CURRENT ENVIRONMENT

- There is an exponential increase in the quantity of available information;
- During the 20th century, the sciences contributed not only to the quantity of information, but also to its quality;
- The arrival of automation (computers) has revolutionized the processing information;
- Automation has improved access to information.
THE ROLE OF LIBRARIES IN AUTOMATION

- With MARC, librarians were some of the first professionals to apply automation to their work;
- The joined the telecommunication and finance industries in using computers to maintain services;
- Implementing computers was a major challenge;
- During the 60s, in large libraries, there was an overabundance of cards to be filed in the manual catalog;
INNOVATIONS DEVELOPED BY LIBRARIES

- Libraries were some of the first institutions to apply computers to textual data;
- Cataloging standards preceded the development of word processing software;
- Cataloging data are linguistically rich (many languages and scripts are represented in library collections);
- Almost all library systems were developed in-house, from the ground up (from zero).
EXAMPLES OF LIBRARY INNOVATIONS:

- Rich character encodings, including characters for a variety of languages (various scripts);
- Databases of variable length records;
- Explicit identification of data elements;
- Support for indexing, sorting, and searching;
- Sophisticated data entry, including prompt screens and validation (no more punched cards!)
THE EARLY YEARS:

- Beginning in 1969 with the first distribution of MARC records, libraries found themselves on the cutting edge of automation;
- A new market for machine-readable products developed;
- Bibliographic agencies like OCLC were born;
- Increase in machine-readable data;
- Increase in the demand for information.
INITIAL LIMITATIONS

- Databases were relatively isolated;
- Despite the quantity of records created, there was little exchange;
- Costs associated with original cataloging provoked exploration of data sharing;
- Initial shared cataloging projects were bases on tape distribution of records;
- The physical media had limitations.
BIRTH OF THE INTERNET

- The suggestion for “inter-networking” was proposed in 1962 for military applications;
- In 1967 four (4) separate computers were linked electronically with ARPANET;
- In 1971 the number of linked computers grew to 23;
- In 1972 the InterNetworking Working Group was established with Vinton Cerf as president;
- 1974 - Telenet; commercial version of ARPANET
CHILDHOOD OF THE INTERNET

- **1982-1987**: Vint Cerf and Bob Kahn develop TCP/IP (software that supports communication between systems)
- **1984**: increase in sales of PCs; appearance of the term cyberspace in the press;
- **1987**: Internet domains surpass 10,000;
- **1988**: First computer virus is spread over the Web;
- **1991**: NSFNET assumes the role of Web “backbone”
- **1993**: MOSAIC – first Web browser.
LIBRARIES CONFRONT NEW REALITIES

- The use of computers in libraries leads to the OPAC (Online Public Access Catalog);
- After the birth of the Internet, pressure on libraries increases to offer to outside users online access to their catalogs;
- The MARC formats continue to develop to handle new forms of material (video, audio, digital technologies, CD-ROM, etc.) and to provide richer bibliographic data.
In 1969, Charles Goldfarb, Edward Mosher and Raymond Lorie (G-M-L) led work at IBM on the Generalized Markup Language (GML);

Their work was based on a project to replace existing textual coding structures with something less proprietary;

ANSI considers a draft American standard in 1978;

In 1980 it becomes and ISO draft under the name: SGML - Standard Generalized Markup Language.
THE CHALLENGE OF SGML TO MARC

- Publication of ISO 8879 (SGML) in 1986;
- Growth in interest in SGML before being noticed by MARC users;
- The Library of Congress first considered its potential usefulness in 1990;
- SGML tagging was found to be compatible with MARC, not a threat;
- 1995 – start of a project to develop MARC-SGML.
SGML IN BRIEF

- SGML is a non-proprietary syntax for explicitly identifying text structures;
- Like MARC, implementation of SGML is based on predetermined tags;
- SGML tags must begin with a letter of the Latin alphabet and are delimited by the signs less-than and greater-than: <p>
- A list of SGML tags is called a “Document Type Definition” (acronym "DTD").
ADVANCED SGML CONCEPTS

- An SGML DTD defined the valid tags and syntax for their use, that is, which tags can be used;
- SGML is recursive, that is, tags can contain tags;
- Attributes are possible; they are part of the tag (e.g.: `<h1 position=“center”>`);
- Start and end tags are possibles, they are not required: `<p>Paragraph text.</p>` (note: "/") is used in the end tag after the character ")"
<h2>Introduction</h2>

This document contains lists codes which have been assigned to an online database after assignment. Entries in the first list are arranged in alphabetical order by the code and consist of the source code followed by the name and address of the organization.

The code consists of a maximum of eight characters as follows:

- 1-2 Country prefix
- 3-4 City prefix
- 5-8 Organization portion of code
Introduction

This document contains lists codes which have been assigned to an online database after assignment.

Entries in the first list are arranged in *alphabetical order* by the code and consist of the source code followed by the name and address of the organization.

The code consists of a *maximum* of eight characters as follows:

* 1-2 Country prefix
* 3-4 City prefix
* 5-8 Organization portion of code
Like MARC, the development of SGML resulted in the introduction of new terms:

- instance: a block of text with tags from one DTD;
- to parser: validate the content of an instance
- entity reference: a series of characters that represents another character or character string (e.g., "&gt;" instead of "&lt;")
- empty tag: an SGML tag without data.
In 1995 a special working group began development on a DTD for MARC;

SGML and MARC experts were part of the group;

They made decisions on the MARC DTD to accommodate SGML;

They decided to defined rigorously SGML tags that followed the MARC-style tags for each MARC element: for example: `<mrcb245-a>`
The working group decided to treat MARC fixed-length elements with SGML attributes:

- `<mrcbldr-bd-06 value="j">`

Special “wrapper” tags were defined to group important MARC fields (for ex., 1XX, 2XX, 3XX)

One problem was how to encode MARC characters in MARC-SGML;

The decision: permit options.
Despite considerable publicity, the MARC user community did not rush to SGML;

Some libraries have experimented with the DTD;

A MARC-to-SGML conversion tool was developed (the tool used PERL scripts)

Due to the decision on tag style, the MARC DTD was very big, which was a source of problems.
**XML - eXTENSIBLE MARKUP LANGUAGE**

- Certain syntactical characteristics of SGML created problems for its implementation;
- SGML minimisation – the possibility of omitting certain final tags, and sometimes even initial tags;
- Empty tags resulted in ambiguity;
- XML – the extensible markup language resolved all these problems and facilitated parsing (syntactic analysis);
- XML is compatible with ISO 8879 and SGML.
XML BASICS

- Every XML start tag must be paired with a corresponding end tag: ex.: <p>paragraph</p>

- Empty tags, which only have attribute values, have a special syntax; the start tag begins with "</"; e.g.,</mrcbldr-br-06 value="j">

- SGML options are not part of XML, thus instances are more sure;

- Parsing of XML is simpler.
MARC XML

- The MARC SGML DTD was modified to conform to XML requirements;
- Existing conversion tools were adjusted to generate XML;
- Alternative simplified DTDs were created to reduce the dimensions of the DTD file and to improve MARC XML instanced;
- The new MARC XML DTD allow local tags;
- Available at: http://www.loc.gov/standards/marcxmxml///
MARC XML CONVERSION ARCHITECTURE

MARC 21 (2709) Records

MARC 21 (XML) Records

Tagging Transformations
Character Set Conversion
Dublin Core Records
MODS Records
Other XML Formats
HTML Output
MARC Validation

MARC & Mark-Up Languages
The development of SGML follows the growth of the Internet, protocols and browsers;

SGML and XML are powerful, but they are sometimes limited by the lack of meaning and style linked to the tags;

The HTML tag set is fairly simple, with a small number of universally understood structures, mostly textual;

HTML define structural tags (ex., <p> for «paragraph») and some functions (ex., links)
Almost all browsers understand the meaning of tags from the HTML DTD;

Not all browsers act the same, but most can display Web documents acceptably;

The HTML tag set is small, but Web site developers have created attractive resources online;

Style sheet technology is improving the use of HTML.
As a result of the need to give access to catalogs via the Internet, many libraries filter their cataloging data through HTML;

Most modern MARC systems provide an HTML view of their MARC data;

Mappings are made between MARC fields and subfields and HTML for interpretation by browsers;

Some systems even support the entry of MARC records by means of HTML/Web interfaces.
IMPLEMENTATION OF TAGGING WITH MARC

- The development of SGML, XML and HTML technology hasn’t stopped people from using MARC;
- The richness and flexibility of the MARC format are supplemented by the alternative SGML structure;
- The existence of thousands of MARC-based systems and the millions of MARC records favor the continued use of the traditional ISO 2709 record structure.
FUTURE OF THE ISO 2709 RECORD STRUCTURE

- It’s not clear if the ISO 8879 (SGML) structure will replace the ISO 2709 (MARC) structure already in place in most library systems;

- In the future, systems based on SGML/XML tags could certainly serve as alternatives;

- Recursion (tags within tags) in SGML and XML could enrich the existing flexibility of MARC;

- Currently, the library community if experimenting.
THE RISE OF METADATA IN THE WORLD OF INFORMATION

- With the progress of technology in libraries, other bibliographic agencies that create and use bibliographic records have joined the development effort;

- Suppliers of non-MARC metadata are studying the use of existing MARC data;

- MARC users are considering ways of using non-MARC data encoded with standards like XML.
XML SCHEMAS

- Schemas are part of the most recent developments in the area of XML;
- Schemas are vocabularies of shared XML elements;
- Schemas allow computers to apply structural rules to documents;
- They define the structure, content and semantics of XML documents;
- XML Schema 1.1 is now a W3C recommendation.
MARC XML “Slim” SCHEMA

- An XML schema for MARC 21 data;
- Supports XML tagging of MARC 21 records;
- MARCXML "Slim" is restrictive to the forms of MARC content designation;
- Alphabetic tags are permitted, as well as signs (ex. %) as subfield codes and locally defined data elements (9XX);
- Available at: http://www.loc.gov/standards/marcxml///
DUBLIN CORE AND METS

- Users outside the MARC community decided to use XML and schemas to handle their bibliographic data;
- Dublin Core: a project to create a brief schema with 15 basic (core) elements;
- METS - Metadata Encoding & Transmission Standard: non-MARC schema for data relative to objects in a digital library.
DUBLIN CORE ELEMENTS:

<table>
<thead>
<tr>
<th>Title</th>
<th>Creator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>Description</td>
</tr>
<tr>
<td>Publisher</td>
<td>Contributor</td>
</tr>
<tr>
<td>Date</td>
<td>Type</td>
</tr>
<tr>
<td>Format</td>
<td>Identification</td>
</tr>
<tr>
<td>Source</td>
<td>Language</td>
</tr>
<tr>
<td>Relation/Link</td>
<td>Coverage</td>
</tr>
<tr>
<td>Copyright</td>
<td></td>
</tr>
</tbody>
</table>
See the Dublin Core Metadata Initiative web site:

http://dublincore.org/

Mappings from MARC to DC were developed by the Library of Congress; this documentation is available at:

http://www.loc.gov/marc/marc2dc.html

http://www.loc.gov/marc/dccross.html
The METS DTD structures XML data into five (5) main section of tags;

- Descriptive metadata (MARC elements are here)
- Administrative metadata (about the machine files)
- File groups (for files relating to digital objects and electronic resources)
- Structural plan (essential for the structure of digital objects)
- Behavior (of software associated with digital objects)
The Library of Congress, maintenance agency for METS;

As a schema, METS is going through a test period;

A special METS Web site is available at:
- http://www.loc.gov/standards/mets/

This Web site has links to documentation and tools that can be used with MARC and METS data.
MODS - Metadata Object Description Schema

- MODS is another XML schema that defines a set of bibliographic data elements;
- It was created for various uses, in particular for use by non-MARC library applications;
- Instances contain data taken from MARC 21 records, but the MARC 21 list of data elements is not required in order to make use of the schema;
- MODS is compatible with MARC 21, essentially a subset of MARC 21 data elements.
<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>titleInfo</td>
<td>note</td>
</tr>
<tr>
<td>name</td>
<td>subject</td>
</tr>
<tr>
<td>ofTypeResource</td>
<td>classification</td>
</tr>
<tr>
<td>genre</td>
<td>relatedItem</td>
</tr>
<tr>
<td>originInfo</td>
<td>identifier</td>
</tr>
<tr>
<td>language</td>
<td>location</td>
</tr>
<tr>
<td>physicalDescription</td>
<td>accessCondition</td>
</tr>
<tr>
<td>tableOfContents</td>
<td>recordInfo</td>
</tr>
<tr>
<td>targetAudience</td>
<td></td>
</tr>
</tbody>
</table>
EXAMPLE OF MODS SUBELEMENTS: "titleInfo"

- titleInfo [required]
  - title
  - subTitle
  - partNumber
  - partName
  - nonSort

- "titleInfo" attributes in the MODS DTD
  - ID, type (abbreviated, translated, alternative, uniform), authority, displayLabel, xlink (to the authority record), xml:land, script, transliteration
SOURCE OF INFORMATION ON MODS

- The Library of Congress is the maintenance agency for the MODS DTD;
- Development of MODS is still ongoing; version 3.0 MODS is now available;
- A special MODS Web site is available at:
  - http://www.loc.gov/mods/
- The Web has links to documentation and tools that can be used with MARC and MODS.
MADS - Metadata Authority Description Schema

- MADS is another XML schema that defines a set of elements for authority records;
- It is intended for many users, in particular non-MARC library applications;
- Instances will contain data taken from MARC 21 records, but the MARC 21 list of elements is not required to use the schema;
- MADS is compatible with MARC 21, essentially a subset of MARC 21 data elements.
HIGHEST LEVEL MADS ELEMENTS

- authority
- refs
- note
- affiliation
- url
- identifier
- fieldOfActivity
- extension
- recordInfo

name
references and tracings
notes
affiliation
Internet identification (address)
identification
field of professional activity
extension/other information
record-level information
EXAMPLE OF MADS SUBELEMENTS: “authority"

- **authority** [required]
  - name
  - **titleInfo** (for a uniform title)
  - **topic** (subject headings)
  - **temporal** (chronological)
  - genre
  - **geographic**
  - **occupation** (profession)
SOURCE OF INFORMATION ON MADS

- The Library of Congress is the maintenance agency for the MADS DTD;
- MADS is still being tested; version 1.0 is now available;
- A special MADS Web site is available at:
  - http://www.loc.gov/mads/
- The Web site has links to documentation and tools that can be used with MARC and MADS data.
DEVELOPMENT STRATEGY

- The MARC (2709) record structure is still popular;
- The MARC 21 data element set (2000+ éléments) remains stable, flexible and is often used as a model;
- MARCXML (8879) seems to be the preferred alternative structure to MARC (2709);
- Conversion and bibliographic data validation tools are now being developed using MARCXML as a central intermediary data structure.
CONCLUSION

- Up to now no other standard for bibliographic data has surpassed MARC, especially among MARC 21 users;
- The MARC 21 data element list has helped the development and implementation of new technologies;
- New technologies supplement the traditional MARC (2709) record structure for migration of data to different environments and systems;
- Cooperation between the MARC and non-MARC communities is still very essential.
SOURCES OF INFORMATION ON MARC 21

U.S. Library of Congress
Network Development and MARC Standards Office,
Washington, DC 20540-4402, U.S.A.
Tel: +1-202-707-6237
Fax: +1-202-707-0115
Email: NDMSO@LOC.GOV
Web Page: HTTP://WWW.LOC.GOV/MARC/

To request technical documentation, contact:
Cataloging Distribution Service
Washington, DC 20541-4910, U.S.A.
Web Page: HTTP://WWW.LOC.GOV/CDS/