EO/GEO World Wide Web Workshop '97

Metadata Standards and MARC

February 4, 1997
Landover, Maryland

Rebecca Guenther, Library of Congress
rgue@loc.gov

"Metadata Standards and MARC"

- What is MARC?

- MARC as Metadata for Internet Resources

- Integration of different data structures

- Metadata in navigating digital collections
  - LC National Digital Library

- MARC and the Dublin Core

- Mapping MARC to Metadata Standards
  - GILS
  - CSDGM
  - EAD
  - SGML
  - Z39.50

MARC = MAchine Readable Cataloging

• **Encoding of information into *machine-readable* format so that it can be processed by computer**

•**Originally designed for the *bibliographic* information found in library catalogs**

•**MARC is not a *system***

•**MARC is a data record *structure***

• **MARC is a structure for communicating *metadata***

MARC Home Page:

http://www.loc.gov/marc

MARC RECORD ELEMENTS:

A MARC record consists of three elements:

1) The record *structure*
- standard form

2) The *content designation*
- uses discreet tags, codes

3) The *data content*
- may be governed by other standards

MARC RECORD STRUCTURE:

A MARC record's structure follows national (Z39.2) and international (ISO 2709) standards. It consists of three (3) structural components:

1) Leader - The first 24 character positions

2) Directory - 13-character entries (1 per field)

3) Variable fields - Of any number or length

STANDARDS GOVERNING DATA CONTENT:

- **Cataloging rules**; **such as *AACR2*, that prescribe how to create catalog entries and provide access**

- **International standards; publications such as the various *ISBD's*, used to control description**

- **Local/national laws; require certain data**

- **Policy statements; control what information is gathered and how it is recorded in databases**

- **Non-bibliographic content standards; conversion to and from MARC of other metadata standards, e.g. CSDGM**

THE MARC FORMATS:

- **A format defines...** the list of valid data elements for specific record types

- **In fixed-length fields...** the character positions;

- **In variable fields...** tags, indicators, subfield codes;

- **Requirements...** mandatory data and optional data;

- **Repeatability...** what is repeatable or not repeatable

SAMPLE OF EXISTING MARC FORMATS:

- **USMARC - Used throughout the world**
- **UNIMARC - Growing use around the world, particularly in Eastern Europe**
- **CCF - Used mostly in Third World countries**
- **CANMARC - Used in Canada ***
- **UKMARC - Used in Great Britain/Europe ***
- **IBERMARC - Used in Spain ***
- **SAMARC - Used in South Africa**
- **Etc., etc. (other MARC formats)**
- **Current format harmonization efforts**

(* Formats closely resembling USMARC)

MARC FORMAT USERS:

1) Libraries and media centers
2) Archives
3) Community information providers (county and state-wide agencies)
4) Museums (apart from their libraries)
5) Government agencies
6) School systems and universities

THE STRENGTHS OF MARC:

- **MARC has been used in libraries since 1968**

- **There is a high level of conformance to a limited number of standard implementions**

- **MARC is the foundation of library networks**

- **MARC has helped to unify the world of bibliographic information**

- **MARC is very flexible (for all kinds of data)**

- **Large bibliographic systems, e.g. LC**
  11 million bibliographic records
  4 million authority records

MARC can contain metadata about Internet resources

• Internet Resources Cataloging Guidelines

• Extensions to MARC:
descriptive terms for type of resource
Electronic Location and Access Field

• Link from record to resource

• INTERCAT database
http://www.oclc.org:6990

MARC for cataloging Internet resources

- incorporate records into library catalog
- system can process in same way
- records shared between systems
- less duplication of effort
- flexibility and familiarity of format
- Use existing record for description; add pointer to digitized version
- High (collection) level description

Integration of different metadata structures

- MARC bibliographic record
- SGML text
- SGML metadata record
- Dublin Core metadata record
- Header in HTML document
- Finding aid in EAD (Encoded Archival Description)
- TEI header in searchable text reproduction

Uses for Metadata

- Resource discovery and retrieval
- Facilitate searching on WWW
- Provide metadata simultaneously with data
- Allow inclusion of templates in software
- Metadata serve as basis for fuller descriptive record
- Ensure common set of elements generally understood across communities
- Allow for interoperability among different data structures (MARC, TEI, SGML, etc.)

Indexes

Importance of organizing structures of metadata

Flat:

A single index of all the words in all the documents of the database.

Supports raw-text searching.

Vs.

Usage-specific

E.g., Author Index, Title Index, Subject Index, Abstract Index.

Supports fielded searches.

Problems with current Web crawlers

•They Waste bandwidth

■Grab entire documents

■Index everything

■visit "mirror" sites

■repeated requests

■multiple, redundant crawlers

•They're annoying

■multi-tasking
■multiple accesses
■skew statistics

Problems with crawlers
(continued)

•They're dumb

- ■don't know where to look, so they look everywhere
- ■download inappropriate data
- ■can't discern relative importance
- ■can't discern context

•They index out-of-date information

•They miss information

•Importance of structured data

Using Metadata for Navigating Digital Collections

Collection: *group of related objects and sub-collections*.

Collections are:

- **Organized thematically (e.g. by subject, creator, historical period),**
- **with diverse objects (maybe in different formats);**
- **with associated description(s)**
- **hierarchical;**
- **distributed.**

Collection
Descriptive Record

Brief text Description

Access Aid
(Associated Description)

Pointers to
Related Collections

List of Members in
Collection

An object or a collection may have one or more Access Aids:

- Finding Aid (textual or marked up)
- EAD
- Cataloging Record
- Exhibition catalog
- GILS Record
- Web Page

Library of Congress National Digital Library

Framework

Home page

HTML text files

Access Aids

Descriptive metadata

MARC bibliographic records

EAD finding aids

Full text

Reproductions

Digital images

Recorded sound/moving image files

Searchable text

Special presentations

Introduce collections

Reproduction and access aid linking

Logical Name
(metadata conveyed in name)

↓

Document Locator System
• Physical location • Associated metadata physical structure rights/permissions terms/conditions administrative

↓

Retrieve resource

OCLC/NCSA Metadata Workshop

· Provide metadata for network-accessible materials

· Bring together stakeholders from various communities to foster common understandings

· Reach consensus on definition of simple data element set

· Attempt to satisfy divergent and specific needs

· Allow information providers to describe own resources

• Facilitate searching; indexing on the WWW

Scope:

• Document-Like Objects (DLOs)
• Support resource discovery on Internet
• Provide simple structure that can be enhanced to more complex records

Assumptions:

· Intrinsicality
· Extensibility
· Syntax Independent
· Optionality
· Repeatability
· Modifiability

Dublin Core

· Thirteen data elements:

· Access points

· Facilitate identification

· Relate object to other objects

· General applicability

· NOT the minimum number of required elements

· Use of qualifiers for more complex description

Dublin Core Meta Data Elements

- **Title**
- **Author or Creator**
- **Subject**
- **Description**
- **Publisher**
- **Contributors** ***(editor, ...)***
- **Date** ***(YYYYMMDD?)***
- **Resource Type**
- **Format**
- **Identifier** ***(url, urn, isbn)***
- **Source** ***(from which derived)***
- **Language**
- **Relation** ***(to other objects)***
- **Coverage** ***(spatial/temporal)***
- **Right management** ***(link)***

MARC and the Dublin Core

- **Simple (default mappings)**
- **Complex (qualifiers)**

Examples:

<u>Author or Creator</u>

Simple:
USMARC: 100$a (Main Entry--Personal Name) (with 1st indicator=1)
Complex:
Qualifiers:
If role=corporate: 110$a (Main entry--Corporate Name)

Subject and Keywords

Simple:

USMARC: 653$a
(Index Term--Uncontrolled)

Complex:

Qualifiers:
If scheme=LCSH: 650$a (Subject added entry--topical term)
If scheme=LCC: 050$a (Library of Congress Call Number/Classification number)
If scheme=DDC: 082$a (Dewey Decimal Call Number/Classification number)

Description

USMARC: 520$a (Summary, etc. note)

Publisher

USMARC: 260$b (Publication, Distribution, etc. (Imprint)/Name of publisher, distributor, etc.)

Warwick Framework

- Dublin Core package
- MARC record package
- Terms and conditions
- Rights
- Structural data
- Administrative data
- Provenance
- Relationship data
- Need for standardized registries

MARC Mappings to other metadata standards

• GILS:

Access to government information resources

Locator records (metadata)

Pointers to information resources

Federal, state, international

GILS Application Profile

Use of Z39.50

MARC, GRS, SUTRS record syntaxes

Mapping of GILS attributes to MARC

MARC Mappings to other metadata standards

- **Content Standards for Geospatial Metadata**

MARC used for catalog access

Crosswalk between CSDGM and MARC fields

MARC changes in June 1994

Alexandria project using MARC

Metadata record in library catalogs (for analog and digital objects)

MARC Mappings to other metadata standards

- **Encoded Archival Description (EAD):**
  electronic finding aids marked up in SGML

- **SGML Document Type Definition for MARC:**
  under developmemt

- **Z39.50 Profile for Digital Library**
  attributes have MARC equivalents

Mapping other data structures to MARC

- more fluid capability to work with same data in different structures

- agree on definition of elements

- need for standardization; proliferation of mappings

- need for metadata in two forms: embedded in document header; independent of item

- implications for resource discovery and retrieval

- navigating digital collections