

From Cataloging to Metadata

Differences in Scope, Skills, and Standards

Carolyn Hansen, Sean Crowe

ALA Midwinter Conference
CaMMS Cataloging Norms Interest Group
January 25, 2014



Introduction

Increasing number of catalogers doing metadata work, as digital initiatives and repositories require more staff bandwidth. What do traditional catalogers need to know to be successful in metadata work?

Structure of presentation – talk about differences between cataloging/metadata, illustrate these differences by describing a case study in which we transformed over 9,000 records in Qualified Dublin Core to VRA.



Carolyn

Metadata
Librarian



Sean

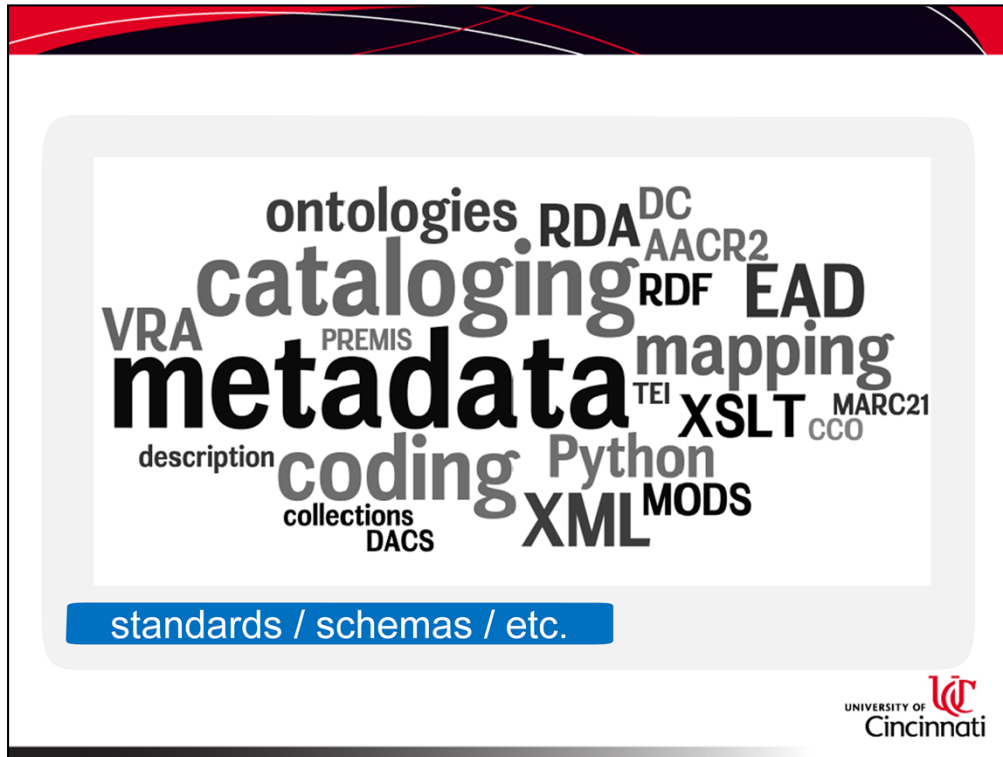
Electronic Resources
Librarian



Talk about our partnership

Carolyn has strong metadata knowledge/experience, Sean has strong technical skills

Partnership illustrates collaborative nature of metadata work



Traditional cataloging

- Bibliographic description in a MARC, AACR2/RDA environment


Metadata

- NISO definition, composed of descriptive, structural, and administrative (includes rights management and preservation) metadata
- Used in many different types of library systems and platforms. At UC, we have various metadata schemas in our Sierra ILS, DSpace repository, Luna repository, Hydra repository, Research Directory, etc.
- Will be talking about metadata projects mostly in the context of metadata transformations

As a cataloger, aren't I doing metadata work already? Yes! Traditional catalogers work with descriptive and some administrative metadata.

Developing a baseline of knowledge of schemas, it's not realistic to be an expert on all schemas, coding languages, etc. Collaborate with subject specialists and technical specialists as much as you can.

Metadata work involves having less control over description, authority, etc. (for example, your system for digital collections may not have authority control)



metadata work is ...

- voluminous
- technical
- collaborative
- contextual
- less controlled

Voluminous:

In general, you'll work with a higher number of records at one time. Metadata entry is likely to be done on a individual record level by student assistants or staff. Metadata librarians are involved with creating input standards, repurposing and transforming existing metadata

Technical:

You'll need to acquire technical skills to work with large batches of metadata. In particular, knowledge of scripting languages is important.

Collaborative:

Nature of the work is less insular than traditional cataloging. Greater collaboration between you and other departments (ex. At UC, I work with Tech Services, Archives and Rare Books, and Digital Repositories/Digital Collections)

Contextual:

Dealing with records in context with each other; often based on collection. You'll need to make descriptive decisions that are consistent across collections. For example, what is the lowest common denominator of subject access for a collection?

Less controlled:

Metadata is often inputted by non-catalogers, which can lead to quality control issues. Digital repositories and digital collection platforms often have no functionality for authority control. You may also deal with legacy metadata that was created in less than ideal circumstances. Metadata work involves trying to create control in a fluid environment. However, metadata work can also be constrictive, since you'll still need to work within the constraints of different schemas when mapping

essentials for metadata transformation

- intellectual framework (mapping)
- automation tools
- computer screen real estate

Prior to building the crosswalk, we felt it would be good to develop a strong mapping document, a “road map” for mapping

Dealing with such a large set of records precludes manual one-by-one editing. With time constraints in mind, we needed tools that can affect reliable and uniform changes to the metadata. Building on a custom code library from previous tasks, we created a point and click tool to process the files.

Examining and comparing spreadsheets is made so much easier with ample screen space, shared time on dual monitor workstation helped review go smoothly.

collection

- Over 9k images
- Rich historical detail
- Streetscapes
- Construction
- Structural damage
- Interiors



Rapid Transit Photographs -- Box 19, Folder 11 (December 2, 1926) -- negative, 1926-12-02, 2:38
P.M. Cincinnati Subway and Street Improvement Records 1916-1955, University of Cincinnati
Libraries Digital Collection < <http://digproj.libraries.uc.edu:8180/luna/servlet/s/5z47x8>>

UNIVERSITY OF
Cincinnati

LSTA grant to digitize images within a sub-collection of the *Register of Office Records of The City Engineer City of Cincinnati, 1851-1957*

Photographic prints, negatives, including unstable nitrate negatives

Images and metadata generated by contractor, reviewed locally, loaded and published in DRC.



DIV. OF HIGHWAYS
11-12-33
FIFTH ST. VIADUCT - LONGVIEW DIST. PHOTOS - BOWLING GREEN ROUTE AFTER IMP.

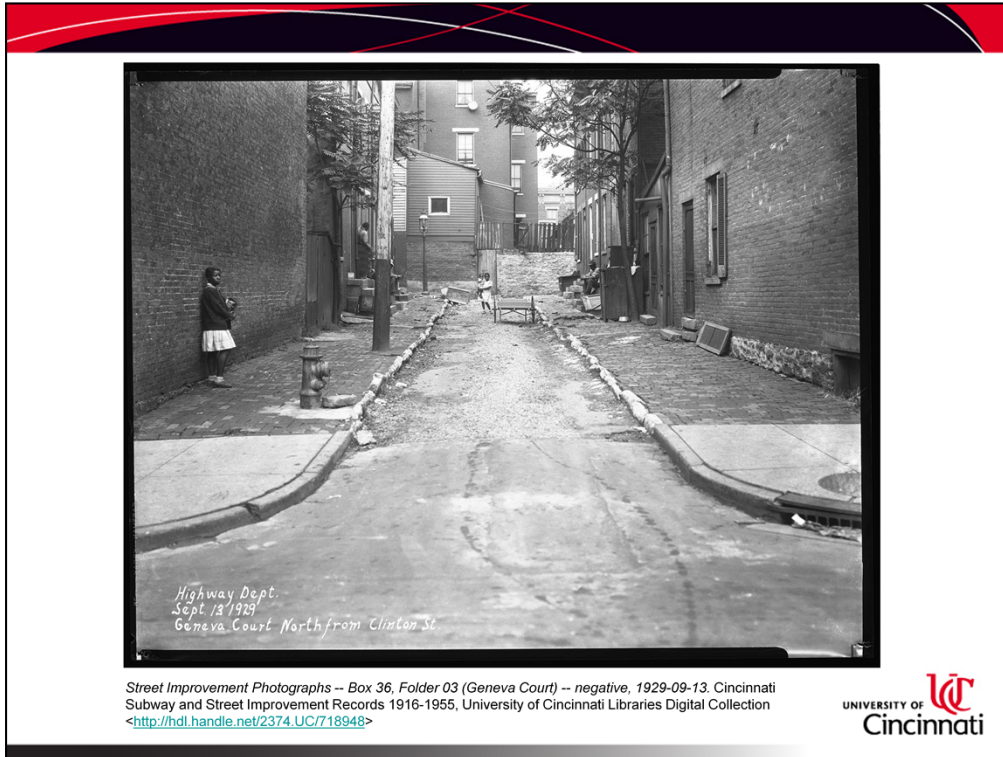
Street Improvement Photographs -- Box 35, Folder 35 (Fifth Street) -- negative, 1938-11-15, Cincinnati Subway and Street Improvement Records 1916-1955, University of Cincinnati Libraries Digital Collection <<http://hdl.handle.net/2374.UC/717193>>



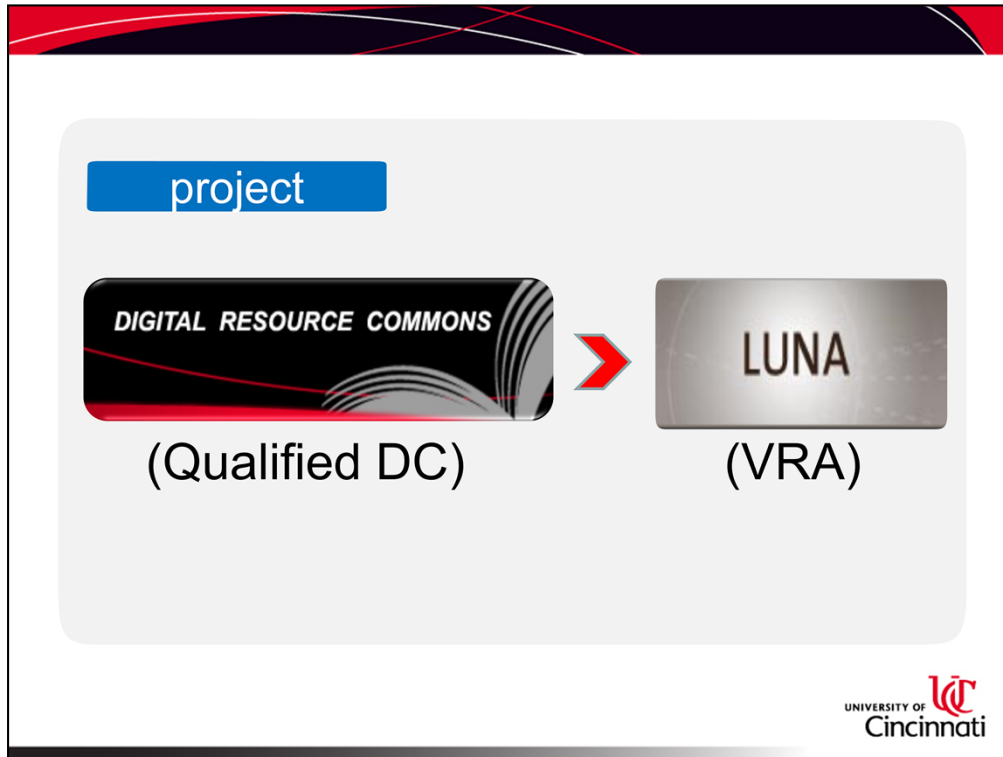


Rapid Transit Photographs -- Box 12, Folder 24 (November 22, 1926) -- print, 1926-11-22, 3:00 P.M. Cincinnati
Subway and Street Improvement Records 1916-1955, University of Cincinnati Libraries Digital Collection
<<http://hdl.handle.net/2374.UC/703593>>





Beautiful and candid images that show not only built world transformation, but also snapshot of every day life – children peaking out a window, workers taking a break, period interiors.



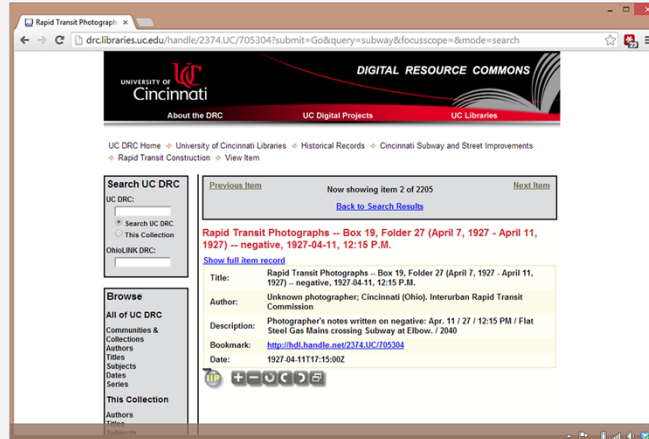
Project: add images and cross-walked metadata to local Luna Repository.

Luna repository generally geared toward image content for DAAP

Luna Repository uses VRA

Project: convert over 9000 Qualified Dublin Core records to VRA, load records and images to Luna repository

Digital Resource Commons



In preparation for the project, the metadata set was exported in XML/Qualified Dublin Core, separated into folders based on box assignment.


mapping



Mapping is an attempt to create a one to one framework between standards but each standard has its own structure/rules, which makes lossless mapping difficult.

DC to VRA

Qualified DC	VRA 3.0	Notes	Vocabularies
Derived from resource location	Work Record ID		
Derived from resource location	Reproduction Record ID		
title	Thumbnail Title		
	For back of photograph, add "(back of photograph)" after title taken from dc title field		
contributor:author	Thumbnail Creator	If photographer's name is known, it should be used	LCNAF
date:created	Thumbnail Date	Format: mm/dd/yyyy, remove timestamp	
series (see below)	Thumbnail Work Type		
format.medium 'photographic prints'	For gelatin silver print, take from: format.medium 'photographic prints'	Plural in DC, use singular form for VRA	
VA	For gelatin silver negative: negative (photographic)		
VA	For acetate negative: negative (photographic)		
VA	For glass negative: negative (photographic)		
type	Work Class	Singular in DC, use plural in VRA	Prefer use of AA1
series (see below)	Work Type		Prefer use of AA1
format.medium 'photographic prints'	For gelatin silver print, take from: format.medium 'photographic prints'	Plural in DC, use singular form for VRA	
VA	For gelatin silver negative: negative (photographic)		
VA	For acetate negative: negative (photographic)		
VA	For glass negative: negative (photographic)		
title	Title		
	For back of photograph, add "(back of photograph)" after title taken from	Both front and back titles are constructed titles	
VA	Title Tune		



Start with a mapping document that expresses one-to-one relationship as much as possible.

Create lowest common denominator subject access, because you won't be able to do much hand editing with such high volume of records. This is an example of a difference between traditional cataloging and metadata work. Whereas in traditional cataloging you generally analyze by the piece (serials excluded), in high volume metadata work you are making more general collection-level subject assignments.

Mapping Challenges: VRA is much more granular than DC; dealing with legacy metadata/ descriptive policies; metadata may have been created by student workers, quality control issues, authority control issues

The slide contains a code editor window on the left and a terminal window on the right. The code editor shows the following Python code:

```

9 class easyCSV:
10     def __init__(self, filename):
11         t = csv.reader(open(filename, 'r'))
12         data = [row for row in t]
13         self.headers = data[0]
14         self.data = data[1:]
15         self.empty_set = [[] for x in xrange(len(data))]
16         self.linkdata = [row[0].split('\t') for row in data]
17
18     def change_header_name(self, data, find, replace):
19         index = data[0].index(find)
20         data[0].remove(find)
21         data[0].insert(index, replace)
22         return data

```

The terminal window shows the following output:

```

Starting number from LinkData.txt sequence (start at 1):
Sean Crowe, Carolyn Hansen -- 09-09-2013

```

Below the code editor, there is a URL: https://github.com/crowesn/DC_to_VRA

In the bottom right corner, there is the University of Cincinnati logo.

Toolkit: Python, OpenOffice, Vim (or any good text editor)

CSV input files – created from DRC xml exports

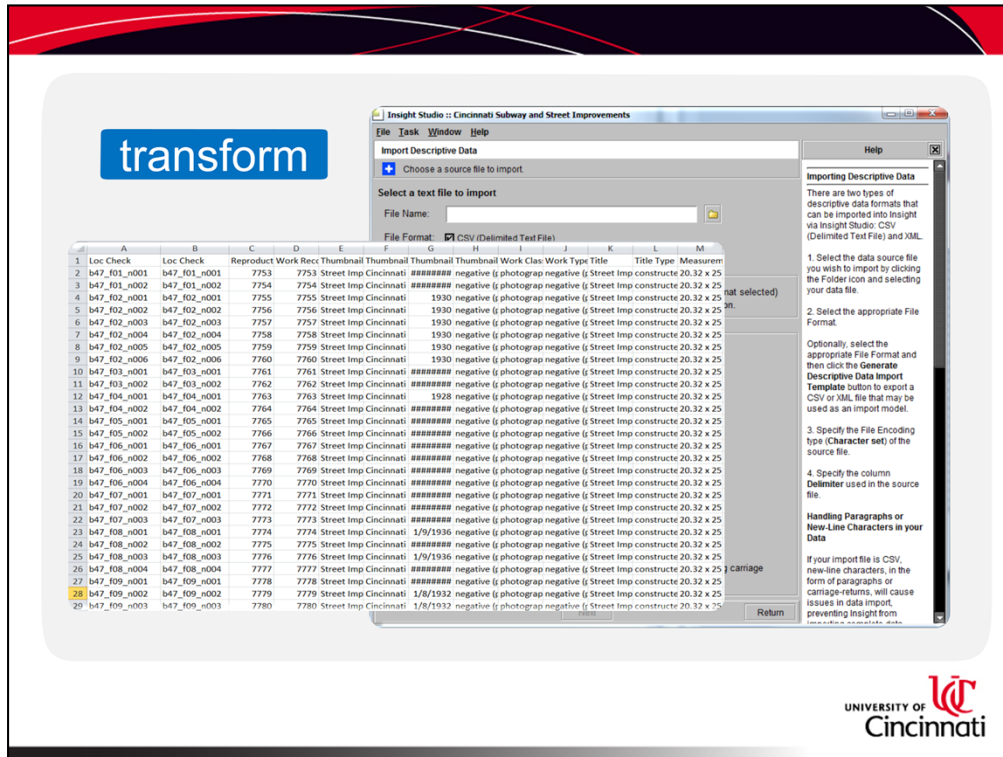
Extra steps for working with Unicode (prefer OpenOffice Calc over Excel).

We created a clickable script to open DC.csv, parse, build, and output VRA.csv

The script uses a custom class for working with data by column. The class includes functions for frequently recurring tasks such as changing headers; adding, copying, inserting and deleting columns; as well as iterating through a column and editing field contents.

Initial development time was ~12 hours. ~2 hours more for tweaking/adjustments

Challenges – adding new fields based on other fields (new content to conform to VRA, terms from Getty, etc) Need clear and comprehensive mapping documentation.



Run script for each of ~50 files

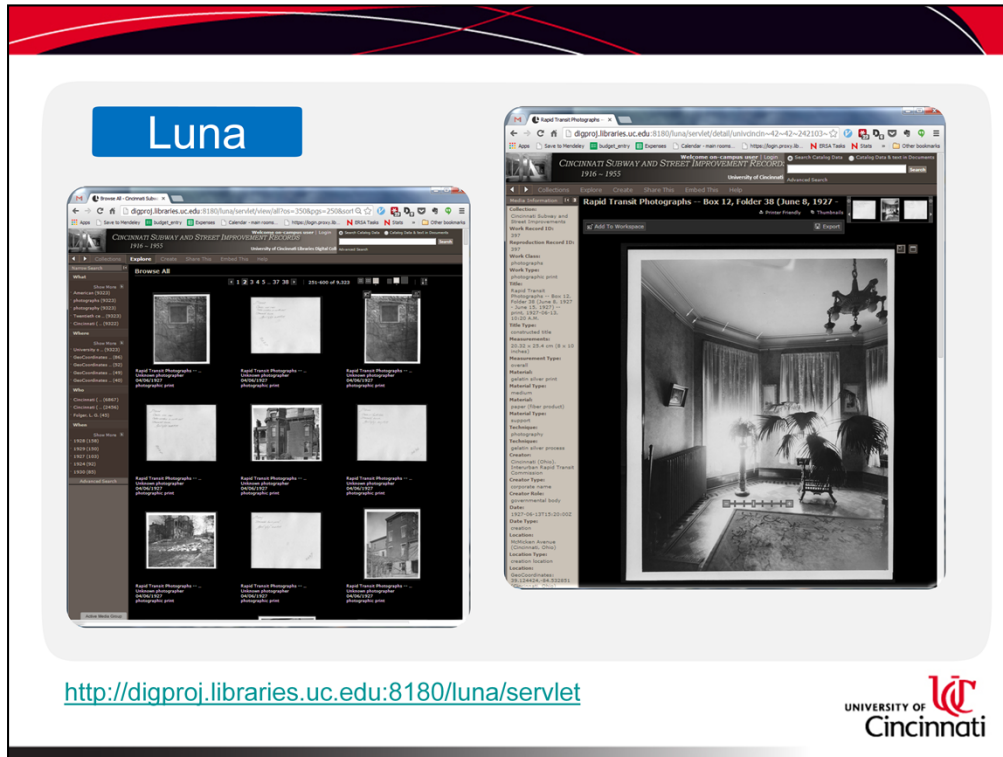
Script maps existing data elements from Qualified DC to VRA as well as creates some VRA conforming elements, conditional on DC elements

Created validation scripts to check for errant or unexpected content as well as duplicate records for print versos

Challenge:

Bottlenecking with Luna software – significant time investment in upload images and metadata to Luna platform.

In addition to metadata and images, Luna uses independent *linking* document, a tab-delimited file that associates IDs for image and record.




Luna is a repository platform that offers tools for viewing and sharing images.

Users can browse, zoom, crop and share images with ease. Searching is well supported with the finer granularity of VRA


Follow-up work will include ingesting metadata for this collection into our discovery layer.

connect



 @meta_caro
carolyn[dot]hansen
@uc[dot]edu



 @s_crowe
sean[dot]crowe@uc
[dot]edu

Resources

Dublin Core

<http://www.dublincore.org>

VRA

<http://www.vraweb.org/>

Python

<http://www.python.org>

Unicode/character encoding

<http://kunststube.net/encoding/>



Thank You!

github

https://github.com/crowesn/DC_to_VRA

Image from: *Rapid Transit Photographs -- Box 12, Folder 07 (September 15, 1926) -- print, 1926-09-15, 10:00 A.M.* Cincinnati Subway and Street Improvement Records 1916-1955, University of Cincinnati Libraries Digital Collection <<http://digproj.libraries.uc.edu:8180/luna/servlet/s/y4418s>>

