Show us where you are!

On the toolbar, click on the yellow star, then click on your location!

Metadata for Digitization and Preservation

Using this software

- Microphone
- Raising your hand
- Green ✓/ Red ✗
- Laughing / Clapping
- Stepping out

- Text chat
- Feedback
- Audio
- Full Screen
- Exiting

Hand   Yes   No   Laugh   Applaud   Step Out   Text Chat   Feedback   Audio   Full Screen

soliNET

Agenda

- Metadata Overview
- Metadata Schemas
- Descriptive Schemas
- Creating Dublin Core Metadata
- Technical/Administrative Schemas
- Preservation Metadata
- Encoding Schemas
- Wrap Up

soliNET

Metadata Overview

- What is metadata?

  - A neutral term for cataloging without the excess baggage of the AACR2 or the MARC formats. (Priscilla Caplan)

  - The sum total of what one can say about any information object at any level of aggregation. (Anne J. Gilliland-Swetland)

soliNET

Data about data?? Well...

- Metadata is data that facilitates the management, description, and preservation of a digital object or aggregation of digital objects. (Judith Cobb)

soliNET

Creating Metadata

- The creation of metadata is governed by a body of standards, best practices and schemas that, when appropriately applied, work together to facilitate the management, description, and preservation of digital objects.

---

soliNET

Traditional Functions

- Acquire material into collection
- Describe material
- Manage the collection
  - Individual items and the entire thing
- Provide Access

---

soliNET

Traditional Functions

- Traditionally, we applied these functions to:
  - Paper based and microform based information resources
    - Monographs, serials, photographs, etc.

- Access provided through local library services
  - Including ILL

New Functions

- Apply these functions to:
  - Web documents, online serials, digital images, digital collections, web sites, digital audio and video, born digital material, etc.

- Access provided via the web and email

Why are digital objects different?

- Information explosion
- Instant access
- Less physical control over the collection
- Some are surrogates
- Increased user expectations
- Preservation is more complex

Why are they the same?

- The traditional functions still apply
  - Acquire
  - Describe
  - Manage
    - Individual item at the collection level
  - Provide Access

soliNET

---

soliNET

Metadata and Users

- Functional Requirements of Records
  - Using data to find materials
  - Using data retrieved to identify materials
  - Using data to select appropriate materials
  - Using data to acquire or obtain access to materials
  - International Federation of Library Associations and Institutions
    - http://www.ifla.org/VII/s13/frbr/frbr.htm

---

soliNET

About Metadata

- Dozens of metadata standards
- Intellectual and physical access
- Increased access
- Beyond description
- Integration
- Generated automatically, manually – or both

About Metadata Schema

- Imposing structure on unstructured information
- Composed of elements
- Standardized presentation and content
- Consistent presentation
- Goal: to use the structured information for access, administration, management, and preservation of resources.

Characteristics of Metadata

- Source
- Method of creation
- Nature
- Status
- Structure
- Semantics
- Level

Source of Metadata

- Internal
  - File names and header information
  - Directory structures
  - File format and compression scheme
- External
  - Cataloging records
  - Rights and other legal information

Method of Metadata Creation

- Automatic
  - Keyword indexes
  - User transaction logs
- Manual
  - Catalog records
  - Dublin Core metadata

Nature of Metadata

- Lay
  - Metatags for a personal Web page
  - Personal filing systems
- Expert
  - Specialized subject headings
  - MARC records
  - Archival finding aids

Status of Metadata

- Static
  - Title, provenance, date of creation
- Dynamic
  - Directory structure, user transaction logs
- Long-term
  - Technical format, rights information, preservation management documentation
- Short-term
  - Transactional

soliNET

Structure of Metadata

- Structured
  - MARC
  - TEI
  - EAD
  - Local database formats
- Unstructured
  - Note fields
  - Annotations

soliNET

Semantics of Metadata

- Controlled
  - AACR2
  - AAT

- Uncontrolled
  - Free-text notes
  - HTML metatags

soliNET

Level of Metadata

- Collection
  - Collection level record (e.g. MARC record or finding aid)
  - Specialized index

- Item
  - Transcribed image captions and dates
  - Format information

soliNET

Types of Metadata

- Descriptive
- Technical/Administrative
- Preservation
- Encoding

---

soliNET

Descriptive Metadata

- Metadata used to describe or identify information resources
  - Cataloging records
  - Finding aids
  - Indexes
  - Hyperlinked relationships between resources

---

soliNET

Technical/Administrative Metadata

- Metadata used in managing and administering information resources
  - Acquisitions information
  - Rights and reproduction tracking
  - Hardware and software documentation
  - Digitization information
  - Authentication and security data
  - Tracking of system response times

Structural Metadata

- Describes internal organization of a resource
- Ties together components of a multimedia entity
- Documents order and format of data elements

Preservation Metadata

- Metadata related to the preservation management of information resources
  - Documentation of physical condition of resource
  - Documentation of actions taken to preserve physical and digital versions of resource

Metadata Schema

- AACR2
- Dublin Core
- Visual Resources Association (VRA)
- Metadata Object Descriptive Schema (MODS)
- Text Encoding Initiative (TEI)
- Encoded Archival Description (EAD)
  And many, many more!

soliNET

Encoding Standards/Schema

- Hypertext Markup Language (HTML)
- Machine-Readable Cataloging (MARC)
- Metadata Encoding Transmission Standards (METS)
- Resource Description Framework (RDF)
- Extensible Markup Language (XML)
- Z39.50

---

soliNET

Metadata Schema and Encoding Schema

- Metadata schema = rules
- Encoding schema = representation

---

soliNET

Digitization and Metadata

- Descriptive metadata for access and administration
- Technical metadata for preservation
- Structural metadata for control over complex digitized objects
- Preservation metadata for management within a digital archive

Choosing a Schema

- Define the descriptive and administrative information you need, then worry about schema
- Is sustainability and preservation your goal? Then administrative/technical metadata will be necessary.
- Interoperability and collaboration considerations

Choosing a Schema: Interoperability

- Open Archives Initiative (OAI)

- Common 3 elements harvested and made searchable from one interface
  - Title, Creator and Description

Choosing an Encoding Schema

- And do you really need one?
- For web access and searchability, your metadata will be created and stored in a database or content management system
- Do you need more than that?

Choosing an Encoding Schema

- The more digitized objects you have, the more complex they are, the more data sharing you do, the more important it will be to utilize an encoding schema
- XML is the most prevalent encoding schema
  - All metadata schema have XML based encoding schema already available

Factors in Metadata Decisions for Digitization Projects

- Audience
- Workflow and timelines
- Technical skills
- Preservation
- Interoperability
- Number of and complexity of digitized objects

Metadata for Digitization

- What information do you need to gather as you select the material?
- What information do you need to create when you scan the source?
- What do you need to know to effectively manage the collection?
- How do you want to provide access?
- How do you want to preserve the digital collection?

soliNET

Metadata for Digitization Projects

- How will all this information be acquired?
- What metadata schema is appropriate? Will it need to be modified?
- Where will it be stored?
- How will it be stored? Who will manage the database?
- How do you envision the web interface? Does that impact metadata decisions?

soliNET

Descriptive Metadata

- Information users will have to gain access to the digitized material
- Should facilitate access to the original source material whenever possible
- Access via a web interface search engine
- User friendly
- Standardized
- Well written

soliNET

Common Descriptive Metadata Schema for Digitization Projects

- Visual Resources Association
- Metadata Object Descriptive Schema
- Encoding Archival Description
- Text Encoding Initiative
- Dublin Core

VRA Core
http://www.vraweb.org/projects/
vracore4/index.html

- Modeled on Dublin Core
- Elements are optional and repeatable
- Single element set
- Describe *works* of visual culture and the *images* that document them.

Encoded Archival Description
http://www.loc.gov/ead/

- Metadata standard for encoding archival finding aids
- Designed as electronic finding aid to resources either electronic or physical
- Created by archivists
- Able to accommodate multiple levels of description
- Strength is its high degree of specificity

MODS
http://www.loc.gov/standards/
mods/

- Is an XML schema designed to be used with METS
- Includes a subset of MARC elements – language based rather than numerical tags
- For digital library objects that require rich descriptions compatible with existing ones in library catalogs

Text Encoding Initiative
www.tei-c.org

- Designed to represent textual materials in electronic form
- Well documented and has set of guidelines and rules for application.
- Mandatory TEI Headers can be integrated with the electronic text or may be separated and used in a database or catalog.

Characteristics of the Dublin Core
http://www.dublincore.org

- All elements optional
- All elements repeatable
- All elements can be displayed in any order
- 15 elements, 10 attributes
- Extensible
- International

15 Dublin Core Elements

- Title
- Creator
- Subject/Keywords
- Description
- Publisher
- Contributor
- Date
- Resource Type

- Format
- Resource Identifier
- Source
- Language
- Relation
- Coverage
- Rights Management

10 Attributes from ISO/IEC11179 Standard

- Name
- Identifier
- Version*
- Registration authority*
- Language*
- Definition
- Obligation*
- Datatype*
- Maximum occurrence*
- Comment

6 Common Attributes

Version: 1.1
Registration Authority: Dublin Core Metadata Initiative
Language: en
Obligation: Optional
Datatype: Character String
Maximum Occurrence: Unlimited

DC:Title

- Name: Title
- Identifier: Title
- Definition: A name given to the resource
- Comment: Typically, a Title will be a name by which the resource is formally known.

DC: Creator

soliNET

- Name: Creator
- Identifier: Creator
- Definition: An entity primarily responsible for making the content of the resource
- Comment: Examples of a Creator include a person, an organization, or a service.

DC: Contributor

soliNET

- Name: Contributor
- Identifier: Contributor
- Definition: An entity responsible for making contributions to the content of the resource
- Comment: Examples of a contributor include a person, an organization, or a service.

DC: Publisher

soliNET

- Name: Publisher
- Identifier: Publisher
- Definition: An entity responsible for making the resource available
- Comment: Examples of a contributor include a person, an organization, or a service.

DC: Date

- Name: Date
- Identifier: Date
- Definition: A date associated with an event in the life cycle of the resource
- Comment: Typically, Date will be associated with an event in the life cycle of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 and follows the YYYY-MM-DD format.

DC: Relation

- Name: Relation
- Identifier: Relation
- Definition: A reference to a related resource
- Comment: Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.

DC: Source

- Name: Source
- Identifier: Source
- Definition: A reference to a resource from which the present resource is derived.
- Comment: The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.

DC: Description

- Name: Description
- Identifier: Description
- Definition: An account of the content of the resource
- Comment: Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.

DC: Subject

- Name: Subject and Keywords
- Identifier: Subject
- Definition: The topic of the content of the resource.
- Comment: Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

DC: Type

- Name: Resource type
- Identifier: Type
- Definition: The nature of genre of the content of the resource
- Comment: Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the working draft list of DC Types) To describe the physical or digital manifestation of the resource, use the FORMAT element.

DC: Format

soliNET

- Name: Format
- Identifier: Format
- Definition: The physical or digital manifestation of the resource
- Comment: Typically, Format may include the media type or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Recommended best practice is to select a value from a controlled vocabulary.

DC: Indentifier

soliNET

- Name: Resource identifier
- Identifier: Identifier
- Definition: An unambiguous reference to the resource within a given context.
- Comment: Recommended best practice is to identify the resource by a string or number conforming to a formal identification system such as URI, URL, DOI, and ISBN.

DC: Language

soliNET

- Name: Language
- Identifier: Language
- Definition: A language of the intellectual content of the resource
- Comment: Recommended best practice for the Language element is defined by RFC 1766 which includes a two-letter language code (taken from the ISO 639 standard) followed optionally by a two letter Country Code (taken from the ISO 3166 standard).

DC: Coverage

- Name: Coverage
- Identifier: Coverage
- Definition: The extent or scope of the content of the resource
- Comment: Coverage will typically include spatial location, temporal period, or jurisdiction. Recommended best practice is to select a value from a controlled vocabulary (such as the Thesaurus of Geographical Names) and use name places or time periods in preference to numeric identifiers such as sets of coordinates or date ranges.

DC: Rights

- Name: Rights management
- Identifier: Rights
- Definition: Information about rights held in and over the resource
- Comment: Typically, a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights, Copyright, and various Property Rights.

Extensibility

- Refining mechanism for elements
  - Improve sharpness of description with qualifiers
- Means for extending element set
  - Complementary packages of other types of metadata (administrative, rights management, discipline-specific, etc.)

Considerations

- Project metadata style guide
- Workflow
- Quality assurance

Implementation

- Metadata decisions
  - What informational elements do you need?
  - What schema will you use?
  - How will you use it?
  - Create style guide for your project

Implementation

- Staff creates the metadata manually according to style manual and established work processes
- Metadata is reviewed for quality
- Metadata is stored and made web accessible via
  - Digital management software (CDM)
  - Database
  - XML

soliNET

Descriptive Metadata Examples

- Ohio Memory
- http://www.ohiomemory.org
- American Memory, Ansel Adams Collection
- http://memory.loc.gov/ammem/aamhtml/aamhome.html
- University of Washington
- http://content.lib.washington.edu/aipnw/index.html

soliNET

Administrative/Technical Metadata

- Information file that facilitates management and preservation of the file
- Technical information about:
  - Master file
  - TIFF
    - Scanning specifications
    - Storage

soliNET

Administrative/Technical Metadata

- NISO Technical Metadata for Digital Still Images
  - http://www.niso.org/standards/resources/Z39 87 trial use.pdf

- If scanning for short term access only, this is not necessary

soliNET

Tagged Image File Format – Background and Metadata

- TIFF is a specification for a file format. The specification is available at http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf
- Spec includes a "directory" or "header" section which consists of several metadata fields
- A TIFF can consist of several images

soliNET

Tagged Image File Format – Background and Metadata

- The TIFF spec is implemented differently by different applications
- Scanning software
  - Usually bundled with your scanner
  - Controls scanner or camera and passes information to computer storage of image editing software
  - Outputs image files in specific image file formats
  - Determines what "flavor" TIFF is produced
  - Determines what metadata fields are utilized and how they are utilized

soliNET

Tagged Image File Format – Background and Metadata

- Other software may add to the TIFF metadata, such as Photoshop
- Tags can be added using particular software
  - Black Ice Software Development Kit
    http://www.blackice.com
  - Captiva's Input Accel
  http://www.captivacorp.com/products/InputAccel.asp

soliNET

NISO Metadata
http://www.loc.gov/standards/mix

- Purpose: To define a standard set of metadata elements for digital images
- Facilitate interoperability
- Support long term management of and continuing access to digital images

---

soliNET

NISO Metadata

- Encoding
- Production
- Assumptions
- XML Schema is available at http://www.loc.gov/standards/mix

---

soliNET

NISO Metadata

- More information is better
- Can add to the cost of the project in the short term
- Will facilitate "scan once" philosophy and save resources in the long run
- Work with software scanning vendor, IT staff as necessary

soliNET

Implementation

- Some of the metadata can be captured at the time of scanning
- You will need to find out the metadata capabilities of your software
- Stored in TIFF header or extracted for external storage

soliNET

Using Dublin Core for Technical Metadata Elements

- Use **Format** element
- Include information such as:
  - File size
  - Quality
  - Extent
  - Compression
  - Checksum value
  - Object producer
  - Creation hardware
  - Creation software

soliNET

Preservation Metadata

- Based on Open Archival Information System (OAIS)
- If you are digitizing with preservation in mind, ALL metadata is preservation oriented
- Metadata must be of the highest quality possible
- Incorporate the creation and management of metadata into your project at the planning stage

soliNET

Preservation Metadata

- Designed to facilitate the process of preservation
- Generally implemented at the time a digital resource is moved to a digital archive
- Several schemas under development for particular operating environments and/or programs
- There is some overlap between descriptive and technical schemas and these schemas

soliNET

Preservation Metadata Sets

- CEDARS – Consortium of University Research Libraries, Exemplars in Digital Archives project
  - Technical, administrative, descriptive, legal elements
  - Does not depend on a particular preservation strategy

http://www.leeds.ac.uk/cedars/indexold.htm

soliNET

Preservation Metadata Sets

- NLA National Library of Australia
  - Does not depend on a particular preservation strategy
  - Resource discovery not a factor

http://www.nla.gov.au/

soliNET

Preservation Metadata Sets

- NEDLIB Networked European Deposit Library
  - Minimum requirements for preservation
  - As generic as possible
  - Does not depend on a particular preservation strategy

  http://www.kb.nl/coop/nedlib/

soliNET

Preservation Metadata Sets

- OCLC Digital Archive
  - Descriptive, technical/structural, and administrative elements
  - Does not depend on a particular preservation strategy

  http://www.oclc.org/digitalarchive/

soliNET

Preservation Metadata Sets

- Inference that there is a core of metadata necessary for preservation regardless of the preservation strategy
- More work needs to be done to identify the particular elements necessary for particular preservation strategies

soliNET

Implementation

- Implementation will occur at the time a digitized resource is placed in a digital archive

---

soliNET

Encoding Schema

- Encoding
  - Best practice
  - Non-proprietary metadata
    - Facilitates data sharing
    - Facilitates preservation
    - Facilitates management
  - Good back-up protocol

---

soliNET

Encoding: XML

- Non proprietary standard mark-up language used to encode information
- Provides meaningful information about the data (much more than HTML formatting codes)
- XML requires technical resources
- Most standard metadata schemas have XML implementation information available

soliNET

What makes XML so important to metadata?

- Non-proprietary method of storing metadata
- Provides meaningful wrapper or envelope for metadata
- Facilitates the creation of a "master metadata record"

soliNET

How to use XML for Encoding Metadata

- Identify established XML encoding schema for your metadata schema
- Move your metadata to XML
- Document the schema you utilize
- If appropriate, utilize a more complex XML encoding schema such as METS

soliNET

Encoding: METS
http://www.loc.gov/standards/mets/ METSOverview.html

- Metadata Encoding and Transmission Standard
- Product of *Making of America* project
- Digital Library Federation Initiative
- Provides an XML schema for encoding metadata necessary for:
  - Management of digital library objects
  - Exchange of those objects (OAIS)

Encoding: METS

- Seven sections of a METS document:
  - Descriptive
  - Administrative
  - File Group/Section
  - Structural Map
  - Behavior
  - Structural Links
  - METS Header

METS Descriptive Metadata

- May point to descriptive metadata external to the METS document

- May imbed the descriptive metadata within the METS document

METS Administrative Metadata

- Information regarding file creation and storage
- Intellectual property
- Metadata regarding the original
- Information regarding provenance of the digital object
- May be externally or internally encoded

soliNET

METS File Groups

- Used to group together related files
- One file group lists all the files which comprise a single electronic version of the digital library object
  - Master (TIFF)
  - Access copy or copies
  - Perhaps a transcript

soliNET

METS Structural Map

- "The heart"
- Outlines the intellectual structure of the digital object
- Links elements of the object to content files and metadata that pertain to each element

soliNET

METS Behavior

- Used to associate executable behaviors with content
- Defines the behaviors
- Can contain executable code to run behaviors

soliNET

METS Structural Links

- Records existence of hyperlinks between nodes in Structural Map
- Used to archive Websites

---

soliNET

METS Header

- Contains metadata describing the METS document itself

---

soliNET

Bringing it to the Web

- EmbARK
  http://www.gallerysystems.com/eb_home.asp
- Cumulus
  http://canto.com/
- CONTENTdm
  http://www.oclc.org/contentdm/default.htm
- XMetaL
  http://www.corel.com/

Wrap Up

soliNET

- Descriptive metadata
  - Use established set like Dublin Core
  - Create project style guide to facilitate standardized, high quality creation
  - Store in content management software or database to provide web access
  - Encode and back up using XML

Wrap Up

soliNET

- Technical Metadata
  - Use TIFF
  - Use as much of the NISO draft standard as possible
  - Contact scanning software creator to see about use of TIFF header fields
  - Document which fields you are capturing in the TIFF header
  - Leave the metadata within the header

Wrap Up

soliNET

- Encoding schema
  - Not necessary for a digitization project to work
  - Best practice for back ups and preservation
  - Appropriate for complex digital objects and digital projects

Wrap Up

soliNET

- Plan for costs
- Create highest quality, most thorough possible
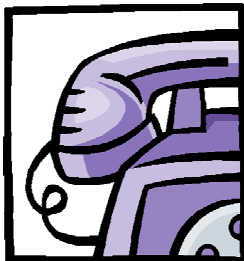- Good metadata is a key to digital preservation

---

Thank You for Attending!

soliNET

Questions?

SOLINET Help Desk
1.800.999.8558
helpdesk@solinet.net
8am-5pm Eastern
Monday-Friday

NELINET
1.800.NELINET