MAKING VENDOR-GENERATED METADATA WORK FOR ARCHIVAL COLLECTIONS USING VRA AND PYTHON

Carolyn Hansen, Metadata Librarian, University of Cincinnati
Sean Crowe, Electronic Resources Librarian, University of Cincinnati

Introduction

Cataloging cultural resources typically requires more descriptive granularity than standard library materials.

Stemming from a project to convert metadata from Dublin Core to VRA, the University of Cincinnati Libraries outlines a successful workflow to improve vendor-generated metadata for a large digital collection of archival materials.

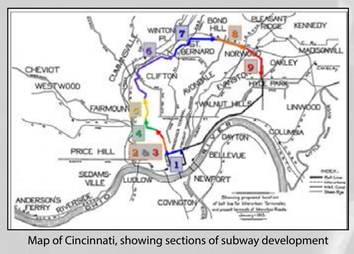
Collection

Content

- Ca. 9,000 digitized images from an extensive archive of the City Engineer of the City of Cincinnati
- Mostly photographic prints and negatives

Subjects

- Unfinished Cincinnati Subway System
- Early 20th century transportation
- Urban history
- Social history



Map of Cincinnati, showing sections of subway development

Metadata & Access

Dublin Core (DC) metadata created by vendor during scanning process; limited library control over metadata mapping and workflow

Prior to project, metadata records and digital objects were accessible on UC's Digital Resource Commons (DRC); not available on UC's LUNA image repository, where most UC photographic collections are stored with VRA metadata.

Challenges

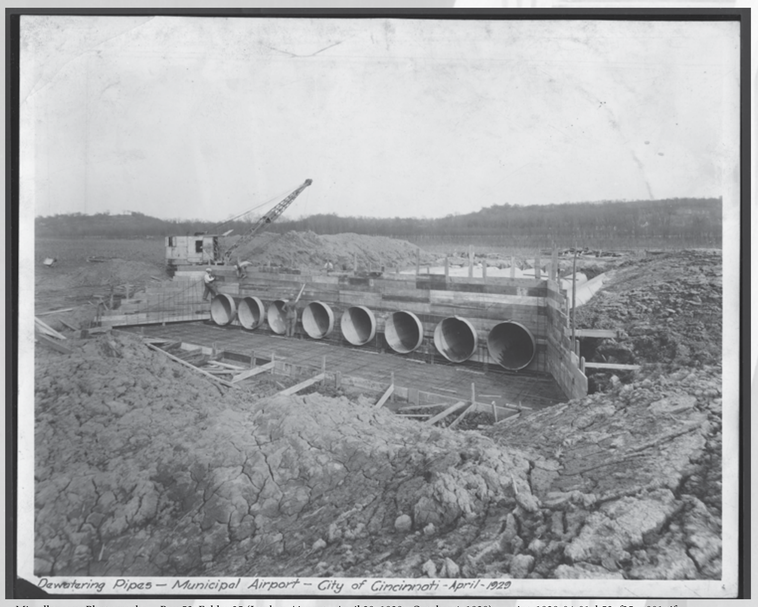
Conversion from Dublin Core to more granular VRA requires thoughtful mapping of data fields and adding additional fields, e.g. VRA describes original object AND digital surrogate. Project involved extensive use of Getty Research Institute's Art & Architecture Thesaurus (AAT) for descriptive vocabulary to fill in gaps in description.

Planning and specification for the contract-scanning project did not include input from Library Technical Services, the vendor-generated metadata has issues with accuracy and consistency.

Miscellaneous Photographs -- Box 53, Folder 13 (Central Parkway) -- print, 1929, b53_f13_p002.tif

Street Improvement Photographs -- Box 35, Folder 42 (Florence Avenue) -- negative, 1926-11-23, b35_f42_n002.tif

Street Improvement Photographs -- Box 36, Folder 03 (Geneva Court) -- negative, 1929-09-13, b36_f03_n001.tif



Miscellaneous Photographs -- Box 53, Folder 25 (Lunken Airport - April 29, 1929 - October 4, 1929) -- print, 1929-04-01, b53_f25_p001.tif

Conversion Project

Convert vendor-generated metadata in DC to VRA, for inclusion in UC's LUNA repository. Records stored in CSV files.

Workflow

1. Research initial vendor metadata creation.

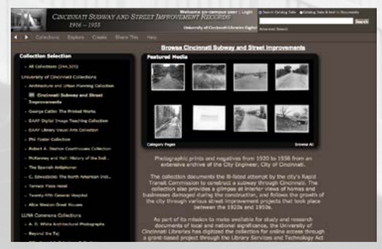
2. Create metadata map from DC to VRA; used as a reference document for the the change script.

	VRA 3.0	Notes
source location	Work Record ID	
source location	Reproduction Record ID	
	Thumbnail Title	
	For back of photograph, add "(back of photograph)" after title taken from dc:title field	
or	Thumbnail Creator	If photographer's name is known, it should be used
	Thumbnail Date	Format: mm/dd/yyyy, remove timestamp
e]	Thumbnail Work Type	
'photographic prints'	For gelatin silver print, take from: format.medium 'photographic prints'	Plural in DC, use singular form for VRA
	For gelatin silver negative: negative (photographic)	
	For acetate negative: negative (photographic)	
	For glass negative: negative (photographic)	
	Work Class	Singular in DC, use plural in VRA
e]	Work Type	
'photographic prints'	For gelatin silver print, take from: format.medium 'photographic prints'	Plural in DC, use singular form for VRA
	For gelatin silver negative: negative (photographic)	
	For acetate negative: negative (photographic)	
	For glass negative: negative (photographic)	
	Title	
	For back of photograph, add "(back of photograph)" after title taken from Both front and back titles are constructed titles	
	Title Type	
	Value: constructed title	

3. Use Getty Vocabularies to specify describe original physical object and digital surrogate.

4. Develop Python script to parse CSV and execute changes to metadata.

5. Batch process changes to descriptive metadata and load into LUNA repository.



Conclusion

Since the collection was posted in Fall 2013, it has received over 17,000 unique page-views in the Luna Repository. This project serves as a template for future shared, interdepartmental projects. Further collaboration is certain as traditional Library Technical Services operations evolve to support local and unique digital content, including research data, archival material, and beyond.

References

University of Cincinnati Libraries (2014). Cincinnati Subway and Street Improvements, 1916-1955. Retrieved from http://digital.libraries.uc.edu/subway/

University of Cincinnati Libraries (n.d.). LUNA Digital Repository. Retrieved from http://dig-proj.libraries.uc.edu:8180/luna/servlet/univcincin~42~42>

Crowe, Sean and Carolyn Hansen (2014). DC_to_VRA. In GitHub. Retrieved from https://github.com/crowesn/DC_to_VRA

Acknowledgements

Linda Newman, Head, University of Cincinnati Libraries Digital Content & Repositories Dept

Elna Saxton, Head, University of Cincinnati Libraries Content Services Division

Leslie Schick, Associate Dean of Library Services, Director Health Sciences Library