Digital Collections at St. Olaf

Metadata Field Selection: Method or shot in the dark?

Jill Strass

Upper Midwest CONTENTdm Users Group

October, 2009

Key findings about metadata selection

- User need
- Are data standards (like Dublin Core and VRA) an answer?
- Is there a connection between metadata and collection maintenance and control?
- Some metadata fields are also system requirements
- Descriptive metadata
- Document project decisions with a data dictionary

A word about the importance of the Data Dictionary

- Data dictionary lists terms used, reasons for use, samples and context for use.
- Shared document that shows all involved parties the data fields.
- Reminds project participants of decisions made.
- Makes consistency in data entry possible.

Case Study #1: Manitou Messenger

- Student lit mag that morphed into newspaper, has serial characteristics.
- Choose metadata with search options in mind
- Created a data dictionary which lists metadata terms and their purpose.
- Used encoded file names for scans.

Case Study #1: Manitou Messenger

- Search and browse considerations without customization or ability to individually examine and catalog each page.
- Use the available interface to emulate volumes and issues.
- Make educated guesses on Dublin Core and homespun fields.

Case Study #1: Manitou Messenger

Mostly likely used search options for end-users:

- Date (both year and decade)
- Volume
- Issue
- Keyword

Metadata for Collection Maintenance or Control:

- Collection Name
- Date Uploaded

Case Study #1: Manitou Messenger Metadata Choices

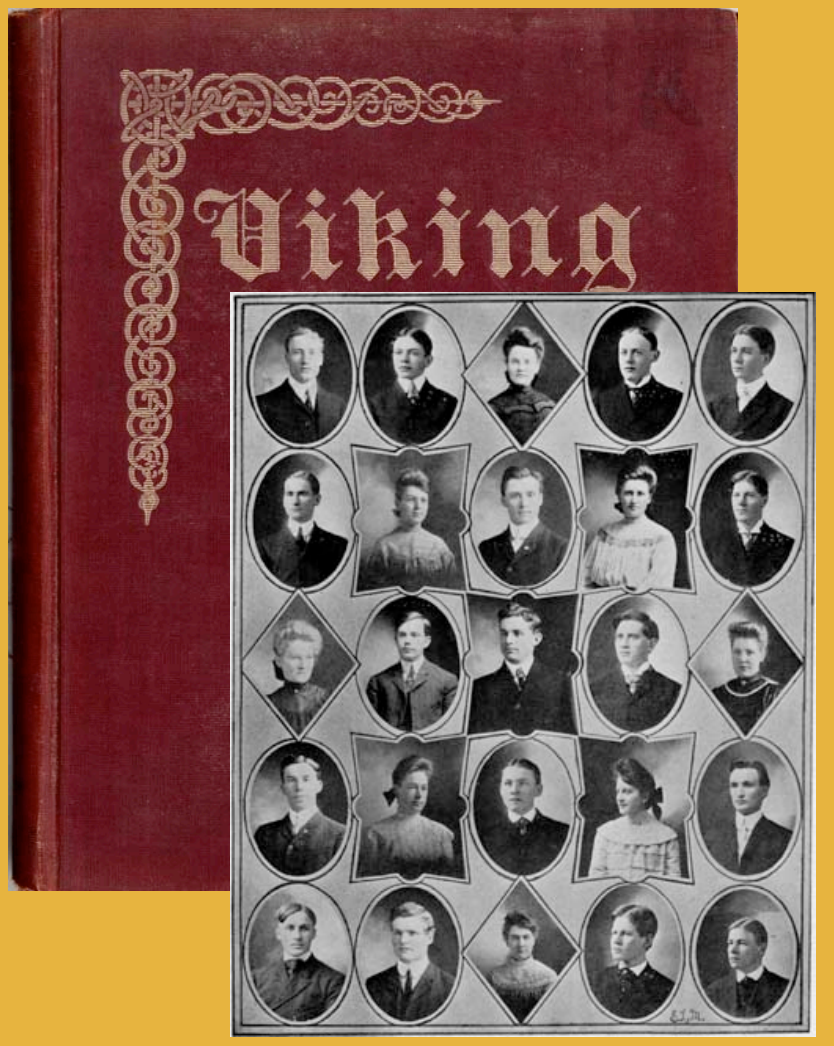
Dublin Core	St. Olaf	CONTENTdm
Title	Collection Name	CDM_LVL
Rights	Volume	CDM_LVL_NAME
Format	Issue	***Title***
Type	Month	Full Text
Creator	Year	Full Resolution
Source	Decade	Local Identifier
Publisher	Page Value	File name
Relation	Frequency	
Coverage	Original Pub. Date	
Language	Date Uploaded	

Case Study #1: Manitou Messenger Complete Data Dictionary

	Data Dictionary for Manitou Messenger	
2	**Field Name**	Field Defined
3		
4	**CDM_LVL**	Required monograph field for CDM
5	**CDM_LVL_NAME**	Required monograph field for CDM
6	**Title**	Name that will appear for individual page in sidebar for user to select
7	**Volume**	Volume number of newspaper
8	**Issue**	Issue number of newspaper
9	**Month**	Month newspaper appeared
10	**Year**	Publication year (year first appeared in print)
11	**Decade**	Decade that newspaper appeared in, i.e. 1800, 1810, 1820....
12	**Page value**	Number of page
13	**Frequency**	How frequently the paper was published, i.e. Monthly, Weekly
14	**Rights**	Copyright statement
15	**Type**	Newspaper
16	**Format**	Printed text
17	**Original Publication Date**	Date paper was originally published i.e. 1887-04 Month and day paper originally appeared, YYYY-MM and then YYYY-MM-DD using W3C.org data standard
18	**Date Uploaded**	Date file was uploaded to CONTENTdm
19	**Full Text**	Used for OCR software
20	**Local Identifier**	Filename of image used for archiving purposes as CONTENTdm deletes the object filename field when producing a backup file.
21	**Full Resolution**	Original images are tiff and are full resolution images
22	**Collection Name**	Name of the Collection i.e. Manitou Messenger
23	**Creator**	Original creators of newspaper content i.e. Students at St. Olaf College
24	**Publisher**	St. Olaf College Northfield MN
25	**Source**	Printed and bound paper copies of original newspaper
26	**Relation**	Manitou Messenger student paper from St. Olaf College http://www.stolaf.edu/library/digital/manitoumess.html
27	**Coverage**	Newspaper published from 1887 through present day
28	**Language**	Language of text i.e. English with occasional Norwegian. If we can ever segment articles, a side project would be adding metadata to show the language for individual articles.
29	**Filename**	Filename of image.
30		

Case Study #2: The Viking Yearbook

- College Yearbook – mostly student-published, hardbound. Published annually.
- Users mostly genealogical researchers, faculty, Archives staff, students.
- Has its own folder in CONTENTdm, each “object” is volume of the yearbook.



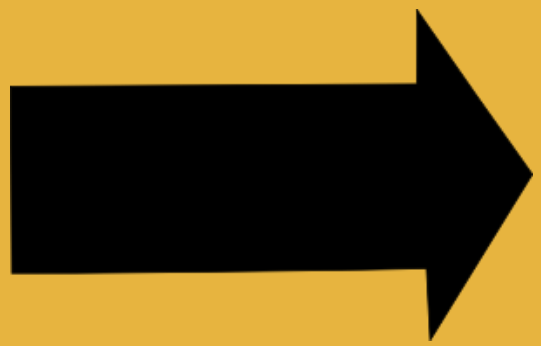
Case Study #2: The Viking Yearbook

- Very similar to Manitou Messenger in process, but we leverage those filenames even more.
- Add info to page metadata about whether pages are of faculty or students.
- Attempted to do freshman, sophomore, junior, senior, but to our surprise, the yearbooks didn't uniformly put classes all together on a page.
- Our student publishers experiment by putting three years into the title, which creates a challenge called 191019111912

Case Study #2: The Viking Yearbook

- To help users who will be browsing the yearbook, we add the letter ‘f’ or ‘s’ for pages that had a name and image of faculty or students, respectively.

Here is some raw metadata for the page names; the names help illustrate the content.



Page 000
Page 000
Page 000 Blank Page
Page 000 Blank Page
Page 000 Blank Page
Page 000
Page 001 Faculty
Page 002
Page 003
Page 004
Page 005
Page 006
Page 007
Page 008
Page 009
Page 010 Blank Page
Page 011
Page 012
Page 013
Page 014
Page 015
Page 016 Faculty
Page 017 Faculty
Page 018 Faculty
Page 019 Faculty
Page 020 Faculty
Page 021
Page 022
Page 023
Page 024
Page 025
Page 026 Blank Page
Page 027
Page 028
Page 029
Page 030 Students
Page 031 Students
Page 032 Students
Page 033 Students
Page 034 Students
Page 035 Students
Page 036 Students
Page 037

Case Study #2: The Viking Yearbook

- For the 191019111912 title, we used some tricks in Excel to parse out the years from the filename, so that the publishers choice would be reflected in the display.

CDM_LVL	CDM_LVL_NAME	Title
		Viking Yearbook: 1910 1911 1912
1	Viking Yearbook: 1910 1911 1912	Page 000
1	Viking Yearbook: 1910 1911 1912	Page 000
1	Viking Yearbook: 1910 1911 1912	Page 000 Blank Page
1	Viking Yearbook: 1910 1911 1912	Page 000 Blank Page
1	Viking Yearbook: 1910 1911 1912	Page 000 Blank Page
1	Viking Yearbook: 1910 1911 1912	Page 000
1	Viking Yearbook: 1910 1911 1912	Page 001 Faculty
1	Viking Yearbook: 1910 1911 1912	Page 002
1	Viking Yearbook: 1910 1911 1912	Page 003
1	Viking Yearbook: 1910 1911 1912	Page 004
1	Viking Yearbook: 1910 1911 1912	Page 005
1	Viking Yearbook: 1910 1911 1912	Page 006
1	Viking Yearbook: 1910 1911 1912	Page 007
1	Viking Yearbook: 1910 1911 1912	Page 008
1	Viking Yearbook: 1910 1911 1912	Page 009
1	Viking Yearbook: 1910 1911 1912	Page 010 Blank Page
1	Viking Yearbook: 1910 1911 1912	Page 011
1	Viking Yearbook: 1910 1911 1912	Page 012
1	Viking Yearbook: 1910 1911 1912	Page 013
1	Viking Yearbook: 1910 1911 1912	Page 014
1	Viking Yearbook: 1910 1911 1912	Page 015
1	Viking Yearbook: 1910 1911 1912	Page 016 Faculty
1	Viking Yearbook: 1910 1911 1912	Page 017 Faculty
1	Viking Yearbook: 1910 1911 1912	Page 018 Faculty
1	Viking Yearbook: 1910 1911 1912	Page 019 Faculty
1	Viking Yearbook: 1910 1911 1912	Page 020 Faculty
1	Viking Yearbook: 1910 1911 1912	Page 021
1	Viking Yearbook: 1910 1911 1912	Page 022
1	Viking Yearbook: 1910 1911 1912	Page 023
1	Viking Yearbook: 1910 1911 1912	Page 024
1	Viking Yearbook: 1910 1911 1912	Page 025
1	Viking Yearbook: 1910 1911 1912	Page 026 Blank Page
1	Viking Yearbook: 1910 1911 1912	Page 027
1	Viking Yearbook: 1910 1911 1912	Page 028
1	Viking Yearbook: 1910 1911 1912	Page 029
1	Viking Yearbook: 1910 1911 1912	Page 030 Students
1	Viking Yearbook: 1910 1911 1912	Page 031 Students
1	Viking Yearbook: 1910 1911 1912	Page 032 Students
1	Viking Yearbook: 1910 1911 1912	Page 033 Students

Case Study #2: The Viking Yearbook Metadata Choices

Dublin Core	St. Olaf Viking	CONTENTdm
Title	Collection Name	CDM_LVL
Rights	Yearbook Date	CDM_LVL_NAME
Format	Decade	***Title***
Type	Page Number	Full Text
Creator	Date Uploaded	Full Resolution
Source		Local Identifier
Publisher		File name
Relation		
Coverage		
Language		

Case Study #3: Archives at St. Olaf goes Digital

- Archives: 111 images
- NAHA (Norwegian American Historical Association): 85 images

But Dublin Core doesn't do it all….

- Find a role model
- Construct a Crosswalk from the role model to the data in use
- Add in a few field concepts that the model doesn't have
- The resulting list of terms is the data dictionary.

Case Study #3: Archives finds a role model: Claremont College

Theatrical production in the Wash, Pomona College

Title	Theatrical production in the Wash, Pomona College
Subject-LCSH	Pomona College (Claremont, Calif.) College theater Theater, Open-air Actors College students
Subject-Local	Blanchard Park (The Wash) - Pomona College
Description	Students perform on a makeshift outdoor stage during the 1904 Wash Day program put on by the senior class. One student standing in the center is reading a piece of paper, while the students sitting behind him are reading books or looking bored.
Notes	For a description of the 1904 Wash Day program at Pomona College, see Los Angeles Times, June 23, 1904, pg. A6.
Publisher	Honnold Mudd Library. Special Collections
Date	1904
Language	eng
Source	Glass plate negative, 7 x 5 inches: no title; The Boynton Collection of Glass Negatives
Accession Number	bce00270
Relation	Boynton Collection of Early Claremont - http://ccdl.libraries.claremont.edu/col/bce
Rights	For more information on copyright or permissions for this image, please contact Honnold Mudd Library Special Collections at http://libraries.claremont.edu/sc
Type	Image
Format	image/jp2
Object File Name	bce00270.tif

Claremont	Archives	CONTENTdm
Title	Title	***Title***
Subject - LCSH	Subject	Local Identifier
College Students	Content Description	File name
Portrait Photography	Owner	
Costume	Original Resource Creator	
Architecture, Domestic	Creator Name	
Description	Rights	
Notes	Resource Type	
Publisher	Format	
Date	Identifier	
Language	Collection Name	
Source	Medium	
Accession Number		
Relation		
Rights		
Type		
Format		
Object File Name		

Case Study #3: Archives sample record

Ball players

Title	Ball players
Subject	Communities: Recreation
Content Description	First St. Olaf baseball team: students and Halvor T. Ytterboe, faculty member.
Owner	StO Archives
Original Resource Creation Date	1888
Creator Name	Felland, O.G. (Publisher: Northfield, Minn.)
Rights	See copyright band on image.
Resource Type	Still Image: Photograph
Format	image-jpg
Collection Name	Photo Collection
Medium	B&W, 4 3/4 x 7 1/4 inches (print from glass negative)
Photo ID Number	f0225
Digital Resource Creation Date	7/29/2008

Case Study #4: Art Museum

- Art Museum – image collection represents museum holdings and originally started as a gift-tracking, and inventory - control system.

Case Study #4: Art Museum – audience and standard to use

- Envisioning the probable end-user was tough for this project. In our choices, there was a division between serving an inventory need and more detailed info about artwork.
- Metadata Standard – We looked at VRA (Visual Resources Association) and used some of their concepts.
- Used the existing database nomenclature, with some experimentation with terms.

Case Study #4: Art Museum – role model

- Luther College was the role model, since its holdings and potential audience most closely matched St. Olaf.
- Also drew (pardon pun) on Five College Museums; Grinnell Art Collection; Smart Museum of Art.

Case Study #4: Art Museum – Luther sample record

The Last Portal (Opus 96)
Fridtjof Schroder

Accession number: 1999:12:01
Date of work: 1958
Subject: Abstract
Category: Painting
Medium: Oil
Artist nationality: American
Artist life dates: 1917-1990
Current location: Korsrud Annex

Dimensions of art:	**Unframed**	**Framed**
English	47 1/2 x 23 1/4	49 1/2 x 25 1/2
Metric	121 x 59 cm.	125.5 x 64 cm.

Case Study #4: Art Museum – adapting VRA for our use

- Examined VRA concepts and used most specific concept within the hierarchy. If the concept didn't make sense on its own, we added words to it.
- Agent (dates) → Life → Artist Life Dates

Case Study #4: Art Museum – sample record

Title	4 x 8 (2)
Artist Name	Ewald '87, Jill
Life Dates	b. 1950
Nationality	United States
Date of Work	2007
Category	Sculpture
Medium	Bronze , Steel
Format	Sculptural
Fabrication	Cast
Culture	American
Description	Eight (8) abstract spherical shapes connected and floating with copper and green oxidation colors
Unframed Height (in)	3.75
Unframed Width (in)	6
Depth (in)	5
Signature	Yes
Credit Line	On Loan from Dan '69 and Nancy Schneider
Digital ID	166_2007_Ewald.jpg
Rights	The image and the text corresponding to this image may only be used for noncommercial educational and scholarly purposes and must be acknowledged as property of Flaten Art Museum of St. Olaf College.
Circulation Rating	Yellow
Accession Number	2007.166
Upload Date	6/18/2009
Contact Information	weselman@stolaf.edu

Key findings about metadata selection

- User need – put this first
- All data standards (like Dublin Core and VRA) we used had to be modified
- Consider how you will control and maintain the collection and how metadata can help (this is a form of user need)
- Some metadata fields are also system requirements
- The reasons above are why descriptive metadata is an afterthought for many collections
- Document decisions with a data dictionary