



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

22-27 August 2004  
Buenos Aires, Argentina

Programme: <http://www.ifla.org/IV/ifla70/prog04.htm>

---

**Code Number:** 024-E  
**Meeting:** 148. IFLA-CDNL Alliance for Bibliographic Standards (ICABS)  
**Simultaneous Interpretation:** -

**Metadata, Protocol, and Identifier Activities: Library of Congress IFLA/CDNL Alliance for Bibliographic Standards Report, 2004, Buenos Aires**

**Sally H. McCallum**  
Library of Congress  
Washington DC, USA

---

### ***Abstract***

*The Library of Congress contribution to the IFLA/CDNL Standards Alliance includes MARC 21 formats, metadata, information retrieval protocols, and identifiers. This report covers a few significant recent activities for each. For MARC 21, the development of MARCXML and the MARC toolkit of format transformations is noted. In the metadata area, the Metadata Object Description Standard (MODS) is described along with its new authority data companion, MADS. Highlighted under information retrieval protocols is the new Z39.50 derivative web service, SRW/SRU, and under identifiers is the new “info” URI established specially for the information community. The status of the Virtual International Authority File model experiment completes the report.*

As you have heard today, ICABS is a new concept for sustaining IFLA work and IFLA participation in a myriad of important information activities. In some cases the programs were already well established in IFLA, in others we are still working out relationships. One basic premise has been that the activities are extensions of those already supported within an Alliance

partner institution. My site, the Library of Congress, is responsible for four areas relating to: MARC 21 and its XML derivatives; Z39.50 information retrieval protocol and its next-generation version; metadata and XML-based metadata schemes; persistent identifiers. We also partner with Die Deutsche Bibliothek in a Virtual International Authority File (VIAF) model experiment.

In the short time I have today I want to pick out one or two items under each of these areas that I think are especially important or that have had special development activities over the last year and report on them.

## **MARC 21**

As you are probably aware, MARC 21 is built on a format structure that is a well-established worldwide standard, ISO 2709. The structure, which is also used for UNIMARC and many other MARC-type formats has held up well and will be important for many more years. But today many new protocols are more “comfortable” with an XML-based structure for the bibliographic record. Responding to that need, alternative structures for the MARC 21 data elements have been developed over the last few years. The most important is MARCXML, as it provides a lossless pathway from MARC 21 to MARC in XML and then back, as needed.

The MARCXML version of MARC 21 is also a good vehicle to move data among other XML formats such as MODS, MADS, Dublin Core, ONIX (bibliographic data content), and potentially others. MARCXML serves well since it is more detailed than most other descriptive data formats so can maintain distinctions that may be important for transformations to other schemas. A set of downloadable transformations are maintained at the MARC 21 web site to facilitate switching among these formats.

These developments have been important for opening up our vast stores of MARC 21 data to easier exchange with XML tools and protocols. Examples of these protocols are the harvesting protocol of the Open Archives Initiative (OAI), the new Z39.50-based search and retrieve protocol (SRW/SRU), and the Metadata Encoding and Retrieval Standard (METS). Using these transformation tools at the Library of Congress, responses to Z39.50 and SRW searches can now be sent in a variety of formats, including MARC 21, MARCXML, MODS, and Dublin Core. We are working with Danish colleagues to establish an ISO standard for representation of any ISO 2709 data in XML in a manner consistent with MARCXML.

I cannot leave the MARC21 action area without mentioning the MARC 21 to FRBR tool that we have developed and offer from the MARC 21 web site for users interested in experimenting with FRBR. FRBR is an important concept coming out of IFLA and supported by another ICABS alliance partner, the British Library. This is a simple tool - it does not claim to be the complete FRBR tool - but it can be very useful in testing possibilities for employing FRBR concepts and the consistency and potential of one's data. It is intended to act on a search retrieval set of records, so it is a “post-search” FRBRizing tool.

## Metadata

Here I want to review two activities from the last year. The first ties in with the MARC 21 initiative since the development of the Metadata Object Description Standard (MODS) grew from the often expressed need for an XML format for electronic resources that could be used by staff to expedite records for those resources. Thus MODS has language-based tags (rather than numeric), some special characteristics for e-resources, and a good compatibility with MARC 21. MODS lacks the detail of MARC but comes close in the core element areas. It does not replace MARC for applications that require the MARC 21 detail for description and retrieval.

It can be used as an alternative view of MARC data for digital conversion items that have already have been cataloged using MARC 21 for the library's catalog. A subset of that metadata may be needed for a repository that archives and disseminates the e-resource version of the item. It is also a good lightweight format for original input for "born-digital" e-resources, and is especially useful for high volume resources like web sites. The format has an important recursive related item capability that enables inclusion of hierarchically related information about the electronic resources. For each data element in MODS, language, script, and transliteration may be indicated.

Another valuable feature of MODS is its very new companion: Metadata Authority Description Standard (MADS). MADS has a relationship to the MARC 21 Authority format, as MODS has to MARC 21 Bibliographic, but it simplifies data and relates to MODS itself in new and useful ways -- influenced by the FRBR development. MADS facilitates recording authorized forms and reference forms of names for basic entities such as names, titles, geographics, genres, topical subjects and temporal subjects.

It is worth noting that the development of both of these formats takes place over the Internet, via an open listserv. In both cases the Library of Congress released the seed format and then took on the editing and web site dissemination (via a web site) functions.

The second really important development in the metadata arena is the development of METS over the last few years. METS can be called a technique for recording, in a highly flexible manner, metadata about all aspects of a digital object. This includes the descriptive metadata that we use MARC and MODS to encode and also technical, preservation, administrative, structural, etc. METS packages for an e-resource include or point to this metadata and specify the format of the metadata, increasing the flexibility. Of course flexibility comes at a price and the early METS users are also defining profiles for different types of resources that will increase the potential for interoperability.

And finally I want to point out a resource that you may find useful as you sort through the different rights expression languages for rights metadata. We commissioned a report to examine the prominent ones today (early 2004), most prominently ODRL, METSrights, Creative Commons and MPEG-21/5. Their underlying goals and assumptions were examined and taxonomy for evaluating them for use established.

## **Information Retrieval : Z39.50 and SRW**

The Z39.50 information retrieval protocol is widely deployed in library systems today. The exciting recent development in this area is, however, the "Z39.50 next generation" of this technology adapted to the web environment, SRW and SRU. SRU stands for Search and Retrieve URL Service and it allows users to send a search using title, name, identifier and other parameters via a URL (HTTP GET) and receive records in response. SRW stands for Search and Retrieve Web Service and it supports a web service that operates over the web base protocol, SOAP, rather than directly through HTTP. It is more robust than SRU is able to be directly over HTTP.

These XML-based search and retrieval protocols are semantically compatible with Z39.50, adapting the most useful parts of that protocol to the web environment. This enables users to provide SRW/SRU gateways to existing Z39.50 applications. Developed by an international editorial group, with review through an open listserv, version 1.1 was released earlier this year (superseding 1.0, an earlier experimental version released in 2002). It has already been incorporated into some products, and my own institution, the Library of Congress, can be accessed via an SRW/SRU gateway.

The effort also resulted in the development of a Common Query Language (CQL). CQL attempts to bridge the gap between different approaches to search languages: those that are powerful but complex, cryptic, and non-user friendly, such as XQuery or SQL; and those that are simple and user-friendly but lacking in functionality such as Google. CQL attempts to combine functionality and user-friendliness.

SRW can combine with OAI and OpenURL to provide powerful and flexible tools for retrieving resources on the Internet.

## **Persistent Identifiers**

The Library of Congress is charged with monitoring the developments of the persistent identifiers. We plan to team up with the National Library of Australia, which has a good site for information on Identifiers and offer a complementary site at LC. At the LC/ICABS site, we are going to mount current information and descriptions of recent activities in the W3C and IETF. We will point to the Australian PADI site for much information is maintained there on this topic. When we started to develop our role in this area we immediately saw that collaboration with another ICABS site was in order.

One significant development in this area in the last year was the "info" URL. Identifiers are a complex topic, but there are three key characteristics for identifiers that the community seeks - persistent, actionable, and unique - and thus far attempts to have then all in one identifier have not been successful. The "info" URI is designed to satisfy the development of identifiers that satisfy the uniqueness characteristic and also to expedite their authoritative allocation for the library and publishing communities. By focusing on the uniqueness characteristic, the "info" URI

can be used to reference digital and non-digital resources, schemas, data elements, codes, etc. It emerged as a requirement for the OpenURL but its usefulness for enabling legacy identification systems to become part of the global web architecture was also recognized. Several of these identifiers have been registered as namespaces under the “info” URI, including the DOI, LCCN (Library of Congress Control Number), and other well-known identifiers.

## **VIAF**

I was asked by the Die Deutsche Bibliothek (DDB) to report on the Virtual International Authority File (VIAF) action item, which is their ICABS responsibility with the Library of Congress. This action item explores a possible model for a “VIAF”. In the last year, an agreement was signed with the partners in the project, which includes OCLC, and files were exchanged for matching records. Several rounds of refinement of the matching algorithms will be needed, followed by the integration of the matching personal name authority records and verification of the matches through checking of the bibliographic records. These DDB, LC, and Bayerischer Bibliotheksverbund (BVB) records will be used to construct an initial VIAF according to this model. The expectation is that the first phase will be completed in 2004. If it is successful, work will be undertaken to support access to the experimental VIAF file server via OAI protocols and through a multilingual interface.

*(2<sup>ND</sup> Version: 1 July 2004)*