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, c1996

Permanent link to this record

<table>
<thead>
<tr>
<th>Copy Status</th>
<th>Details</th>
<th>Find Similar Items</th>
<th>Full Record</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Description**: xxxviii, 340 p.; 23 cm
- **Notes**: Includes bibliographical references
- **ISBN**: 0872202151 (alk. paper)
  0872202143 (pok. : 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 Hauiti Potaka of New Zealand: a biographical essay
Creators: Young, Bruce
Keywords: Potaka, Louis Hauiti
Issue Date: Dec-2005
URI: http://hdl.handle.net/1811/5984
Appears in Collections: E-Publications

Files in This Item:

<table>
<thead>
<tr>
<th>File</th>
<th>Description</th>
<th>Size</th>
<th>Format</th>
<th>View/Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>potaka.pdf</td>
<td></td>
<td>329Kb</td>
<td>Adobe PDF</td>
<td>View/Open</td>
</tr>
</tbody>
</table>
Why do libraries use it?

• Share information about collections
  – MARC
  – OCLC
  – OAIster…
Why do libraries use it?

Record 1 of 1

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...


The Aims, Methods and Activity of the

LEAGUE

OF

NATIONS

GENEVA 1935
SECRETARIAT OF THE LEAGUE OF NATIONS

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
### Full metadata record

<table>
<thead>
<tr>
<th>DC Field</th>
<th>Value</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>creator</td>
<td>Young, Bruce</td>
<td>-</td>
</tr>
<tr>
<td>date.accessioned</td>
<td>2006-03-02T17:48:03Z</td>
<td>-</td>
</tr>
<tr>
<td>date.available</td>
<td>2006-03-02T17:48:03Z</td>
<td>-</td>
</tr>
<tr>
<td>date.issued</td>
<td>2005-12</td>
<td>-</td>
</tr>
<tr>
<td>identifier.uri</td>
<td><a href="http://hdl.handle.net/1811/5984">http://hdl.handle.net/1811/5984</a></td>
<td>-</td>
</tr>
<tr>
<td>format.extent</td>
<td>337401 bytes</td>
<td>-</td>
</tr>
<tr>
<td>format.mimetype</td>
<td>application/pdf</td>
<td>-</td>
</tr>
<tr>
<td>language.iso</td>
<td>en</td>
<td>en</td>
</tr>
<tr>
<td>subject</td>
<td>Potaka, Louis Hauiti</td>
<td>-</td>
</tr>
<tr>
<td>title</td>
<td>Dr. Louis Hauiti Potaka of New Zealand: a biographical essay</td>
<td>en</td>
</tr>
<tr>
<td>type</td>
<td>Article</td>
<td>en</td>
</tr>
</tbody>
</table>
Descriptive:

<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>
</mods>
## Technical: TIFF Header

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

- “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: \textit{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...
<panel n="3">
  <panelDescription characters="cap" />
  <caption>
    <paragraph>
      Before the hammer descends on cap, his shield demolishes the evil mechanism!
    </paragraph>
    <emphasis style="bold">
      demolishes
    </emphasis>
    <emphasis>
      the evil mechanism!
    </emphasis>
  </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!
Where does metadata come from?
Content Management Systems (CMS)

- Descriptive metadata:
Content Management Systems (CMS)

- Technical metadata:

<table>
<thead>
<tr>
<th>Size</th>
<th>File Format</th>
<th>Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,166,272 bytes</td>
<td>Microsoft Powerpoint (known)</td>
<td>5555d635213a608d325beb85d4878ae5 (MD5)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Book Title</td>
<td>Segments</td>
<td>Naming convention</td>
<td>Files (list)</td>
<td>Pages (sequential)</td>
</tr>
<tr>
<td>2</td>
<td>Life Records of Chaucer</td>
<td>“Parts”</td>
<td>part1, part2, etc.</td>
<td>LIFE RECORDS OF CHAUCER_part1.pdf</td>
<td>1-99</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>LIFE RECORDS OF CHAUCER_part2.pdf</td>
<td>100-220</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>LIFE RECORDS OF CHAUCER_part3.pdf</td>
<td>221-261</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>LIFE RECORDS OF CHAUCER_part4.pdf</td>
<td>262-477</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>Dictionnaire des antiquites grecques et romaines, v. 1, pt. 1 pg151-300.pdf</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>Dictionnaire des antiquites grecques et romaines, v. 1, pt. 1 pg301-450.pdf</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Spreadsheets and databases

• Pros
  – Easy to use
  – Flexible

• Cons
  – Limited functionality
  – More sophisticated functionality requires expertise to create
XML
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

MARC Mapping to MODS
Version 3.2

3. Mapping

titleInfo

245 $a$s$fb$g$k
245 $b
245 $n (and $f$g$k following $n)
245 $p (and $f$g$k following $p)
245 ind2 is not 0

<title> with no <titleInfo> type attribute and <subTitle>
<title> with <titleInfo> type="abbreviated" and <subTitle>

[If $f$g$k follow $b$ they go with <subTitle>. If they follow $a$ they go with <title>.]
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
Alphabet Soup: A Metadata Overview

NINES

A networked infrastructure for nineteenth-century electronic scholarship

About 9s | People & Projects | Tools & Interfaces | Join 9s

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

For Children: The Gates of Paradise, copy D (object 4)

Source: The William Blake Archive

Water Published by W. Blake 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: Holl, 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…

…and 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

Knowledge Bank at Ohio State University (OSU)
1 record

Record 1 of 1

<table>
<thead>
<tr>
<th>add to bookbag</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong>: October 17, 2006</td>
</tr>
<tr>
<td><strong>Publisher</strong>: Ohio State University Libraries</td>
</tr>
<tr>
<td><strong>Year</strong>: 2007-06-11T21:17:29Z</td>
</tr>
<tr>
<td><strong>Year</strong>: 2006-10-17</td>
</tr>
<tr>
<td><strong>Resource Type</strong>: Newsletter</td>
</tr>
<tr>
<td><strong>Resource Format</strong>: 46780 bytes, application/pdf</td>
</tr>
<tr>
<td><strong>Language</strong>: English</td>
</tr>
<tr>
<td><strong>Subject</strong>: Ohio State University. Libraries.</td>
</tr>
<tr>
<td><strong>URL</strong>: <a href="http://hdl.handle.net/1811/28497">http://hdl.handle.net/1811/28497</a></td>
</tr>
<tr>
<td><strong>Data Contributor</strong>: Knowledge Bank at Ohio State University (OSU)</td>
</tr>
</tbody>
</table>
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

  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/

Questions?

• schlosser.40@osu.edu

• Thank you to
  – Jenn Riley, Metadata Librarian, Indiana University
  – Maureen Walsh, Metadata Librarian, OSU