



Alphabet Soup: A Metadata Overview

Melanie Schlosser
Metadata Librarian

Contents

- What is metadata?
- Types of metadata (function)
- Types of metadata (format)
- Where does metadata come from?
- What kind should I use?
- Who's that guy on the horse? (sharing metadata)



Definitions of metadata

- “...structured data about digital (and non-digital) resources that can be used to help support a wide range of operations. These might include, for example, resource description and discovery, the management of information resources (including rights management) and their long-term preservation.” (UKOLN)
- “...the sum total of what one can say about any *information object* at any level of aggregation.” (Gilliland)
- Metadata is data about data

Who uses metadata?

- Scientists
 - Manage datasets and digital objects
- Business
 - Organization and protection of proprietary data
 - Communication and industry standardization
- Museums, other cultural heritage institutions
- Libraries

Why do libraries use it?

- Traditional cataloging: Organize and access local collections

Title Existentialism : basic writings / edited, with introductions, by Charles Guignon and Derk Pereboom
Publish Info Indianapolis : Hackett, c1995
[Permanent link to this record](#)

Copy Status	Details	Find Similar Items	Full Record
Description	xxxviii, 340 p. ; 23 cm		
Notes	Includes bibliographical references		
ISBN	0872202151 (alk. paper) 0872202143 (pbk. : alk. paper)		
OCLC #	31815210		
LCCN	94048670		
Subjects	Existentialism		
Add'l Authors	Guignon, Charles B., 1944- Pereboom, Derk, 1957-		

Why do libraries use it?



Please use this identifier to cite or link to this item: <http://hdl.handle.net/1811/5984>

Title: Dr. Louis Hauti Potaka of New Zealand: a biographical essay

Creators: Young, Bruce

Keywords: Potaka, Louis Hauti

Issue Date: Dec-2005

URI: <http://hdl.handle.net/1811/5984>

Appears in Collections: [E-Publications](#)

Files in This Item:

File	Description	Size	Format
potaka.pdf		329Kb	Adobe PDF View/Open

[Show full item record](#)

Why do libraries use it?

- Share information about collections

- MARC  The Library of Congress
Network Development
& MARC Standards Office

- OCLC  WorldCat® Beta

- OAIster...

Why do libraries use it?

The screenshot shows the OAlster interface with a navigation bar (Home, Search, Help, About, Using OAlster, News) and a search result for a record from the Knowledge Bank at Ohio State University (OSU). The record details are as follows:

Record 1 of 1	
add to bookbag	
Title	October 17, 2006
Publisher	Ohio State University Libraries
Year	2007-06-11T21:17:29Z
Year	2006-10-17
Resource Type	Newsletter
Resource Format	46780 bytes, application/pdf
Language	English
Subject	Ohio State University. Libraries.
URL	http://hdl.handle.net/1811/28497
Data Contributor	Knowledge Bank at Ohio State University (OSU)

Why do libraries use it?

- Manage complex digital objects
 - e.g. Digitized photograph w/ TIFF master image, screen-sized JPEG, and thumbnail-sized JPEG



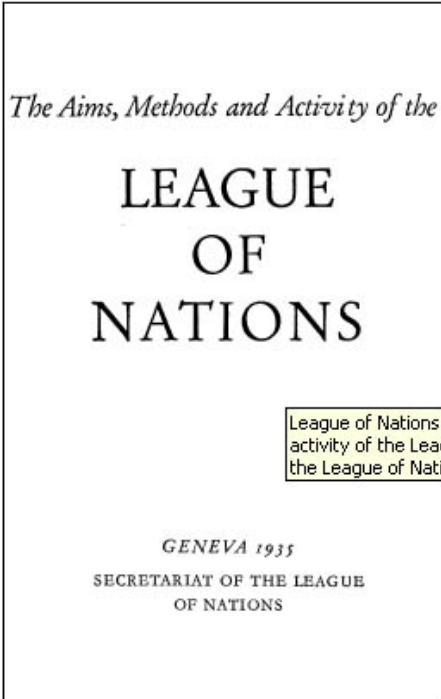
- e.g. Digitized book...

Metadata at work...

League of Nations. Secretariat. The aims, methods, and activity of the League of Nations.
Geneva : Secretariat of the League of Nations, 1935.

View full record: [IUCAT](#)

|< << 1 of 224 >> >| [\[view larger image\]](#)



The Aims, Methods and Activity of the
**LEAGUE
OF
NATIONS**

GENEVA 1935
SECRETARIAT OF THE LEAGUE
OF NATIONS

League of Nations. Secretariat. The aims, methods, and activity of the League of Nations. Geneva : Secretariat of the League of Nations, 1935.: Page 1 of 224

|< << 1 of 224 >> >| [\[view larger image\]](#)

- Front Matter
 - **Title page**
 - Verso
 - Table of Contents
- Body
 - Part 1
 - Chapter 1
 - Chapter 2
 - Chapter 3
 - Chapter 4
 - Part 2
 - Chapter 1
 - Chapter 2
 - Chapter 3
 - Chapter 4
 - Part 3
 - Chapter 1
 - Chapter 2
 - Conclusion
 - Part 4
 - Chapter 1
 - Chapter 2
 - Chapter 3
 - Chapter 4
- Back Matter
 - Publications
 - Preservation colophon

What kinds of metadata are there?

Types of metadata (function)

- Metadata often classified by the function of the information it captures
- There are 5 basic types
 - Descriptive
 - Technical
 - Preservation
 - Rights
 - Structural
- Markup

Descriptive: Dublin Core



KnowledgeBank

UNIVERSITY LIBRARIES AND OFFICE OF THE CHIEF INFORMATION OFFICER



Full metadata record

DC Field	Value	Language
creator	Young, Bruce	-
date.accessioned	2006-03-02T17:48:03Z	-
date.available	2006-03-02T17:48:03Z	-
date.issued	2005-12	-
identifier.uri	http://hdl.handle.net/1811/5984	-
description	On essay: December 31, 1999. Revised December, 2005.	en
format.extent	337401 bytes	-
format.mimetype	application/pdf	-
language.iso	en	en
subject	Potaka, Louis Hauti	-
title	Dr. Louis Hauti Potaka of New Zealand: a biographical essay	en
type	Article	en

Descriptive:

MODS



```
<mods version="3.0">
  <titleInfo>
    <title>Hiring and recruitment practices in academic libraries</title>
  </titleInfo>
  <name type="personal">
    <namePart>Raschke, Gregory K.</namePart>
    <displayForm>Gregory K. Raschke</displayForm>
  </name>
  <typeOfResource>text</typeOfResource>
  <genre>journal article</genre>
  <originInfo>
    <place>
      <placeTerm type="text">Baltimore, Md.</placeTerm>
    </place>
    <publisher>Johns Hopkins University Press</publisher>
    <dateIssued>2003</dateIssued>
  </originInfo>
```

Technical: TIFF Header

The Directory Entries, 12 bytes each entry				
Tag Name	Tag Code (2 bytes)	Type (2 bytes)	Length (4 bytes)	Value; Pointer (4 bytes)
newsubfiletype	254	LONG	1	0
imagewidth	256	SHORT	1	IMAGEWIDTH
imagelength	257	SHORT	1	IMAGEHEIGHT
bitspersample	258	SHORT	1	8; 8-bit, single plane data
compression	259	SHORT	1	1; no compression, pack bytes
photointerp	262	SHORT	1	3; image is color, single plane
make	271	ASCII	7	COMPIX
model	272	ASCII	5	6000
stripoffsets	273	LONG	1	offset; to actual image data
samplesperpixel	277	SHORT	1	1; 'Palette Color' or bilevel
rowsperstrip	278	LONG	1	imageheight; all rows in 1 strip
stripbytecount	279	LONG	1	imageheight*imagewidth; all in 1 strip
xresolution	282	RATNL	1	75 dots per inch
yresolution	283	RATNL	1	75 dots per inch
planarconfig	284	SHORT	1	1; single image plane
resunit	296	SHORT	1	2 res is in dots per INCH
software	305	ASCII	24	offset; "COM6 original file type"
colormap	320	SHORT	768	offset; 3 x 255 colormap table
compixprivate	32970	SHORT	23	offset; 23 words in tag
compixprivate	32971	SHORT	21	offset; 21 words in tag
compixprivate	32972	SHORT	varies	offset; "WinTES" files only

Preservation: **P R E M I S**

- “Preservation metadata is information that supports and documents the digital preservation process:
 - Provenance: *Who has had custody/ownership of the digital object?*
 - Authenticity: *Is the digital object what it purports to be?*
 - Preservation Activity: *What has been done to preserve the digital object?*
 - Technical Environment: *What is needed to render and use the digital object?*
 - Rights Management: *What intellectual property rights must be observed?”*

(PREMIS Information Sheet)

Rights:



- CDL's *Assessment of Common Rights Metadata Encoding Schemes and Gap Analysis* suggests that the following are needed:
 - Creator, personal
 - Creator, corporate
 - Creator (unspecified)
 - Creator death date
 - Copyright holder
 - Date of publication
 - Date of copyright
 - Date of creation
 - Date of renewal
 - Publisher
 - Country of publication
 - Published/unpublished
 - Type of resource
 - Contact information
 - Other terms and conditions
- Rights metadata found in many other schemas

Structural: *METS*

- ‘Wrapper’ schema for descriptive, administrative and structural metadata
- Used to manage complex digital objects
 - How many files are there?
 - What are they called?
 - Where are they stored?
 - How do they fit together?

Markup:



- A way to make text machine-readable
 - Publishing (print and online)
 - Humanities research
- HTML
- TEI (Text Encoding Initiative)
- CBML (Comic Book Markup Language)
 - Based on TEI
 - Allows representation of unique structures and conventions in comic books...

Markup:



```
<panel n="3">
  <panelDescription characters="cap" />
  <caption>
    <paragraph>
      Before the hammer descends on cap, his shield
      <emphasis style="bold">
        demolishes
      </emphasis>
      the evil mechanism!
    </paragraph>
  </caption>
  <soundEffect>
    KRAK!
  </soundEffect>
</panel>
```

Types of metadata (content)

- Many standards have been created to describe particular types of content
 - Images (MIX)
 - Learning objects (LOM, SCORM)
 - Geospatial data (FGDC)
 - Archival finding aids (EAD)
 - ...and many more!

MIX
←

<ead>


Where does metadata come from?

Content Management Systems (CMS)

- Descriptive metadata:

Knowledge

Bank


UNIVERSITY LIBRARIES AND OFFICE OF THE CHIEF INFORMATION OFFICER

Describe
Describe
Describe
Upload
Verify
License
License

Submit: Describe Your Item

Please fill in the requested information about your submission below. In most browsers, you can use the tab key to move the cursor having to use the mouse each time. ([More Help...](#))

Enter the name(s) of the author(s) of this item below.

<i>Last name</i>	<i>First name(s) + "Jr"</i>	
e.g. Smith	e.g. Donald Jr	
Creator		Add More

Enter the name(s) of any other entity with secondary responsibility for the creation of the resource (e.g., editors, photo)

<i>Last name</i>	<i>First name(s) + "Jr"</i>	
e.g. Smith	e.g. Donald Jr	
Contributor		Add More

Content Management Systems (CMS)

- Technical metadata:



Size	File Format	Checksum
2,166,272 bytes	Microsoft Powerpoint (known)	5555d635213a608d325beb85d4878ae5 (MD5)

[Click here if this is the wrong format](#)

[Click here if this is the wrong file](#)

Content Management Systems (CMS)

- Pros
 - Easy to use
 - Ability to customize
 - Ability to apply controlled vocabularies
 - Metadata is integrated with other functions
- Cons
 - Requires software
 - Not infinitely flexible

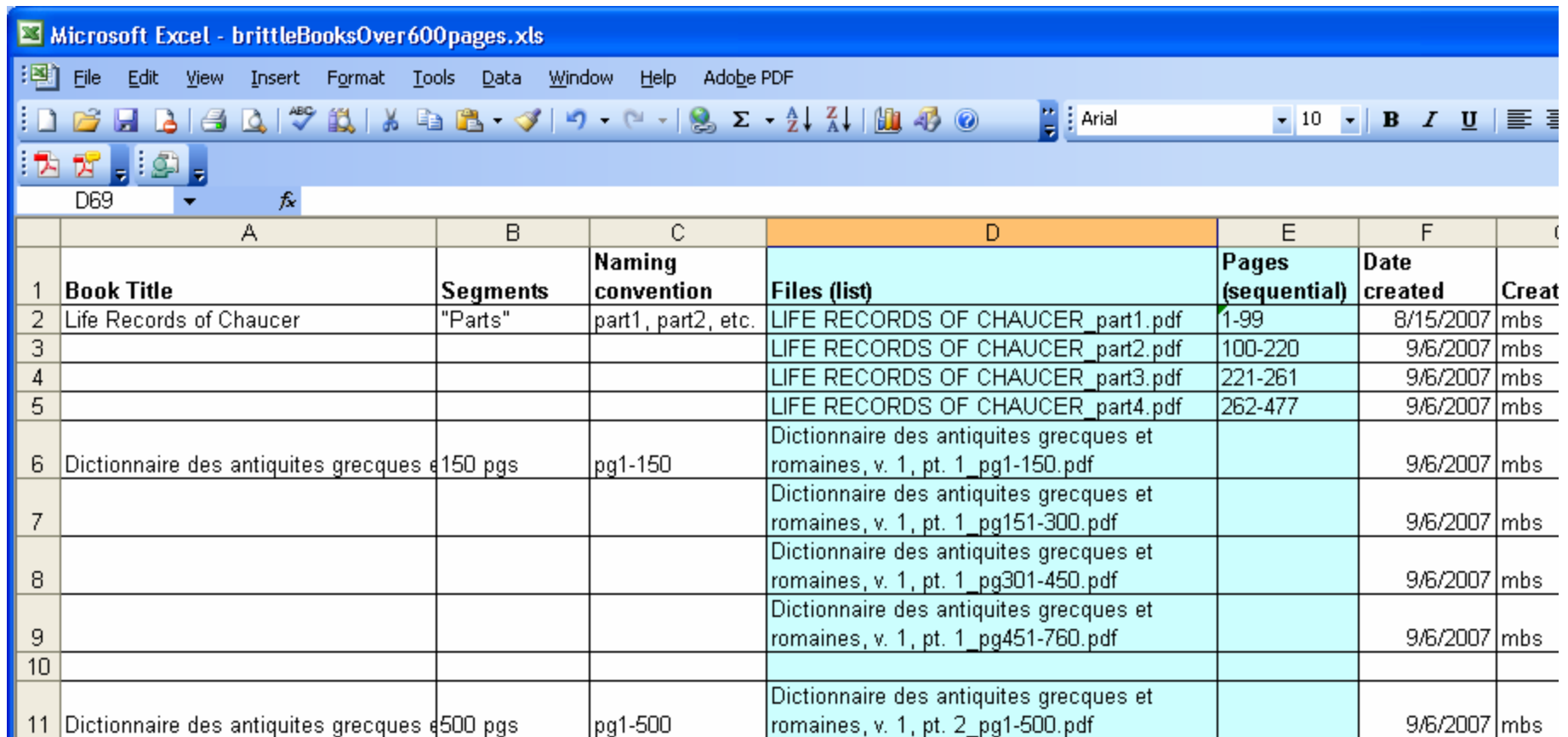
Image capture devices

- Scanners, cameras, etc.
- Automatic capture of technical metadata
 - Date of capture
 - Settings
 - Color information
- Usually a proprietary format

Image capture devices

- Pros
 - Automatic capture
 - Embedded and reusable
 - Aids in preservation
- Cons
 - Usually proprietary format – not a standard
 - Can be difficult to extract
 - Limited to technical metadata

Spreadsheets and databases

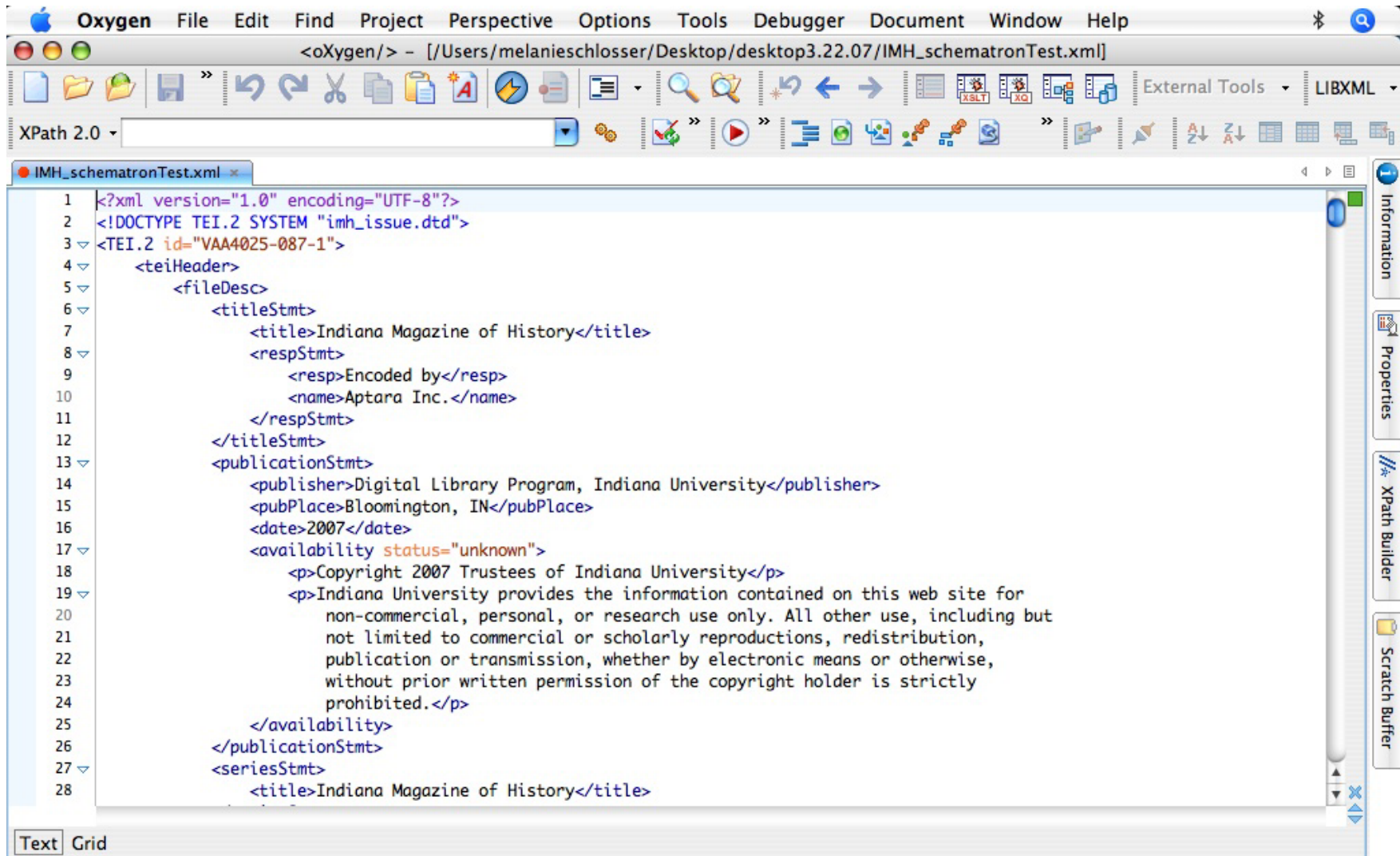


	A	B	C	D	E	F	G
1	Book Title	Segments	Naming convention	Files (list)	Pages (sequential)	Date created	Creat
2	Life Records of Chaucer	"Parts"	part1, part2, etc.	LIFE RECORDS OF CHAUCER_part1.pdf	1-99	8/15/2007	mbs
3				LIFE RECORDS OF CHAUCER_part2.pdf	100-220	9/6/2007	mbs
4				LIFE RECORDS OF CHAUCER_part3.pdf	221-261	9/6/2007	mbs
5				LIFE RECORDS OF CHAUCER_part4.pdf	262-477	9/6/2007	mbs
6	Dictionnaire des antiquites grecques et romaines, v. 1, pt. 1_150 pgs	pg1-150	pg1-150	Dictionnaire des antiquites grecques et romaines, v. 1, pt. 1_pg1-150.pdf		9/6/2007	mbs
7				Dictionnaire des antiquites grecques et romaines, v. 1, pt. 1_pg151-300.pdf		9/6/2007	mbs
8				Dictionnaire des antiquites grecques et romaines, v. 1, pt. 1_pg301-450.pdf		9/6/2007	mbs
9				Dictionnaire des antiquites grecques et romaines, v. 1, pt. 1_pg451-760.pdf		9/6/2007	mbs
10							
11	Dictionnaire des antiquites grecques et romaines, v. 1, pt. 2_500 pgs	pg1-500	pg1-500	Dictionnaire des antiquites grecques et romaines, v. 1, pt. 2_pg1-500.pdf		9/6/2007	mbs

Spreadsheets and databases

- Pros
 - Easy to use
 - Flexible
- Cons
 - Limited functionality
 - More sophisticated functionality requires expertise to create

XML



```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE TEI.2 SYSTEM "imh_issue.dtd">
3 <TEI.2 id="VAA4025-087-1">
4   <teiHeader>
5     <fileDesc>
6       <titleStmt>
7         <title>Indiana Magazine of History</title>
8       <respStmt>
9         <resp>Encoded by</resp>
10        <name>Aptara Inc.</name>
11      </respStmt>
12    </titleStmt>
13    <publicationStmt>
14      <publisher>Digital Library Program, Indiana University</publisher>
15      <pubPlace>Bloomington, IN</pubPlace>
16      <date>2007</date>
17      <availability status="unknown">
18        <p>Copyright 2007 Trustees of Indiana University</p>
19        <p>Indiana University provides the information contained on this web site for
20        non-commercial, personal, or research use only. All other use, including but
21        not limited to commercial or scholarly reproductions, redistribution,
22        publication or transmission, whether by electronic means or otherwise,
23        without prior written permission of the copyright holder is strictly
24        prohibited.</p>
25      </availability>
26    </publicationStmt>
27    <seriesStmt>
28      <title>Indiana Magazine of History</title>

```

XML

- Pros
 - Very powerful
 - Very flexible
 - Does not require proprietary software
- Cons
 - Requires expertise
- Alternative: have something else create it for you
 - CMS, database, etc.
 - Then have to transform it somehow...

Existing metadata

Metadata Object Description Schema (MODS) [Official Web Site](#)



MARC Mapping to MODS Version 3.2

3. Mapping

titleInfo	<titleInfo>
245 \$a\$\$g\$k	<title> <i>with</i> no <titleInfo> type attribute <i>and</i>
245 \$b	<subTitle>
245 \$n (<i>and</i> \$\$g\$k following \$n)	<partNumber>
245 \$p (<i>and</i> \$\$g\$k following \$p)	<partName>
245 ind2 is not 0	<nonSort> around characters excluded from sort as indicated in indicator value
<i>[If \$\$g\$k follow \$b they go with <subTitle>. If they follow \$a they go with <title>.]</i>	
210 \$a	<title> <i>with</i> <titleInfo> type="abbreviated" <i>and</i>
210 \$b	<subTitle>

Existing metadata

- Pros
 - Efficient
 - Utilizes existing assets
 - Tools exist to help convert
- Cons
 - Most projects won't fit an existing mapping exactly
 - Requires expertise to customize or create from scratch

What kind of metadata should I use?

Standards: Your best friend

- Allow you to share data
 - Benefits users
 - One-stop shopping
 - Aggregation of subject-specific resources
 - Benefits institutions
 - Increased exposure
 - Broader user base
 - Bringing together distributed collections

NINES a networked infrastructure for nineteenth-century electronic scholarship

[About 9s](#)

[People & Projects](#)

[Tools & Interfaces](#)

[Join 9s](#)



collect

[OCCASIONED BY A BASKET OF WATER LILIES BEING PRESENTED TO A YOUNG LADY FROM THE RIVER *****](#) →

site: [British Women Romantic Poets](#) →

name: Spencer, Mrs. Walter. (Author)

date: 1812

genre: Poetry [1528 items](#)



collect

[For Children: The Gates of Paradise, copy D \(object 4 \)](#) →

Water *Published by WBlake 17 May 1793*

site: [The William Blake Archive](#) →

name: William Blake (Author)

Catherine Blake (Author)

William Blake (Publisher)

Catherine Blake (Publisher)

date: 1793

genre: Visual Art [62 items](#)

Poetry [1528 items](#)



[The Water-Cress Girl](#) →

source: The Amulet: a Christian and Literary Remembrancer

site: [The Poetess Archive](#) →

name: Hall, S. C. (Samuel Carter) (Editor)

Moir, D. M. (David Macbeth) (Author)

Standards: Your best friend

- Well designed, documented, and supported
 - Designing an effective scheme is difficult
 - Your collections aren't dependent on any staff member
 - Better chance of preserving for the future
 - Official documentation and user communities for troubleshooting

Which standard should I use?

- Institution
 - Infrastructure and capabilities
 - Community
- Standard
 - What does it do well?
 - Who uses it?
 - What is required to create it?

Which standard should I use?

- Materials
 - Genre, format, etc.
 - Existing metadata
- Project
 - Level of description needed
 - Need relationships between records and versions?
 - Other uses for metadata? (Sharing, reuse, aggregation)

Which standard should I use?

Most importantly...

...what will the users of the collection need?

Sharing metadata

- You've scanned
- You've created metadata
- You used standards...

- You're ready to share, right?

Who's that guy on the horse?



(Smithsonian National Portrait Gallery)

Sharing metadata

- Can still have the ‘on a horse’ problem (Wendler)
 - Individual records rely on local context
 - Example: ‘Images of Theodore Roosevelt’ collection
 - Title: ‘On a horse’
 - When removed from original context (the collection), metadata is incomplete

Sharing metadata

The screenshot shows the OAlster digital library interface. At the top left is the OAlster logo with the tagline "...find the pearls". A navigation bar contains links for Home, Search, Help, About, Using OAlster, and News. On the left side, there is a box for "Knowledge Bank at Ohio State University (OSU)" with a link to "1 record". The main content area displays "Record 1 of 1" and a link to "add to bookbag". Below this is a metadata table with the following entries:

Title	October 17, 2006
Publisher	Ohio State University Libraries
Year	2007-06-11T21:17:29Z
Year	2006-10-17
Resource Type	Newsletter
Resource Format	46780 bytes, application/pdf
Language	English
Subject	Ohio State University. Libraries.
URL	http://hdl.handle.net/1811/28497
Data Contributor	Knowledge Bank at Ohio State University (OSU)

Sharing metadata

- How to avoid this problem
 - Collection-level description
- Other important stuff...
 - Internal consistency
 - Pay attention to community best practice
 - Use controlled vocabularies
 - Keep sharing in mind while creating metadata schemes
 - And finally...

Sharing metadata

We're only human...

Insert quality control into your metadata processes
whenever possible

Thank you!

References

- Gilliland, Anne J. *Setting the Stage: Introduction to Metadata*, Online edition, version 2.1, ed. by Murtha Baca, J. Paul Getty Trust.
http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html
- UKOLN Metadata <http://www.ukoln.ac.uk/metadata/>
- W3C Metadata <http://www.w3.org/Metadata/>
- Wendler, Robin. The Eye of the Beholder: Challenges of Image Description and Access at Harvard. In Hillman, Diane I. and Westbrook, Elaine L., eds., *Metadata in Practice*. American Library Association, Chicago, IL, 2004, 51-69.

Questions?

- schlosser.40@osu.edu
- Thank you to
 - Jenn Riley, Metadata Librarian, Indiana University
 - Maureen Walsh, Metadata Librarian, OSU

