

Spring 2011

Metadata at Work

Kathryn Lybarger
University of Kentucky, kathryn.lybarger@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/libraries_present



Part of the [Library and Information Science Commons](#)

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Repository Citation

Lybarger, Kathryn, "Metadata at Work" (2011). *Library Presentations*. 16.
https://uknowledge.uky.edu/libraries_present/16

This Presentation is brought to you for free and open access by the University of Kentucky Libraries at UKnowledge. It has been accepted for inclusion in Library Presentations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Guest Lecture
LIS 656, Spring 2011
Kathryn Lybarger

METADATA AT WORK

“data about data”

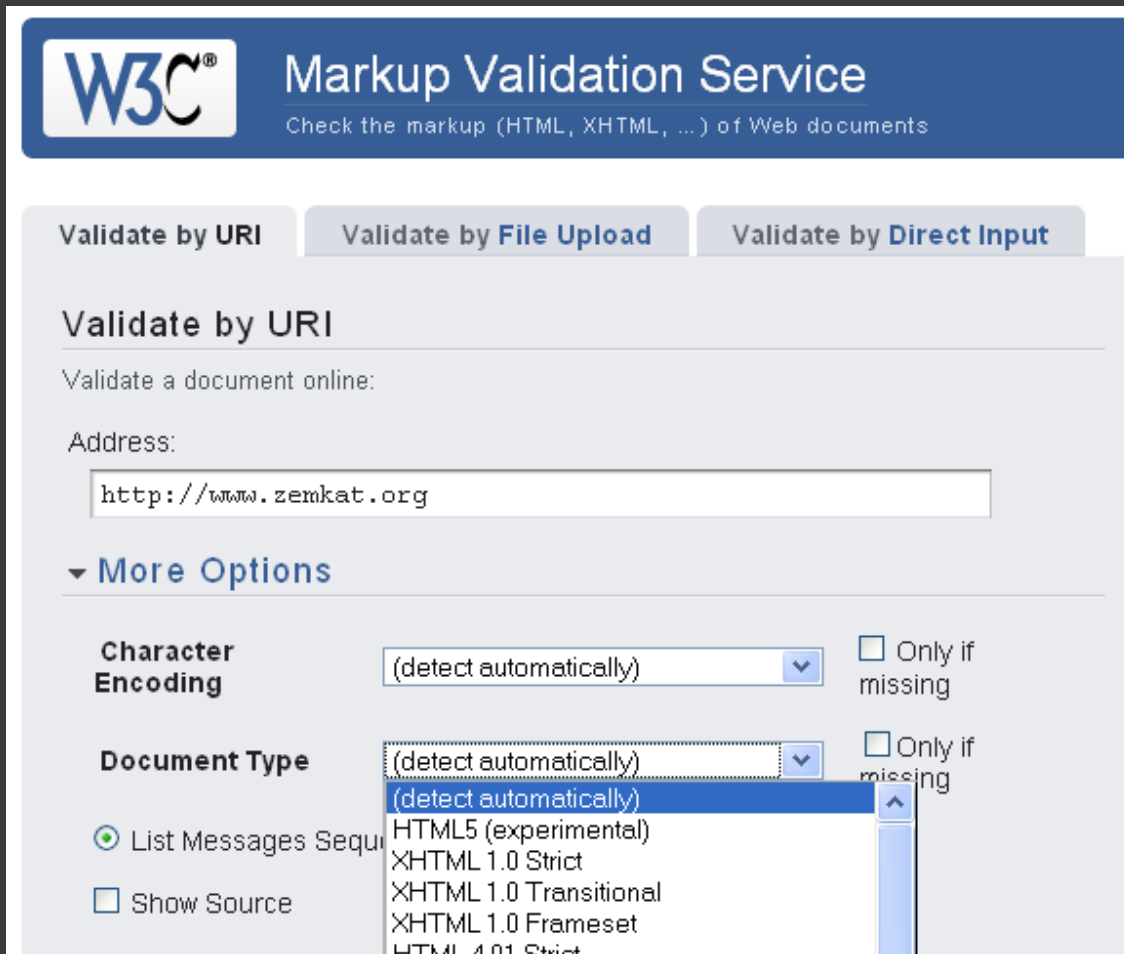
When choosing a standard...

- ⦿ Appropriate for format?
- ⦿ Sufficient granularity?
- ⦿ Where can I find examples?
- ⦿ Who maintains the standard?
- ⦿ Do I have a system which uses it?
- ⦿ Can I validate my documents?

Validators

- ⦿ May be web-based
- ⦿ May be a DTD/schema
- ⦿ May be (part of) a separate program
 - ILS validates MARC records

Validator example (XHTML)



W3C[®] Markup Validation Service
Check the markup (HTML, XHTML, ...) of Web documents

Validate by URI Validate by File Upload Validate by Direct Input

Validate by URI

Validate a document online:

Address:

▼ More Options


Character Encoding (detect automatically) Only if missing

Document Type (detect automatically) Only if missing

- List Messages Sequentially
- Show Source

HTML5 (experimental)
XHTML 1.0 Strict
XHTML 1.0 Transitional
XHTML 1.0 Frameset
HTML 4.01 Strict

An XHTML document with problems...



Markup Validation Service

Check the markup (HTML, XHTML, ...) of Web documents

Jump To: [Validation Output](#)

Error found while checking this document as XHTML 1.1!

Result:	1 Error	
Address:	<input type="text" value="http://www.zemkat.org/"/>	
Encoding:	utf-8	<input type="text" value="(detect automatically)"/>
Doctype:	XHTML 1.1	<input type="text" value="(detect automatically)"/>
Root Element:	html	
Root Namespace:	http://www.w3.org/1999/xhtml	

Specific problems

Validation Output: 1 Error

 *Line 11, Column 40:* **required attribute "alt" not specified**




```
<div></div>
```

The attribute given above is required for an element that you've used, but you have omitted it. For instance, in most HTML and XHTML document types the "type" attribute is required on the "script" element and the "alt" attribute is required for the "img" element.

Typical values for type are type="text/css" for <style> and type="text/javascript" for <script>.

A valid XHTML document

**Markup Validation Service**
Check the markup (HTML, XHTML, ...) of Web documents

Jump To: [Congratulations](#) · [Icons](#)

This document was successfully checked as XHTML 1.1!

Result:	Passed
Address :	<input type="text" value="http://www.zenkat.org/"/>
Encoding :	utf-8 <input type="text" value="(detect automatically)"/>
Doctype :	XHTML 1.1 <input type="text" value="(detect automatically)"/>
Root Element:	html
Root Namespace:	http://www.w3.org/1999/xhtml

DTD / Schema

- ⦿ Can validate in an XML editor
 - <oxygen/>
 - XMLSpy

- ⦿ Free tools:
 - Xmlint
 - Xmlstar
 - Programming language libraries

Encoded Archival Description (EAD)

- Electronic finding aids for archival description
- Maintained by Society of American Archivists (SOA)
- Allows for more structured description than MARC

Using EAD

- XML language (formerly SGML)
- Validates against a DTD
- Many people use DACS content

Describing Archives: A Content Standard (DACS)

- ◉ Expands on archival description rules from AACR2
- ◉ US implementation of ISAD(G)
- ◉ Can be used with MARC or EAD

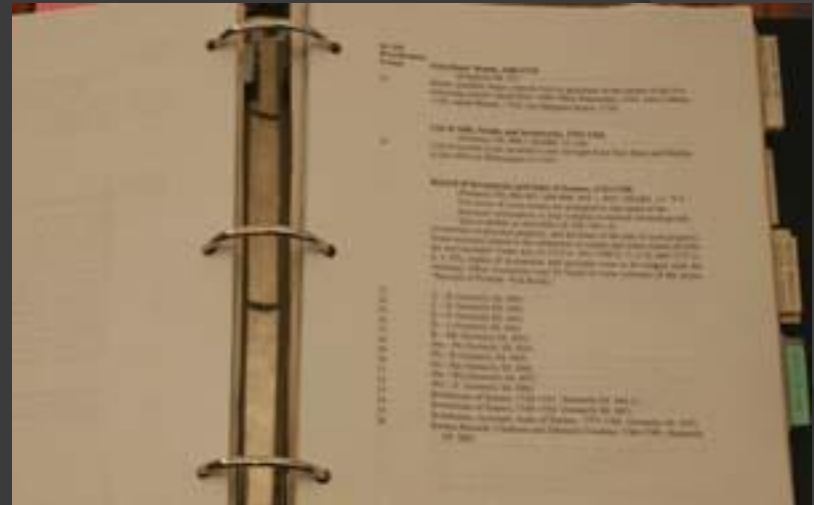


Kentuckiana Digital Library

- Displays finding aids for our own collections
- Collects finding aids from around the state
- Links to digitized manuscripts, diaries, photos, oral histories
- Example: [Postcard Collection](#)
[Stereoscopic Gems](#)

Finding aids

- May or may not:
 - Be electronic
 - Be encoded in EAD
 - Contain item-level detail
 - Link to digital collections



Finding aids let you know

- What unique documents exist in your area of interest
- If you want to travel to visit the archive
- If you want to bring the collection back from remote storage

Dublin Core

- ⦿ A content standard for bibliographic description
- ⦿ All elements optional
- ⦿ All elements repeatable

Simple Dublin Core

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

Qualified Dublin Core

- ⦿ Three more elements:
 - Audience Provenance RightsHolder
- ⦿ Qualifiers:
 - Date
 - Date.Created
 - Date.Valid
 - Date.Available
 - Date.Issued
 - Date.Modified

A Content Standard

- Has syntax guidelines for text, HTML, XML and RDF
- Schemas available for XML
- Recommended frameworks not a requirement – if you are using these elements, you are using DC!

A valid DC document (plain text)

- Title = "The Card Turner"
- Creator = "Louis Sachar"

Dublin Core in XML

```
<metadata>  
  <dc:title>The Card Turner</dc:title>  
  <dc:creator>Louis Sachar</dc:creator>  
</metadata>
```

Dublin Core in HTML/XHTML

```
<html>
  <head>
    <meta name="DC.title" content="The Card
Turner" />
    <meta name="DC.creator" content="Louis
Sachar" />
  </head>
  <body>
    ...
  </body>
</html>
```

Metadata Object Description Schema (MODS)

- XML format
- maintained by Library of Congress
- has corresponding authority format
MADS

MODS elements

- ⦿ Elements similar to MARC
 - Named fields instead of numeric
- ⦿ All elements optional
- ⦿ All elements repeatable

Advantages to MODS (XML)

- ⦿ More powerful nesting (see relatedItem)
- ⦿ XML is easier to process than MARC
 - More general tools
 - XSLT
- ⦿ XML can be more easily included in framework documents

MARCXML

```
<marc:datafield tag="245" ind1="1" ind2="0">  
  <marc:subfield code="a">How baking  
works</marc:subfield>  
  <marc:subfield code="h">[electronic  
resource] :</marc:subfield>  
  <marc:subfield code="b">exploring the  
fundamentals of baking science /  
</marc:subfield>  
  <marc:subfield code="c">Paula Figoni.  
</marc:subfield>  
</marc:datafield>
```

Metadata Encoding and Transmission Standard (METS)

- XML standard
- Maintained by Library of Congress
- A “framework” standard
- Good for large complicated objects, such as newspapers

METS framework

```
<mets>
```

```
  <metsHeader>
```

(data about this METS file)

```
  </metsHeader>
```

```
  <dmdSec>
```

(descriptive metadata, often
MODS)

```
  </dmdSec>
```

METS framework (continued)

```
<amdSec>
```

(administrative / technical metadata)

```
</amdSec>
```

```
<fileSec>
```

(file metadata)

```
</fileSec>
```

```
<structMap>
```

(structural metadata)

```
</structMap>
```

```
</mets>
```

National Digital Newspaper Program (NDNP)

- ⦿ Joint project of NEH/LC to digitize newspaper content from microfilm
- ⦿ Creation of website *Chronicling America*
 - 3,474,161 searchable newspaper titles, 1859-1922
- ⦿ Uses several standards in a METS framework
- ⦿ Batches are validated with a standalone validator program (DVV)

NDNP at University of Kentucky

- Kentucky among first six states
- All work was done in-house (other states outsourced)
- Project at UK run by a UK SLIS graduate, Kopana Terry
- Other states will get awards, looking for project managers

METS profiles

- ⦿ Individual applications use METS in different ways, described by profiles
- ⦿ Narrative documents describing what documents should contain
- ⦿ Nice profiles contain example documents
- ⦿ Not “machine actionable” – cannot validate against a profile

Other METS projects at UK

- ◎ New implementation of Kentuckiana Digital Library using BlackLight
 - Uses Dublin Core for bibliographic description
 - Files in a METS framework

- ◎ Enterprise Repository
 - Data stored in a proprietary system
 - Parallel METS version for flexibility

Text Encoding Initiative

- For encoding of literary and linguistic texts
- International standard maintained by TEI Consortium
- XML format

TEI guidelines

- Current version of guidelines is P5 (2007)
- Two large volumes
- Detailed rules about drama, manuscripts, etc
- Not all projects need all features

Make your own schema!

- Make an ODD file (“One Document Does it all”)
- Submit to online application Roma
- Generate your own TEI schema to validate against

Encoding levels

- Level 1: Fully automated conversion and encoding
- Level 2: Minimal encoding
- Level 3: Simple analysis
- Level 4: Basic content analysis
- Level 5: Scholarly encoding projects

TEI projects at University of Kentucky

- ◎ Kentuckiana Digital Library
 - Digitized books
- ◎ Carolingian Canon Law project
 - medieval legal manuscripts
 - Collaboratory for Research in Computing for Humanities

“Words don’t have meaning,
they have usage.”

The same is true of
metadata standards.

References

- ⦿ Encoded Archival Description (EAD)

<http://www.loc.gov/ead/>

- ⦿ Text Encoding Initiative (TEI)

<http://www.tei-c.org/index.xml>

- ⦿ National Digital Newspaper Program
Technical Specification

<http://www.loc.gov/ndnp/techspecs.html>

References

- Dublin Core

<http://dublincore.org/>

- Metadata Object Description Schema (MODS)

<http://www.loc.gov/standards/mods/>

- Metadata Coding and Transmission Standard (METS)

<http://www.loc.gov/standards/mets/>

References

- W3C Markup Validation Service

<http://validator.w3.org>

- Kentuckiana Digital Library

<http://kdl.kyvl.org/>