



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-G
Meeting:	148. IFLA-CDNL Alliance for Bibliographic Standards (ICABS)
Simultaneous Interpretation:	-

Aktivitäten zu Metadaten, Protokoll und Identifiers: der ICABS-Bericht der Library of Congress

Sally H. McCallum

Library of Congress
Washington DC, USA

Abstract

Der Beitrag der Library of Congress zur IFLA-CDNL Alliance for Bibliographic Standards (ICABS) umfasst Aktivitäten zum Format MARC21, Metadaten, Protokolle zum Information-Retrieval und Identifiers. Dieser Bericht bezieht sich auf einige wesentliche Aktivitäten der jüngsten Vergangenheit zu jedem dieser Themen. Für MARC21 wird über MARCXML und das MARC-Toolkit für Formattransformationen berichtet. Im Bereich Metadaten wird der Metadata Object Description Standard (MODS) gemeinsam mit seinem neuen Pendant für Normdaten MADS (Metadata Authority Description Standard) beschrieben. Hervorzuheben bei den Protokollen zum Information-Retrieval sind die neuen von Z39.50 abgeleiteten Protokolle SRW/SRU, die auf dem http-Protokoll aufsetzen, und bei den Identifiers wurde der neue „Info“-URI speziell für das Informationswesen entworfen. Ein Sachstandsbericht über den Modellversuch eines Virtual International Authority File vervollständigt den Bericht.

Wie Sie bereits gehört haben, ist ICABS ein neues Konzept zur Unterstützung der IFLA-Arbeit und der Beteiligung von IFLA an einer Fülle von wichtigen Informationsaktivitäten. In einigen Fällen sind die Vorhaben von ICABS innerhalb der IFLA bereits gut etabliert, in anderen arbeiten wir noch daran, sie richtig einzubetten. Eine wesentliche Voraussetzung hierfür ist, dass die Aktivitäten Erweiterungen derjenigen Aktivitäten sind, die bereits innerhalb einer Partnerinstitution der Allianz vorhanden sind. Mein Haus, die Library of

Congress, ist verantwortlich für vier Bereiche: MARC21 und seine XML-Ableitungen, das Z39.50-Information-Retrieval-Protokoll und dessen nächste Generation, Metadaten und XML-basierte Metadaten-Schemas und Persistent Identifiers. Gemeinsam mit Der Deutschen Bibliothek sind wir auch Partner bei dem Modellversuch eines Virtual International Authority File (VIAF).

In der kurzen Zeit, die mir heute zur Verfügung steht, möchte ich aus jedem dieser Bereiche einen oder zwei Punkte herausgreifen, von denen ich glaube, dass sie besonders wichtig sind, oder die spezielle Entwicklungen im Verlauf des letzten Jahres erfahren haben, und über sie berichten.

MARC21

Wie Sie wahrscheinlich wissen, baut MARC21 auf einer international standardisierten Formatstruktur auf, der ISO 2709. Diese Struktur, die auch für UNIMARC und viele andere MARC-ähnliche Formate genutzt wird, war sehr dauerhaft und wird ihre Bedeutung noch viele weitere Jahre haben. Aber heutzutage sind viele neue Protokolle durch eine XML-basierte Struktur „komfortabler“ für die Titelaufnahme. Während der letzten Jahre wurden als Antwort auf dieses Bedürfnis alternative Strukturen für die MARC21-Datenelemente entwickelt. Die wichtigste ist MARCXML, weil es eine Übertragung ohne Informationsverlust von MARC21 zu MARC in XML und auch wieder zurück ermöglicht, wenn dies notwendig ist.

Die MARCXML-Version von MARC21 ist auch ein gutes Mittel, um Daten zwischen anderen XML-Formaten wie MODS, MADS, Dublin Core, ONIX (Inhalte bibliografischer Daten) und möglichen anderen zu bewegen. MARCXML leistet dabei gute Dienste, weil es detaillierter ist als die meisten anderen beschreibenden Datenformate und so Unterscheidungen beibehalten kann, die zur Transformation in andere Schemas notwendig sein können. Ein Transformations-Set zum Herunterladen wird auf der MARC21-Website angeboten, um das Wechseln zwischen diesen Formaten zu erleichtern.

Diese Entwicklungen sind wichtig, um unsere umfangreichen Speicher von MARC21-Daten für einen leichteren Datenaustausch mit XML-Werkzeugen und –Protokollen zugänglich zu machen. Beispiele für diese Protokolle sind das Harvesting-Protokoll der Open Archives Initiative (OAI), das neue Z39.50-basierte Search- and-Retrieval-Protokoll (SRW/SRU) und der Metadata Encoding and Retrieval Standard (METS). Durch die Anwendung dieser Transformations-Tools in der Library of Congress können Antworten auf Anfragen über Z39.50 und SRW nun in einer Vielzahl von Formaten geliefert werden, einschließlich MARC21, MARCXML, MODS und Dublin Core. Wir arbeiten gemeinsam mit dänischen Kollegen daran, einen ISO-Standard zur XML-Darstellung jeglicher nach ISO 2709 erfasster Daten in einer mit MARCXML konsistenten Form einzuführen.

Ich kann den Bereich MARC21 nicht abschließen, ohne das Tool zu erwähnen, das eine automatische Konvertierung von MARC21 zu FRBR unterstützt, welches wir entwickelt haben und auf der MARC21-Website für Anwender anbieten, die mit den FRBR experimentieren möchten. Die FRBR sind ein wichtiges Konzept, das innerhalb der IFLA entwickelt wurde und von einem anderen ICABS-Partner, der British Library, unterstützt wird. Es handelt sich um ein einfaches Tool, das keinen Anspruch darauf erhebt, ein perfektes FRBR-Tool zu sein, aber beim Testen von Möglichkeiten, das FRBR-Konzept einzusetzen und zur Überprüfung der Konsistenz und dem Potenzial von Daten kann es sehr nützlich sein. Es ist gedacht für den Einsatz als Search- and-Retrieval-Set für Datensätze, es ist also ein Tool zur FRBRisierung, das nach der Suche zum Einsatz kommt.

Metadaten

An dieser Stelle möchte ich zwei Aktivitäten des letzten Jahres ansprechen. Die erste steht in Verbindung mit der MARC21-Initiative, weil die Entwicklung des Metadata Object

Description Standard (MODS) aus dem häufig geäußerten Bedarf an einem XML-Format für elektronische Ressourcen erwachsen ist, das vom Personal bei der Erstellung von Aufnahmen für diese Ressourcen benutzt werden könnte.

Deshalb hat MODS sprachbasierte (statt numerische) Tags sowie einige besondere Kennzeichnungen für elektronische Ressourcen und ist gut mit MARC21 kompatibel. Der Standard MODS ist nicht so detailliert wie MARC, aber kommt ihm im Bereich der Hauptelemente sehr nahe. Er ersetzt MARC nicht für Anwendungen, welche den hohen Detaillierungsgrad von MARC21 für die Beschreibung und das Retrieval erfordern.

Er kann als alternative Sicht auf MARC-Daten für Digitalisate genutzt werden, die für den Bibliothekskatalog bereits nach MARC21 katalogisiert wurden. Eine Teilmenge dieser Metadaten ist für einen Repository-Server erforderlich, auf dem die elektronische Version archiviert und verbreitet wird. MODS ist ebenso ein gutes „abgespecktes Format“ zur Erfassung genuin elektronischer Ressourcen, und besonders sinnvoll ist es für umfangreiche Ressourcen wie Web-Sites. Mit dem Format ist die Abbildung hierarchischer Strukturen über elektronische Ressourcen besonders gut möglich.

Für jedes Datenelement in MODS kann die Sprache, Schrift und Transliteration angezeigt werden.

Eine andere wertvolle Besonderheit von MODS ist sein neues Pendant: Metadata Authority Description Standard (MADS). MADS ist mit dem Format MARC21/Authority verwandt, so wie MODS mit MARC21/Bibliographic, aber es vereinfacht Daten und bezieht sich selbst, beeinflusst durch die FRBR-Entwicklung, in neuer und nützlicher Weise auf MODS. MADS erleichtert das Erfassen von Ansetzungs- und Verweisungsformen für grundlegende Entitäten wie Namen, Titel, Geografika, Genres, Themen- und Zeitschlagwörter.

Es ist erwähnenswert, dass die Entwicklung dieser beiden Formate durch eine offene E-Mail-Diskussionsliste über das Internet stattfindet. In beiden Fällen hat die Library of Congress das Grundformat freigegeben und dann die Editier- und Verbreitungsfunktionen (über eine Web-Site) übernommen.

Die zweite, wirklich bedeutende Entwicklung im Bereich Metadaten ist die Entwicklung von METS während der letzten Jahre. METS kann als Methode zur Erfassung von Metadaten aller Facetten eines digitalen Objekts in einer sehr flexiblen Art bezeichnet werden. Dies umfasst die beschreibenden Metadaten, die wir in MARC und MODS kodieren, und ebenso technische, strukturelle sowie Erhaltungs- und Verwaltungsmetadaten etc. METS-Pakete für eine elektronische Ressource schließen Metadaten ein oder verweisen auf sie. Sie spezifizieren das Format der Metadaten und steigern die Flexibilität. Flexibilität hat natürlich ihren Preis und die ersten METS-Anwender definieren Profile für verschiedene Ressourcentypen, die das Potenzial für die Interoperabilität steigern.

Schließlich möchte ich auf eine Ressource hinweisen, die im Zusammenhang von Metadaten für die Rechteverwaltung nützlich sein kann. Wir haben einen Bericht in Auftrag gegeben, um die aktuell (Anfang 2004) wichtigen dieser Metadatenansätze zu untersuchen, hauptsächlich ODRL, METSrights, Creative Commons und MPEG-21/5. Es wurde untersucht, welche Ziele und Annahmen ihnen zugrunde liegen, und ein Kriterienkatalog zur Evaluation ihrer Nutzung wurde eingeführt.

Information Retrieval: Z39.50 und SRW

Das Z39.50-Information-Retrieval-Protokoll ist gegenwärtig in Bibliothekssystemen weit verbreitet. Die vielversprechenden jüngsten Entwicklungen auf diesem Gebiet sind jedoch die „Z39.50 Next Generation“ dieser Technologie, die an eine Web-Umgebung angepasst sind: SRW und SRU. SRU steht für „Search and Retrieve URL Service“ und erlaubt es den

Benutzern, eine Suche durchzuführen, bei der sie Titel, Namen, Identifiers und andere Parameter mittels einer URL (HTTP GET) eingeben und dafür als Antwort die entsprechenden Datensätze erhalten. SRW steht für „Search and Retrieve Web Service“ und unterstützt einen Web-Dienst, der über das Web-basierte - Protokoll SOAP arbeitet statt direkt über das http-Protokoll zu gehen. SRW ist stabiler als SRU, das direkt über das http-Protokoll angesprochen wird.

Diese XML-basierten Search- and-Retrieval-Protokolle sind semantisch kompatibel mit Z39.50 und die nützlichsten Teile dieses Protokolls sind an die Web-Umgebung angepasst. Dies ermöglicht den Anwendern, SRW/SRU-Gateways zu existierenden Z39.50-Anwendungen anzubieten. Entwickelt durch eine internationale Gruppe mit einer Prüfung durch eine offene Listendiskussion, wurde Version 1.1 zu Beginn dieses Jahres freigegeben (die Version 1.0, eine ältere Testversion aus dem Jahr 2002 wurde ersetzt). Diese Version wurde bereits in einige Produkte integriert und meine eigene Institution, die Library of Congress, ist über ein SRW/SRU-Gateway zugänglich.

Die Arbeit mündete auch in die Entwicklung einer allgemeinen Suchsprache (Common Query Language, CQL). CQL versucht, die Lücke zwischen verschiedenen Herangehensweisen an Suchsprachen zu schließen: zwischen denen, die leistungsfähig, aber komplex, kryptisch und benutzerunfreundlich sind, wie XQuery oder SQL, und andererseits denjenigen, die einfach und nutzerfreundlich sind, denen aber Funktionalitäten fehlen, wie z. B. Google. CQL versucht, Funktionalität mit Benutzerfreundlichkeit zu kombinieren.

SRW kann mit OAI und OpenURL kombiniert werden und bietet damit ein leistungsfähiges und flexibles Retrieval-Werkzeug für Web-Ressourcen.

Persistent Identifier

Die Library of Congress wurde damit beauftragt, die Entwicklung im Bereich der Persistent Identifiers zu beobachten. Wir planen, mit der National Library of Australia, die bereits eine gute Webpräsenz mit Informationen über Identifikatoren hat, zusammenzuarbeiten und eine ergänzende Site bei der LoC anzubieten. Auf den LoC/ICABS-Seiten werden wir aktuelle Informationen und Beschreibungen aktueller Aktivitäten beim W3C und bei IETF anbieten. Wir werden auf die Seiten des australischen PADI verweisen, weil dort viele Informationen zu diesem Thema angeboten werden. Als wir begannen, unsere Aufgabe für diesen Bereich zu planen, haben wir sofort erkannt, dass die Zusammenarbeit mit einem anderen ICABS-Bereich sinnvoll war.

Eine bedeutende Entwicklung in diesem Bereich innerhalb des letzten Jahres war der „Info“-URI. Identifiers sind ein komplexes Thema, aber es gibt drei Hauptmerkmale für Identifiers, nach denen alle streben – dauerhaft, benutzbar und weltweit eindeutig – und bis dahin waren keine der Bemühungen erfolgreich, diese Merkmale in einem Identifier zusammenzuführen. Der „Info“-URI wurde konzipiert, um die Entwicklung der Identifiers zu forcieren, die das Merkmal der Eindeutigkeit erfüllen, und auch um sie in ihrer Zugehörigkeit zu den Normdaten für das Bibliotheks- und Informationswesen zu fördern. Durch die Konzentration auf das Merkmal der Eindeutigkeit kann der „Info“-URI dazu benutzt werden, digitale und nicht digitale Ressourcen, Schemas, Datenelemente, Codes etc. zu identifizieren. Er ist zwar als Anforderung aus dem OpenURL-Kontext hervorgegangen, aber es wurden auch die Vorteile eines standardisierten Identifier-Systems als Bestandteil der globalen Web-Architektur erkannt. Mehrere dieser Identifiers, einschließlich DOI, LCCN und anderer bekannter, wurden als Namespaces unter dem „Info“-URI registriert.

VIAF

Die Deutsche Bibliothek hat mich gebeten, über den Bereich des Virtual International Authority File (VIAF) zu berichten, der in die gemeinsame Verantwortung Der Deutschen Bibliothek und der Library of Congress fällt. In diesem Projekt wird ein mögliches Modell für einen „VIAF“ untersucht. Im letzten Jahr wurde von den Projektpartnern, zu denen auch OCLC gehört, eine Vereinbarung unterzeichnet, und für einen Datenabgleich wurden Norm- und Titeldaten ausgetauscht. Mehrere Runden einer Verfeinerung der Algorithmen für den Abgleich werden nötig sein, gefolgt durch die Vervollständigung der abgeglichenen Normdaten für Personennamen und die Überprüfung der Übereinstimmungen anhand der Titeldaten. Die Titeldaten Der Deutschen Bibliothek, der Library of Congress und des Bayerischen Bibliotheksverbundes (BVB) werden benutzt, um einen VIAF nach diesem Modell aufzubauen. Es wird erwartet, dass diese erste Phase im Jahr 2004 abgeschlossen werden wird. Wenn der Test erfolgreich verläuft, wird daran gearbeitet, den Zugang zum VIAF-Test-Server via OAI-Protokoll und mit einer mehrsprachigen Oberfläche zu unterstützen.

Übersetzt von Susanne Oehlschläger
Die Deutsche Bibliothek, ICABS-Sekretariat