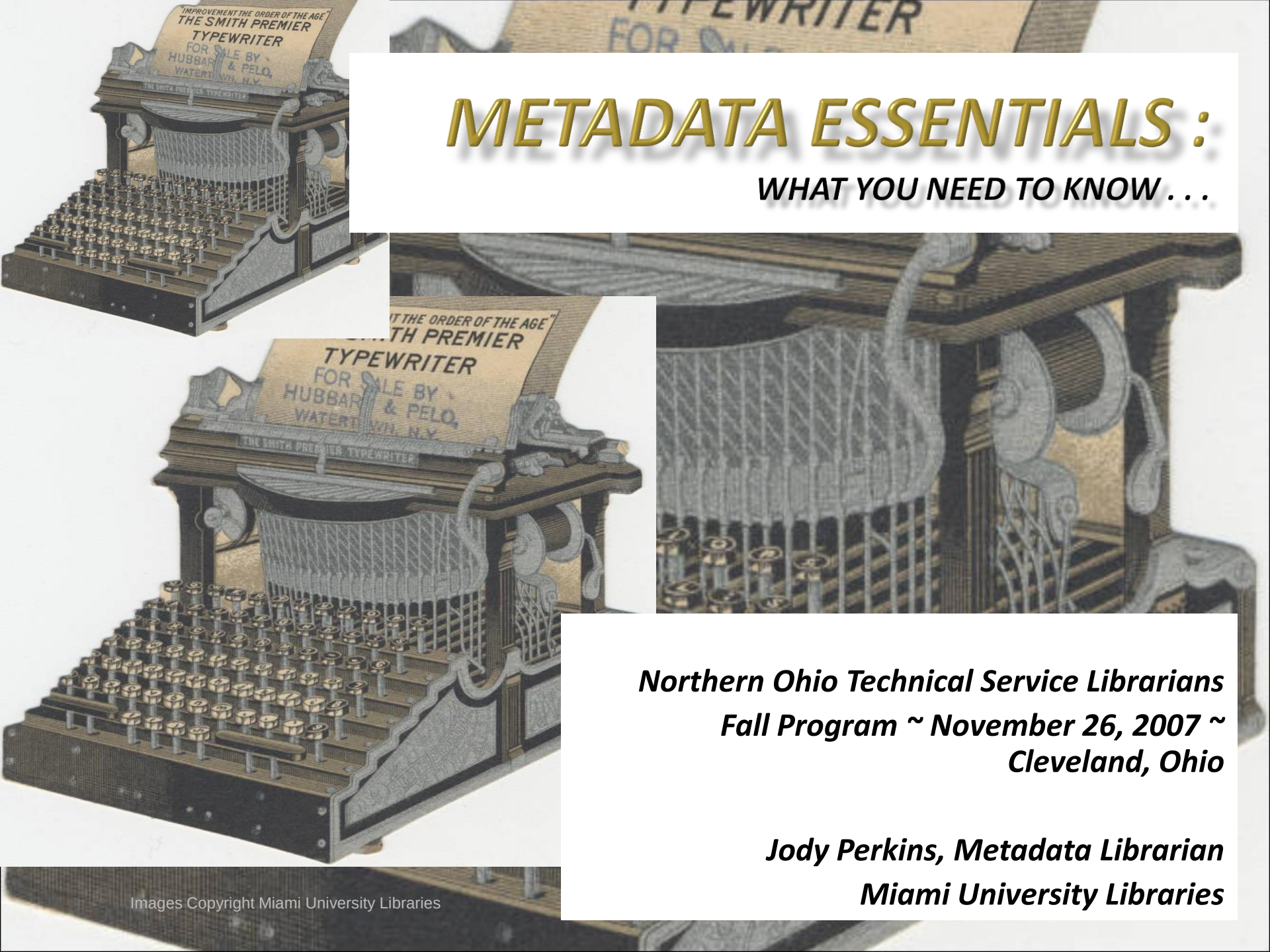


METADATA ESSENTIALS :

WHAT YOU NEED TO KNOW . . .

Northern Ohio Technical Service Librarians
Fall Program ~ November 26, 2007 ~
Cleveland, Ohio

Jody Perkins, Metadata Librarian
Miami University Libraries

Introductions

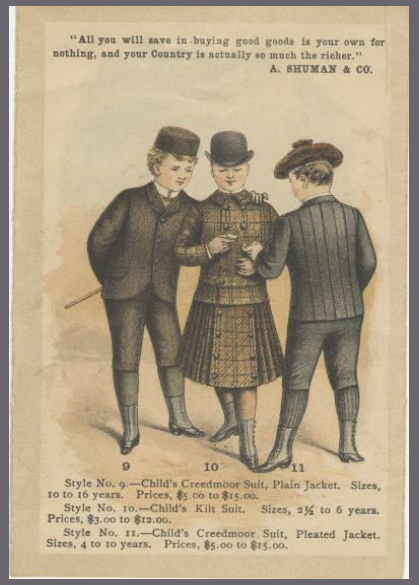
please . . .

Metadata Essentials:

what you need to know . . .

. . . in three parts:

- PART I : Building Your Toolkit
- PART II : Understanding Your Environment
- PART III : Making Choices



QUICK REVIEW OF CONCEPTS AND TERMS

What is Metadata?

- "Metadata consists of *statements we make about resources to help us find, identify, use, manage, evaluate, and preserve them.*"

- "Answers come from three traditions:
  - –Database Management Systems ("Schemas of relational databases")
  - –Library Cataloging Traditions (MARC & AACR2)
  - –The World Wide Web (since the mid-1990's)"

Stuart A. Sutton, Basic Semantics, International Conference on Dublin Core and Metadata Applications—Singapore, 2007 http://www.dc2007.sg/T1-BasicSemantics.pdf

Functions/Types of Metadata

- Administrative
- Descriptive
- Structural

Caplan, Priscilla. “Metadata Fundamentals for All Librarians”. American Library Association, Chicago, 2003. p. 3-5

Definitions

- Schema
- Element
- Vocabulary Encoding Schemes
- Granularity
- Application Profile
- Data Dictionary
- Crosswalks
- Interoperability

Woodley, Mary. “DCMI Glossary”
http://dublincore.org/documents/usageguide/glossary.shtml#S

Definitions continued . . .

- Element
- Schema
- Vocabulary Encoding Schemes
- Granularity

Definitions continued . . .

- Application Profile
- Data Dictionary
- Crosswalk

Data Dictionaries, Application Profiles and Crosswalks

- University of Washington DD Index
- Snyder Collection
- OhioLINK DMC Application Profile
- Colorado Digitization Project
- Metadata Standards Crosswalk (Getty)
- Library of Congress MARC to Dublin Core

Definitions continued . . .

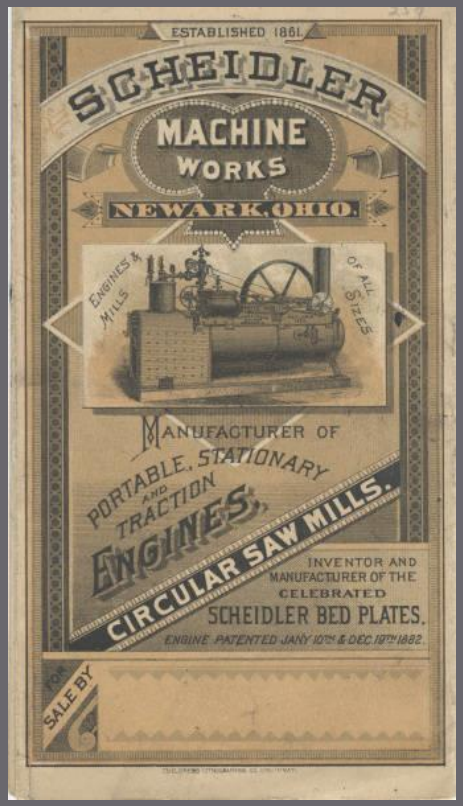
- Interoperability

  – the ability of different types of computers, networks, operating systems, and applications to work together effectively, without prior communication, in order to exchange information in a useful and meaningful manner.

- Three Types:

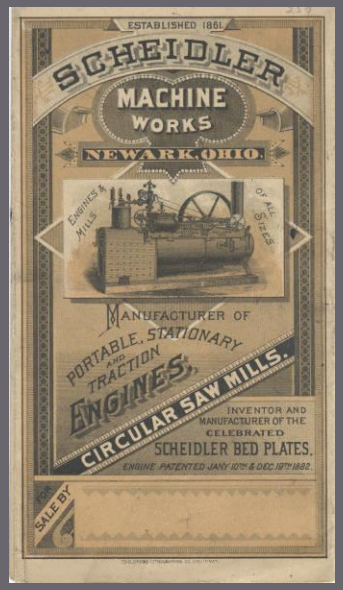
  Semantic, Structural, Syntactical

PART I :
Building Your Toolkit



Tools & Metrics

Basic Components



- Schemas
- Content Standards
- Controlled Values
- Standards for Encoding, Transmission and Harvesting
- Data Models

Metrics

- Best Practices
- Measures of Quality
- Shareable Metadata

Schemas

- General Purpose
  - Dublin Core
  - MARC, MODS
- Specialized
  - VRA Core
  - EAD
  - Many others . . .

Dublin Core

- Library community led it's development
- Originally designed for document-like internet resources
- Flat Element Set
- Facilitates Interoperability
- Syntax (encoding) Independent
- No "official" content standard
- 2 versions
- A registered international standard
- Can be used as a switching mechanism

Fundamental Principles of Dublin Core

- Flexible
- Extensible
- All elements are optional
- All elements are repeatable
- 1:1 principle
- Dumb down principle

Dublin Core (Simple) Elements

- Title
- Subject
- Description
- Type
- Source
- Relation
- Coverage
- Creator
- Publisher
- Contributor
- Rights
- Date
- Format
- Identifier
- Language

VRA Core

- Developed and maintained by the Visual Resources Association
- Highly specialized – designed for a specific community of practice
- Has a related content standard (VRA-CCO)
- Hierarchical
- Follows the 1:1 rule

VRA Core Elements (3.0)

- Record Type
- Type
- Measurements
- Material
- Technique
- Creator
- Date
- Location
- ID Number
- Style/Period
- Culture
- Subject
- Relation
- Description
- Source
- Rights

VRA Core Elements (3.0)

- Record Type
- Type
- Measurements
- Material
- Technique
- Creator
- Date
- Location
- ID Number
- Style/Period
- Culture
- Subject
- Relation
- Description
- Source
- Rights

More specialized schemas

- EAD – Encoded Archival Description
- TEI – Text Encoding Initiative

Content Standards

- General Purpose
  - AACR2, RDA
  - DCMI Dublin Core Usage Guide
- Specialized
  - Cataloging Cultural Objects (CCO)
  - Describing Archives (DACS)

Controlled Values

- General Purpose
  - LCSH, FAST, TGM
  - DCMI Type Vocabulary

- Special
  - AAT
  - TGM (both general and special)
  - ULAN
  - TGN
  - Code lists and local vocabularies

General purpose controlled vocabularies/values

- LCSH – Library of Congress Subject Headings
- FAST – Faceted Application of Subject Terminology
- TGM – Thesaurus for Graphic Materials
- DCMI Type Vocabulary

Special controlled vocabularies/values

- AAT – Art and Architecture Thesaurus
- ULAN – Union List of Artists Names
- TGN – Thesaurus of Geographic Names
- TGM – Thesaurus for Graphic Materials
- Code lists
- Locally developed controlled vocabularies

Standards for Encoding, Transmission and Harvesting

- XML
- OAI-PMH
- METS

Best Practices

- Application Profiles
- Standards Documentation
- Consortia Documents
- Professional Associations
- National and International Library Agencies

Measures of Quality

- Completeness
- Provenance
- Accuracy
- Conformance to Expectations
- Logical Consistency and Coherence
- Timeliness
- Accessibility

Thomas R. Bruce and Diane I. Hillman, 2004. "The continuum of metadata quality: defining, expressing, exploiting," In: Diane I. Hillman and Elaine Westbrooks (editors). Metadata in Practice. Chicago: ALA Editions, pp. 238-256.

Shareable Metadata

- Human Understandable
- Machine Processable
- Useful and Useable
- Supports Search Interoperability
- Meets Quality Criteria (Hillman)

Moving towards shareable metadata by Sarah L. Shreeves, Jenn Riley, and Liz Milewicz in First Monday, volume 11, number 8 (August_2006) , URL:http://firstmonday.org/issues/issue11_8/shreeves/index.html

Characteristics of Shareable Metadata

- Content
- Consistency
- Coherence
- Context
- Communication
- Conformance to Standards

Moving towards shareable metadata by Sarah L. Shreeves, Jenn Riley, and Liz Milewicz in First Monday, volume 11, number 8 (August_2006) , URL:http://firstmonday.org/issues/issue11_8/shreeves/index.html

Exercise 1

- Matching

Matching

Standards & Tools	Traditional Cataloging	Metadata Cataloging
Schemas		
Content Standards		
Value Spaces / Controlled Vocabularies		
Encoding Standards		
Transmission Standards		
Data Model / Framework		
Best Practices & Guidelines		

AACR2
AAT
CCO
CDWA
DACS
DC
DDC
EAD
FRBR
ISBD

LCSH
MARC
METS
MODS
OAI-PMH
RDF
TGM
TGN
VRA Core
XML
Z39.50
Others?

AACR2	Anglo-American Cataloguing Rules (AACR2)
AAT	Art & Architecture Thesaurus
CCO	Cataloging Cultural Objects
CDWA	Categories for the Description of Works of Art
DACS	Describing Archives: a content standard
DC	Dublin Core
DDC	Dewey Decimal System
EAD	Encoded Archival Description
FRBR	Functional Requirements for Bibliographic Records
ISBD	International Standard Bibliographic Description
LCSH	Library of Congress Subject Headings
MARC	MAchine-Readable Cataloging
METS	Metadata Encoding and Transmission Standard
MODS	Metadata Object Description Schema
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
RDF	Resource Description Framework
TGM	Thesaurus for Graphic Materials
TGN	Thesaurus of Geographic Names
VRA Core	Visual Resources Association Core Categories
XML	Extensible Markup Language
Z39.50	NISO standard for an application layer protocol for information retrieval

Standards & Tools	Traditional Cataloging	Metadata Cataloging
Schemas	MARC	EAD, CDWA, MODS, VRA, DC
Content Standards	AACR2	CCO, DACS, RDA
Value Spaces / Controlled Vocabularies	LCSH, Sears, DDC	AAT, TGM, TGN
Encoding Standards	MARC	XML, METS, EAD (XML)
Transmission Standards	Z39.50	METS, OAI-PMH
Data Model / Framework	ISBD	FRBR, RDF
Best Practices & Guidelines	LCRI, Local Guidelines	Application Profiles, Best Practice Documents produced by various communities

Understanding Interoperability

- Interoperability is critical for . . .
- Barriers
- Approaches

Interoperability is critical for. . .

- Federated Searching
- Harvesting
- Inter and Intra Institutional Collaboration
- Future Proofing

Barriers to Interoperability:

- Semantic differences
- Different practices
- Differences in representation
- Different vocabularies
- Multiple versions

Priscilla Caplan, “Metadata Fundamentals for All Librarians,” (Chicago: ALA, 2003): 41-42.

Barriers . . . continued

- Different data models
- Differences in granularity

Some ways to facilitate interoperability:

- Compliance with standards and best practices
- Application Profiles
- Framework
- Conversion
- Integration
- Registries

Lois Mai Chan and Marcia Lei Zeng, "Metadata Interoperability and Standardization - A Study of Methodology Part I." *D-Lib Magazine*, 12, no. 6 (2006). http://www.dlib.org/dlib/june06/chan/06chan.html (accessed August 30, 2007) and "Metadata Interoperability and Standardization - A Study of Methodology Part II." *D-Lib Magazine,* 12, no. 6 (2006). http://www.dlib.org/dlib/june06/zeng/06zeng.html (accessed August 30, 2007)

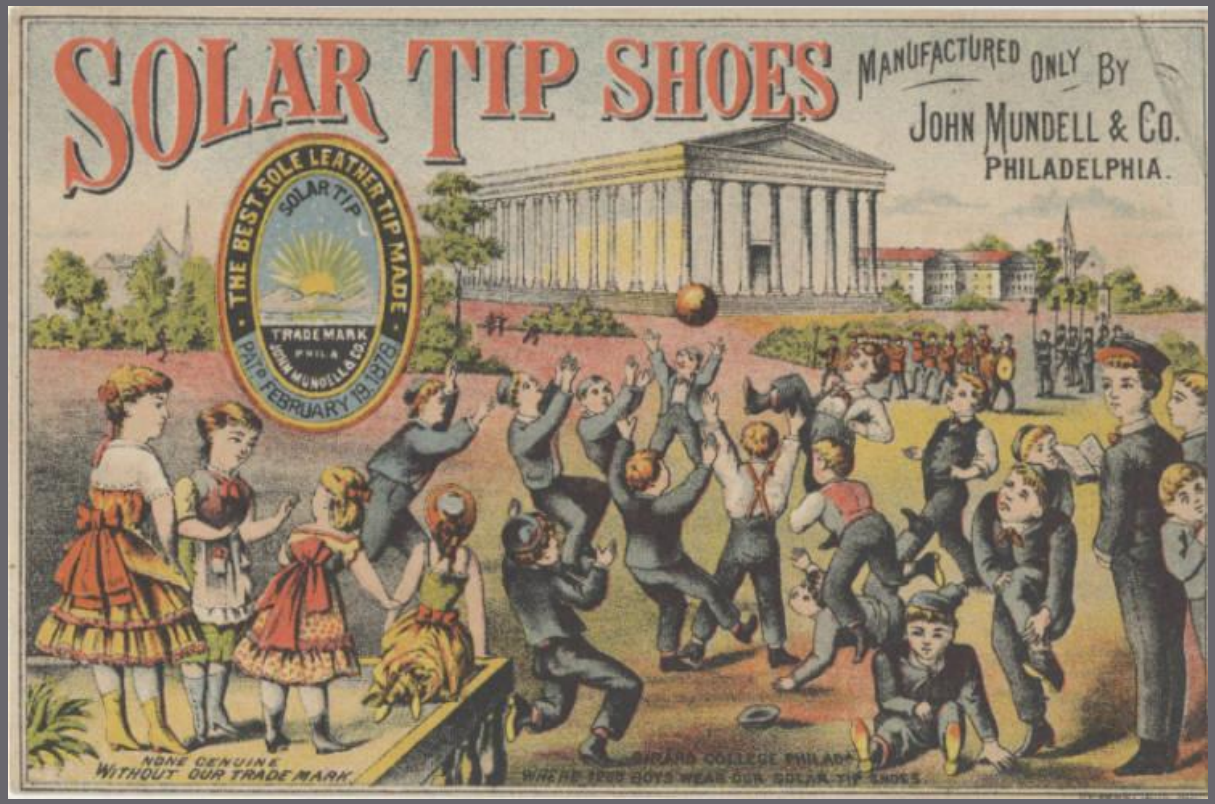
Exercise 2

- Mapping

Mapping

DC Element	MARC Tags
Title	
Creator	
Subject	
Description	
Contributor	
Publisher	
Date	
Type	
Format	
Identifier	
Source	
Language	
Relation	
Coverage	
Rights	

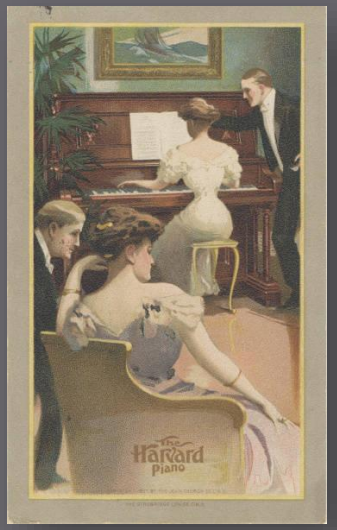
Breaktime



DC Element	MARC Fields
Title	245
Creator	100, 110, 111, 700, 710, 711
	720
Subject	600, 610, 611, 630, 650, 653
Description	500-599, except 506, 530, 540, 546
Contributor	Not used
Publisher	260ab
Date	260$c
Type	Leader06, Leader07
	655
Format	856$q
Identifier	856$u
Source	786ot
Language	008/35-37
	546
Relation	530, 760-787ot
Coverage	651
	752
Rights	506, 540

MARC to DC Crosswalk, Library of Congress: http://www.loc.gov/marc/marc2dc.html

PART II : Understanding Your Environment



PART II : Your Environment

- Communities of Practice
- Other communities
- The Library
- The Department
- Project Teams
- Collections

Community of Practice

- Library (MARC, MODS, AACR2)
- Archival (EAD, DACS)
- Museum (CDWA, VRA Core, CCO)
- Education (IEEELOM, CanCore)
- Others

Other Communities

- Consortia
- Repositories of any kind
- Dark Archives
- Harvesting Agencies

Inside the Library

- Work place culture, politics, organization
- Support
- Collaboration
- Standards & Systems

Inside the Department

- Current policies, procedures and legacy practices
- Adaptability
- Available expertise
- Workflow

Project Teams

Diverse perspectives of:

- Team members
- Stakeholders

Impact on:

- Approach to planning
- Decision making
- Implementation

MARC and Non-MARC Metadata Creation: Similarities

- Production oriented
- Tangible results
- Requires coordination of multiple processes
- Based on same theoretical principles
- Standards driven
- May have the same reporting lines

MARC and Non-MARC Metadata Creation: Differences

	MARC	Non-MARC
Workflow	Routine	Project based
Funding	Ongoing	Project based
Stakeholders	Fixed	Project based
Formats	Homogenous	Heterogeneous
Standards	Stable	Evolving
	Plug & Play	Mix & Match
Local Documentation	Fewer choices	More choices
	Less important	More important
Environment	Stable	Shifting landscape

Related Collections

- Legacy Collections
- Orphaned Project Collections
- Consortia Collections

Breaktime

PART III :
Making Choices

Things to consider when . . .

- Selecting a Schema
- Selecting Elements
- Selecting Vocabulary/Value Schemes
- Selecting a Content Standard

Beall's Points of Comparison for Schemas

- Granularity
- Ability to handle Multiple Formats
- Connection to Content Standards
- Software that supports it
- Community Specificity
- Interoperability
- Proven Success
- Training Required
- Viability of Sponsoring Organization
- Ability to handle Specific Functions
- Adaptability to Local Needs
- Scalability
- Surrogacy

Selecting a Schema

- Authoritative Source
- Related Content Standard
- XML DTD or XML Schema
- OAI-PMH compliant
- Endorsed by the relevant COP
- Suitability for the local environment
- Supports needs of the collection / project
- Supports desired system functionality
- Appropriate for the designated user community
- Appropriate to the materials being described

Selecting Elements

- Represent universal attributes across all or almost all items in the collection
- Provide any required context needed to “interpret” the resource
- Represent key access points . . .
- Represent all necessary information for location, management and use of the resource
- Support the needs of the user

Selecting Controlled Values/Vocabularies

- Developed and maintained by an authoritative body
- Regularly up-dated
- Follows recommended best practice for your community
- Appropriate for the designated element
- Appropriate to the project/collection
- Supports interoperability

Footnote on topical access issues

- LCSH
- Specialized Controlled Vocabularies
- Local Controlled Vocabularies
- Keywords
- Folksonomies

Selecting a Content Standard

- Developed and maintained by a recognized authority
- Regularly up-dated
- Recommended by relevant Community of Practice
- Appropriate for the Schema(s) you've chosen

Selected discrepancies between major content standards:

- Methods of deriving titles
- Treatment of derived titles
- Unknown creators
- Uncertain dates
- Use of abbreviations
- Use of qualifiers

Footnote on adding local elements

- Hard to avoid
- Look for guidance in related Best Practices
- Not necessary to make all elements available for harvesting or sharing
- Consider interoperability across local collections

Exercise

- Decisions
- Documentation

Metadata Documentation

Project Name:									
Content Standard:									
Schema(s)									
Field Label	**Mapping**	**Data Type**	**Search-able**	**Hidden from User Display**	**Controlled Vocabulary**	**Repeatable**	**Obligation**	**Responsibility or Data Source**	**Input Guidelines**

Project Name: Camden Local History Photograph Collection									
Content Standard: AACR2, DACS									
Schema(s): DC qualified									
Field Label	**DC Mapping**	**Data Type**	**Search-able**	**Hidden from User Display**	**Controlled Vocabulary**	**Repeatable**	**Obligation**	**Responsibility or Data Source**	**Input Guidelines**
Title	Title	Text	Yes	No	No	Yes	Mandatory	Derived	Enter [who/what, where, when].
Caption	Title.alternative	Text	Yes	No	No	Yes	Required if available	Hand lettered caption on original	Enter as it appears on item
Creator	Creator	Text	Yes	No					
Description									
Subject									
Rights									

Miami University Libraries Frank Snyder Collection Data Dictionary (draft)

Field Name	Dublin Core Mapping	data Type	Large field	Searchable	Hidden	CV	Obligation	Repeatable Values?	Suggested Definition and Usage Guidelines	Examples	Notes
Miami University Libraries Collection	-		-	-					[CD=Frank Snyder Collection]	Frank Snyder Collection	CONSTANT DATA
Collection Description	-		-	-					[CD=narrative description - from Bob]	http://content.lib.muohio.edu/site-templates/about.html	CONSTANT DATA
Title	Title	Text	No	Yes	No	No	required	no	A brief but unique descriptive name for the photo [who, what, where, when]. Enter personal names in direct order. Capitalize only proper nouns. If dates are ambiguos use circa [e.g. "Alumni Hall ca. 1900"].	Alumni on college commons, 1910.	Mapped from Archives "Subject". Requires some clean-up.
Photographer	Creator	Text	No	Yes	No	No	required	no	[CD] Name of photographer.	Snyder, Frank	CONSTANT DATA
Abstract	Description-Abstract	Text	Yes	Yes	No	No	optional	no	Narrative details of the what appears in the photo. Enter any descriptive information not included elsewhere.	Western College in background	Mapped from Archives "Abstract".
Subject	Subject	Text	No	Yes	No	TGM	optional	yes; enter a semi-colon after each value	Terms chosen from the internal thesaurus (TGM) that relate to the content represented in the image.	Alumni and alumnae; Women's education; portrait photographs;	No mapping. Archives will add these later as time permits.
Physical Description of Original	Format	Text	No	Yes	No	No	required where available		Dimensions of the original photograph are in inches [entered as w x h in. ; materials]. Separate dimensions from materials with a space semicolon space. Where possible use standard terms for materials.	8 x 10 in. ; glass negative	Mapped from Archives "Size". Requires some clean-up.
Original Creation Date	Date-Created	Text	No	Yes	No	No	required where available	no	Approximate date original photo was taken.	1909 (June)	Mapped from Archives "Date".
Resource Type	Type	Text	No	Yes	No	No	-	-	[CD] Select from DCMI list of terms for DC.type	Image	CONSTANT DATA / Will need to do a global change to comply with DC.type term (to change from Photo to DC term - Image).
Snyder Identification No.	Source	Text	No	Yes	No	No	-	-	Identifying number on photo and digital image.	9413	For this collection DC.identifier and DC.source are the same.
Repository	None	Text	No	No	No	No	-	-	[CD and/or automatically generated during load]	Miami University Archives	CONSTANT DATA
Digital Publisher	Publisher	Text	No	No	No	No	-	-	[CD and/or automatically generated during load]	Miami University Libraries-Digital Library Program	CONSTANT DATA
Digital Creation Date	Date-Available	Date	No	No	Yes	No	-	-	[CD and/or automatically generated during load]	81008	CONSTANT DATA
File Format	Format	Text	No	No	Yes	No	-	-	[CD and/or automatically generated during load]	jpeg	CONSTANT DATA
Copyright	Rights	Text	No	No	No	No	required	no	[CD] Brief copyright statement.	Copyright Miami University Libraries	CONSTANT DATA
Archives Internal Note	None	Text	No	No	Yes	No	optional	-	Use for internal notes not intended for public view.	Journal not clear - shot sometime in 1901.. entered in journal in June 1905	Open text field. For internal use only
Digital Services Internal Note	None	Text	No	No	Yes	No	optional	-	Document processing details here.	Rotate; Remove	Open text field. For internal use only
LC Collection Level Subject	Subject	Text	No	Yes	No	LCSH	required		Enter one to three subject heading that represent the content for the entire collection.	Universities and colleges--Alumni and alumnae; Miami University--Alumni and alumnae; Miami University (Oxford, Ohio)--History;	CONSTANT DATA

Draft 8/28/2006

Data Dictionaries & Application Profiles

- University of Washington DD Index
- Snyder Collection
- OhioLINK DMC Application Profile
- Colorado Digitization Project

Questions?


H.J. HEINZ COMPANY PICKLES FOOD PRODUCTS.
STAIRWAY
AUDITORIUM
57 VARIETIES
Auditorium Main Plant, H.J. HEINZ CO. PITTSBURGH, U.S.A.