Objectives of this lecture

- Explore course concepts in the context of the world of digital information
- Highlight current developments that point the way forward
- Motivate you to think how these developments may affect your future professional life

Two areas of professional responsibility

- Connect users to information
- Instruct users to use tools and resources

- Both of these require awareness and knowledge of:
  - Available resources
  - Information organization practices
  - Tools to access those resources
  - Standards and technologies used by the tools
  - How the tools work

Expanding roles for info professionals

Some things change . . .

Many types of information professionals are now responsible for information organization

- Library cataloger
- Library/museum archivist
- Classificationist
- Thesaurus developer
- Indexer/abstractor
- Digital bibliographer
- Metadata specialist
- Standards developer
- Systems specialist
- Network specialist
- Information architect

Changes in technology and their use demand lifelong learning

Cataloging librarian

- responsible for coordinating all activities involved in providing bibliographic access and associated digital access to LRC hard copy, non-print, and digital material; supervision of copy cataloging, original cataloging, managing the bibliographic database
- minimum of 2 years of experience doing original cataloging of multiple formats; knowledge of current and emerging trends of technical services in academic library services/resources; knowledge and proficiency in using AACR2r, LC Classification, LCSH, SuDoc organization, MARC21 formats, and OCLC Connexion; knowledge of FRBR, RDA, Dublin Core and other emerging practices and standards

Metadata librarian

- investigates, evaluates, and recommends metadata systems and standards for use with digital collections; leads the development of metadata standards and best practices; provides recommendations on the types of metadata required for effective access to electronic publications; tracks developments of metadata standards and facilitates access to resources through metadata creation; implements suitable metadata schemes for digital projects, putting measures in place to ensure quality and consistency, in order to support ease of access to digital objects; creates templates and other tools to support metadata creation for digital objects; develops metadata crosswalks as necessary, implements them using automated data conversion routines as digital projects and collections warrant; provides training in the area of metadata implementation and creation for digital objects
- knowledge of national cataloging standards including MARC21, AACR2, LCSH, Dublin Core and other emerging descriptive metadata standards; knowledge of metadata implementation and management techniques, specifically in the context of Dublin Core; working knowledge of standard computer office applications
The library’s diminishing market share

- Think in terms of value-added services
  - What value do we add that save potential users time, money, effort, etc?
- We have valuable resources but are users using them?
  - Library catalog is being bypassed
  - Large allocation of budget for commercially provided resource (licensed databases, etc.)
- We make users use our systems that are not easy to use
  - Think of the various interfaces of the licensed databases
  - Are we driving them away?

Taylor’s categories of added value

- Ease of use
- Noise reduction
- Quality
- Adaptability
- Time-saving
- Cost-saving

A working definition

The processes, practices, and activities of describing and representing information content and containers, as well as identifying the connections and relationships between and among information containers, content, and the people responsible for the creation and/or production of the information.

These processes, practices, and activities serve information users by providing them with ways to identify, locate, access, retrieve, and make judgments about information in response to their information needs.

Does this change in the networked environment?

Challenges of information organization

Some things remain the same . . .

- Goal is to connect users to information
- Representation is central concept
- Create representations that serve users
- Accept that there is not one right representation
- Build upon past understanding and practices

BUT

- The digital environment with its more complex and multifaceted information objects will provide new opportunities and challenges

But some things do change...

Technology and Buckland’s 3 types of libraries

- Paper Library
  - library materials based on paper
  - paper-based technical operations
- Automated Library
  - library materials primarily based on paper
  - computer-based technical operations
- Electronic Library
  - library materials primarily electronic/digital
  - computer-based technical operations

Information Organization Horizons

What the Web has wrought...
- Increased user expectations for access
- Decreased user expectations for precision
- Increased user expectations for noise
- Decreased user patience with difficult interfaces
- Increased user expectations for global access
- Increased amounts of information in various/new formats
- Decreased order in the bibliographic universe
- Increasing invisibility of digital resources

Networked information discovery & retrieval
- Networked Information Resources
  - Digital objects, collections of digital objects, or information services on the network
- Resource Description
  - Representing and describing resources
  - Metadata
- Resource Discovery
  - Searching distributed databases of different resources
  - Complex collection of searching activities
  - Metadata

The challenge of terminology
- Networked information
- Resource discovery
- Metadata
- Qualifiers
- Extensibility
- Crosswalks
- Interoperability
- Protocols
- Semantics
- Syntax
- Schemas
- DTDs
- XML
- RDF
- Your favorite buzzterm!

Models of resource description
- Full-text indexing approach:
  - Completely automated
  - Assumes that the resource is available for indexing
- Metadata approach:
  - A structured representation that describes the resource
  - Manual (or semi-automated) creation
  - Automated generation
- Self-describing resources
  - Extensible Markup Language (XML) provides a potential solution
  - Automated extraction of structured metadata

Librarians no longer solely create and define the terms, concepts, standards, and technologies that drive information organization and access in the networked environment.
Information Organization Horizons

**Markup of data and metadata**
- HyperText Markup Language (HTML)
  - Syntax: `<SomeTag>`some text`<EndSomeTag>`
- Extensible Markup Language (XML)
  - Syntax: `<SomeTag>`some text`<EndSomeTag>`
- Machine-Readable Catalog record

**HyperText Markup Language**
- Defines a set of tags and syntax for preparing a document to be rendered by a web browser
- Syntax:
  - `<SomeTag>`some text`<EndSomeTag>`
- Tags:
  - http://www.w3.org/TR/html4/index/elements.html
- Tags, except for Title, are about the structure of the document

**HTML example**
- `<H1>Forest elephants</H1>`
- `<P>In this section, we discuss the lesser known forest elephants. ...</P>`
- `<H2>Habitat</H2>`
- `<P>Forest elephants do not live in trees but among them.</P>`

**MARC as markup language**
- Machine-Readable Catalog record
- Provides content designation in the form of numeric tags, indicators, subfield codes to indicate the “meaning” of data
- 00741cam 2200205a 45000010030000000804100130001017001540100270007104003009008050017001700720010014401002700154245005050018126004030231300027002744400290030150090125003303004005404563002005090AA-0404 69069s1995 mdu b 000 0 eng d

**MARC example**
- 001 1101757
- 008 960200a1995 mdu b 000 0 eng d
- 035 $a (OCoLC)ocm34164430
- 035 $9 AFN2212CU
- 040 $a NCS $c NCS $d CUY $d ISM
- 043 $a n
- 049 $a ISMM
- 090 $a Z674.8 $b .M63 1995
- 092 $a 005.758 $b N277i, guide
- 100 1 $a Moen, William E.
- 246 30 $a Z39.50
- 246 75 $a ANSI/NISO Z39.50-1995
- 300 $a 12 p. ; $c 28 cm.
- 500 $a Cover title.
- 504 $a Includes bibliographical references (p. 11-12).
- 500 $a Library information networks $x Standards $x United States $x Guidebooks.
Extensible Markup Language -- XML

- Uses basic markup conventions of HTML
- Allows users to use tags defined by them
- Oriented to semantic markup of text
- Hierarchical tree structure

The fundamental purpose of good markup is to clarify meaning and facilitate intelligent use of information.

XML is a way of structuring data

Ron Gilmour, XML: A Guide for Librarians

XML example

```
<menu>
  <menu_item>
    <item_name>Subgum Wonton</item_name>
    <item_description>Crabmeat, sliced chicken, roast pork, shrimp, lobster, imported straw mushrooms, bamboo shoots, waterchestnuts and Chinese veg. baby corn, served with crispy wonton.</item_description>
    <price>$9.15</price>
  </menu_item>
</menu>
```

XML is the (near) future

- Basic syntax for data in the web environment
- Critical for information and metadata exchange

Library use of XML

- MARCXML
- MODS (Metadata Object Description Schema)
- METS (Metadata Encoding and Transmission Standard)

Resource Discovery

- Resource discovery consists of...
  - Finding
  - Identifying
  - Selecting
  - Accessing/Acquiring
  - Other tasks?
- Using representations or surrogates for networked information resources...
  - People
  - Organizations
  - Products
  - Services
  - Texts
  - Images
  - Sounds
  - Collections of objects

The Web

Centralized

- Single information retrieval system
- Harvesting or collecting resources
- Central indexing
- Tight integration of user interface and IR system
- Typically full-text indexing

Decentralized

- Multiple databases/IR systems
- Search multiple resources as single logical resource
- Local indexing/search logic
- Results come from source databases
- Typically well-structured metadata
Centralized Model of Resource Discovery

- Digital Info Resources
  - HTML Resources
  - Other Digital File Types
- Resources Gathered & Represented (full text)
- Access Mechanisms
- Search Engine

Decentralized Model of Resource Discovery

- Digital & Analog Info Resources
  - Printed Materials
  - HTML Resources
  - Digital Images
  - Databases
- Represented by Structured Metadata
- Metadata Access Mechanisms
- Search Interface

World’s Largest Information Database

What happens if Google...

- Acquires or licenses for global access the key commercial information resources
- Indexes the resources
- Provides single, easy to use search interface to all those resources
- Charges $10/month for users to have access to all of that
- Who will use our hard-to-use resources with all those different interfaces?
- Google Scholar – Advanced Search

Metalsearch (or federated search)

- Single search interface
- Concurrent searching of two or more resources
- Uses various technologies
  - Standards such as Z39.50 information retrieval protocol; Search and Retrieve Web Service
  - Proprietary Connectors (e.g., WebFeat, Muse Global)
- Helps users get started discovering resources

The Networked Information Landscape According to Google

- Open Web
- Licensed Databases
- WorldCat
- Digital Repositories
- Digitized Books
- Google

Search Interface
Metadata harvesting: OAI-PMH

- Open Archives Initiative Protocol for Metadata Harvesting
  http://www.openarchives.org/
- Defines a protocol for harvesting metadata from repositories
  - Partitions the world into:
    - Data providers
    - Service providers
  - Uses Dublin Core Metadata Element Set as standard metadata representation for exchange
  - Uses XML for exchanging the metadata records
OAIster

- **A union catalog of digital resources**
- Contains nearly 11,000,000 records describing freely-available and restricted-access digital resources
- Uses the Open Archives Initiative Protocol for Metadata Harvesting
- Harvets the descriptive metadata (records) and makes those searchable
- Currently harvesting from over 700 digital repositories

Digital repositories

- Digital repository (sort of a generic term)
  - An application that stores and makes available digital resources and their associated metadata
- Image repository
  - (e.g., http://pro.corbis.com/)
- Learning objects repository
  - (e.g., http://carea.ucalgary.ca/cgi-bin/WebObjects/CAREO.woa?theme=careo)
- Data repository
  - (e.g., http://www.public.asu.edu/~huanliu/DHub/bioinformatics.html)
- Institutional repository
  - (e.g., http://txspace.tamu.edu/)
- Differentiated by
  - Types of objects
  - Types of metadata
  - Purpose
  - …

Repositories – The technical side

- Database component
- **Metadata component**
- Search and browsing component
- Web interface component
- Submission component
- Administration component
- …

Institutional repositories

- A repository application
- Purpose
  - To make available the scholarly and research output of an academic organization
- Characterized by
  - Organizational commitment to long-term stewardship
  - Open access

Potential contents for IRs

- Pre-prints and post-prints
- Technical reports, working papers
- Theses & dissertations
- Books or chapters of books
- Conference proceedings
- Presentations
- Sound and video files
- Digital research materials (e.g. simulations, code)
- …

Metadata – The key

- Boundaries between information communities are porous
- The world will not be made up of MARC
- Many metadata schemes:
  - To describe and manage resources
  - Provide structured representations of the resources that can be processed by machines
  - Serving needs of different information communities
- Typically using Extended Markup Language (XML)
  - Syntax for encoding metadata for exchange and reuse

I've often said librarians should like any metadata they see.  
(R. Tennant)
Metadata supports operations on resources

- Know what resources are available
- Know how to play a resource
- Know provenance of a resource
- Know what use policy governs a resource
- Know how to ingest a resource
- Know how to interact with a resource
- Know how to compose/decompose resources
- ...

And relieves the user of having to have advance knowledge of the characteristics or existence of the resource.

(from Lorcan Dempsey 2004)

Dempsey’s acronymic density or this is the present future!!

- Metadata schemes: DC, MODS, CDWA, VRA, etc.
- Metadata content standards: AACR, CCO, DACS, etc.
- Metadata encoding standards: MARC, XML, RDF, etc.
- Metadata container/wrapper standards: METS, MPEG, etc.
- Discipline specific metadata schemes: GILS, CSDGM, GEM, IEEE-LOM, etc.
- Other schemes of interest: TEI, EAD, etc.

Current state of affairs

- Many metadata schemes and element sets
  - Well known & documented (e.g., AACR2, ISBD)
  - Less known and little public documentation
- Multiple content standards
  - Well known & documented (e.g., AACR2)
  - Less known &/or emerging (e.g., CCO)
- Similar/same content described by different metadata schemes and vocabularies
  - No canonical metadata record for an object
- Varied syntaxes for encoding metadata
  - No canonical syntax (e.g., MARC, XML)

Current state of affairs

- A vital and diverse metadata ecology!
- Which reflects:
  - Community practices, needs, meaning
  - Cost barriers to adopting common standards
  - Lack of knowledge of available standards
  - Not invented here syndrome

Interoperability

System-oriented definition:
- The ability of two or more systems or components to exchange information and use the exchanged information without special effort on either system

User-oriented definition:
- The condition achieved when two or more technical systems can exchange information directly in a way that is satisfactory to users of the systems (AAP)

More of our present future

- A complex information landscape in the networked information environment
- A complex metadata ecology with diverse life forms (L. Dempsey)
- Metadata choices reflect community needs and practices
- Libraries are one community among many; we are engaged with multiple metadata schemes already
- Users interacting with multiple metadata records
- Our systems need to enable “ingesting, merging, indexing, enhancing, and presenting to the user, metadata from a variety of sources describing a variety of objects”. (R. Tennant)
Information Organization Horizons

Users organizing the resources

- Folksonomies
- User tagging
- Social bookmarking

The changes are here and coming

- The library as network centric not building centric
  - A node on the network in competition for users
- The library catalog as one metadata repository
  - A rich repository of detailed metadata, which needs to interact with other network systems
- The network centric library
  - Exposes, transforms, reuses, and aggregates metadata
  - Supports interoperability of metadata
- The library cataloger as metadata maven
  - Organizing and managing a wide variety of resources in multiple repositories with different metadata schemes
  - Digital libraries
  - Institutional repositories
  - Image databases
  - ...

Concepts and terms

- Automated extraction
- Automated indexing
- Digital library
- Digital repositories
- Extensible markup language
- Folksonomies
- Full-text indexing
- Harvesting
- Interoperability
- Metadata repositories
- Metasearch
- Networked information environment
- Networked information resources
- Open archives metadata harvesting protocol
- Protocol
- Resource description
- Resource discovery
- Self-describing resources
- Standards
- Z39.50

Links

- Aquabrowser – offered by TLC systems
  - http://www.acornweb.org/
- Endeca – at NCSU libraries
  - http://www.lib.ncsu.edu/catalog/
- OCLC’s new Worldcat
  - Via Firstsearch (UNT Libraries)
  - http://www.worldcat.org/
- Google Scholar
  - http://scholar.google.com/ (back to Worldcat)
- Social tagging – del.icio.us
  - http://del.icio.us/

Links

- Metasearch
  - http://libraryoftexas.org/ (via UNT E-resources)
  - http://mikey.indexdata.com/demo/
- Institutional repositories
  - https://drum.umd.edu/dspace/index.jsp
  - http://bospaces.tamu.edu/
  - http://labs.di.tamu.edu:8080/geofolios/handle/123456789/2
- Metadata harvesting
  - http://www.oaister.org/