headings. LC and SACO will use the same input mechanism.

Young reported that more than 700 existing authority records were added or changed to create LCGFT. The LCGFT authority records have a prefix of 'gf' in the LCCN, and the LCCN from the LCSH authority record has been retained as an invalid LCCN in a $z. LC implemented law genre/form terms on June 15, 2011. LC continues to partner with the Music Library Association on music genre/form terms. LC has partnered with the American Theological Library Association for the development of religious terms. These terms should be approved some time in 2012. LC may begin working on literature terms sometime in 2011. Developing literature genre/form terms will be a huge project and may make LC rethink the way they catalog literature.

Someone from the audience asked how general terms will be added to LCGFT. Young responded that SAC has been working on these, and have completed their report. The report is awaiting a vote by SAC. Young completed her part of the program by stating that LC is currently working with Ex Libris on the issue of how the LCGFT terms display.

The next speaker, Karen Coyle, is a distinguished librarian and consultant, with over 30 years of experience with library technology. Coyle began her talk with the disclaimer that she is not a cataloger, and the presentation is based on her observations. Her presentation, "Authorities: The Things of Library Data," focused on the concept that the information world is made up of Things that have relationships to other Things. Her talk highlighted some of the differences between how the Semantic Web and libraries have managed their Things, and how those differences can be overcome.

Coyle began with Lois Mai Chan's definition of authority control, "To ensure consistency, an authority record is created for each authorized heading for a proper name or subject." She continued by explaining the difference between a heading and a person. A heading is a 'string,' and a person is a 'thing'. Coyle used William Shakespeare as an example. The heading or 'string' did not write "As you like

it." But the 'thing,' has a name "William Shakespeare" and wrote "As you like it." Authority records usually convey information about the name (string), not the person (thing). Things get identifiers that are consistent, persistent, and unique in some context. There should be rules for identifiers to ensure consistency. An identifier should always mean the same thing. Identifiers should not be reused. Identifiers should not be names or words, because words are ambiguous. Identifiers are made for machines to understand and must be precise and able to be displayed in any language. People, books and places are things.

Coyle continued to explain that strings are endpoints, but things can connect to other things. We can control the names (strings) of things by using identifiers for things and creating labels (strings). Labels are for people, machines need identifiers. Coyle concluded with the thought that we (catalogers) need to see the machine as the first user, and see ourselves as the second user.

The third speaker, Gordon Dunsire, is currently a freelance consultant. Dunsire has a background in cataloging and systems librarianship, and is a member of the FRBR Review Group and ISBD/XML Study Group which are developing Semantic Web representations of IFLA standards. Dunsire's presentation, "Authority Control, New Library Standards, and the Semantic Web" discussed the place of authority control in new bibliographic standards such as the Functional Requirements for Bibliographic Records (FRBR), authority data (FRSAD), and subject authority data (FRSAD), Resource Description and Access (RDA), and the consolidated edition of the International Standard Bibliographic Description (ISBD). A brief-non-technical introduction to the concept of triples and linked data was followed by a discussion of the importance of authority control concepts in the Semantic Web and their application to linked data.

Dunsire continued with an explanation of the RDA implementation scenario 1: a relational/object-oriented database structure. He argued that if we used identifiers in bibliographic records instead of text strings it would be safe to link to the identifier, because the identifier would not change. Dunsire went on to explain that linked data is not a new idea. Linked data has been around since the idea of authority control. Resource Description Framework (RDF) was designed for the Semantic Web. Because machines are dumb, everything must be disambiguated. RDF expresses metadata as "atomic" statements, constructed in 3 parts. The example Dunsire used to illustrate a triple was:
- The title of this book is "Cataloguing is fun!"
  - Subject of the statement = *Subject:* This book
  - Nature of the statement = *Predicate:* has title
  - Value of the statement = *Object:* Cataloguing is fun!
- This book – has title – "Cataloguing is fun!"
  - Subject-predicate-object
In order for this to work, there must be an unambiguous way of identifying each part of the triple for efficient machine-processing. This can be done with a Uniform Resource Identifier (URI). A URI can be any unique combination of numbers and letters. A URI can look like a URL, but it does not lead to a web page. RDF requires that the subject and predicate of a triple be a URI, but the object can be a URI, or a literal string ("Cataloguing is fun!") Predicates are called properties in RDF. Properties are the links in linked data.

Dunsire maintains that the use of URIs and RDF would allow for distributed authority control, but without the need to copy and maintain headings in the local catalog. Different labels for the same

thing could be linked. VIAF is an example of this. Dunsire suggests that we need a paradigm shift. All our current cataloging concepts should be re-examined.

Someone in the audience asked what can we do now to prepare for this new bibliographic world. Dunsire answered that we need to be able to express these concepts to colleagues in simple terms.

The final speaker, Robin Johnson, is a senior standards editor in the Getty Vocabulary Program at the Getty Research Institute. Her presentation, "The Getty Vocabularies: Issues in Authority Control for Art and Architecture" was an overview of the Getty Vocabularies, focusing on the ART & Architecture (AAT), the Getty Thesaurus of Geographic Names (TGN), the Union List of Artist Names (ULAN), and the new vocabulary in development, the Cultural Objects Name Authority (CONA). The Getty Vocabularies may be used as non-authoritarian authorities for cataloging art and architecture.

The AAT, ULAN and TGN were developed in the mid-1980s to meet the needs of the art museum, visual resources, archives and art library communities. The vocabularies are considered to be authoritative and contain a suggested term/name based on literary warrant, but users can use variant terms if appropriate. Over the years these thesauri have been regarded as standards for the indexing and retrieval of art related resources. Although each thesauri was conceived differently, they have been united with the same structure and editorial rules. At the request of the art community, the 4th vocabulary, Cultural Objects Name Authority (CONA) was begun in 2004 and should be ready for use in late 2011.

All four vocabularies focus on the visual arts and architecture, grow through contributions from the user community, and are compiled, maintained and distributed by the Getty Vocabulary Program, a unit of the Getty Research Institute. AAT terms represent generic concepts in art, TGN represents inhabited places and geological features, ULAN represents artists and firms, and CONA will represent titles or names of art works and buildings. Each vocabulary record is identified by a unique, persistent numeric ID. If a term changes or records are merged, the original ID number is retained.

The vocabularies grow and change over time. New terms come from Getty projects and from outside contributers, which include museums, libraries, and archives. The growth of the vocabularies is contingent on contributions from the art community. The vocabularies remain authoritative through training and compliance with editorial rules. The vocabularies comply with national and international standards (NISO and ISO) for thesaurus construction. The Getty vocabularies are used by catalogers and indexers who are describing works of art, archival materials, visual surrogates or bibliographic records, by researchers, and by systems implementers creating search tools to enhance access to online resources.

Johnson then went through each of the vocabularies, highlighting the scope, showing some examples, and discussing the vocabulary's current development. Because CONA is the newest vocabulary, she spent extra time going through its structure and development. CONA is the first Getty vocabulary to be linked to the other vocabularies. The other vocabularies use controlled lists based on terms from the other vocabularies. The Getty vocabularies are not multilingual, but many terms now included Terms/Names/Titles and descriptive and scope notes that are repeated in multiple languages.

Johnson concluded her presentation with information about the Getty vocabularies and the Virtual International Authority File (VIAF) and linked data. The Getty Institute was invited by the Library

of Congress to join the VIAF project and became an official member in March 2010. ULAN records are now a part of VIAF and work is currently being done to add TGN. The Getty vocabularies can be used as linked data using the unique ID numbers, as well as the semantic links (e.g. hierarchical, equivalence, associative relationships, links to sources, etc.) within the records themselves. The vocabularies also map well to Simple Knowledge Organization System (SKOS), and they are currently exploring how to express and disseminate the data as SKOS linked data.

This program was developed by the LITA/ALCTS Authority Control Interest Group (ACIG) and co-sponsored by ALA, Cataloging & Classification Quarterly, and MARCIVE, Inc.

Lynnette M. Fields
Southern Illinois University Edwardsville
Edwardsville, IL