

□

Marrying local metadata needs with accepted standards:

The creation of a data dictionary at
UIC Library

Kristin E. Martin and Peter Hepburn

Nov. 16, 2010

Introduction

Who we are:

- Kristin Martin, Metadata Librarian, Catalog Department
- Peter Hepburn, Digitization Librarian, Digital Programs Department

Many of the images in this presentation come from the CITY 2000 collection.

“On a quest to create a lasting record of life in Chicago In the Year 2000 more than 200 photographers spent 366 days canvassing the city and chronicling its people, places and personality. The project's eclectic mix of styles and approaches blended into an historical document that will inform those in the future long after it teaches us about ourselves. “

Outline

- Working with metadata
- History of metadata guidelines at UIC Library
- Considerations in approaching the data dictionary concept

Outline

Given my longer history at UIC Library and yet relative newness to the world of metadata, I will speak largely on background and workflow matters. I will also draw in how I, as a non-cataloger, have had to learn about and work with metadata.

This will take me into the last point mentioned on this slide. Kristin will take over from that point.

Outline

- **Writing the data dictionary**
- **Examples from the data dictionary**
- **Issues**
- **Conclusions/questions**

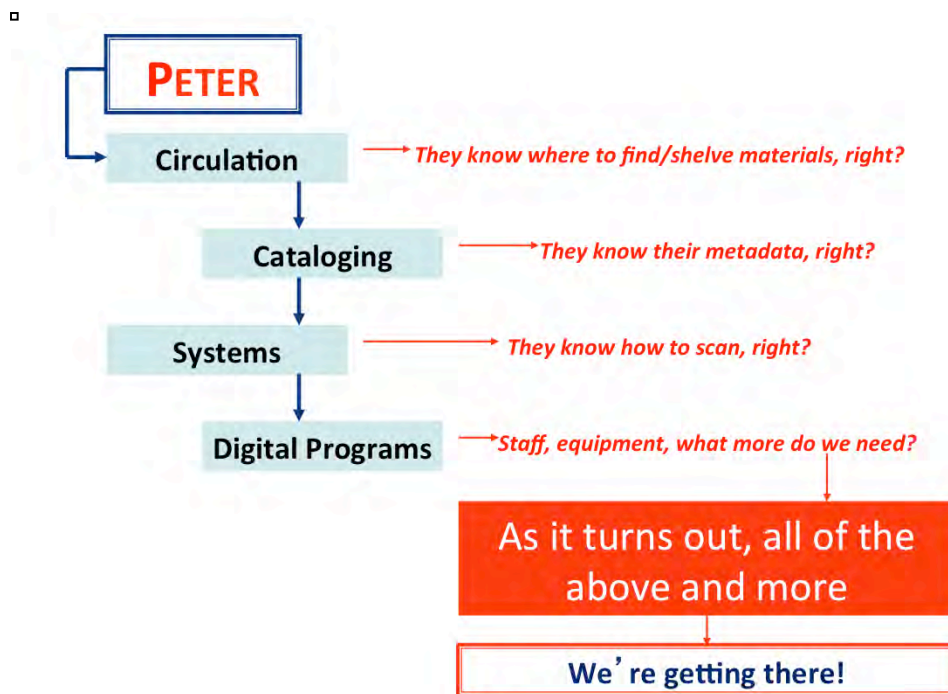
Overview (continued)

The process of writing the dictionary

Examples from the data dictionary

Issues

Conclusions (questions left until after all of us have presented?)



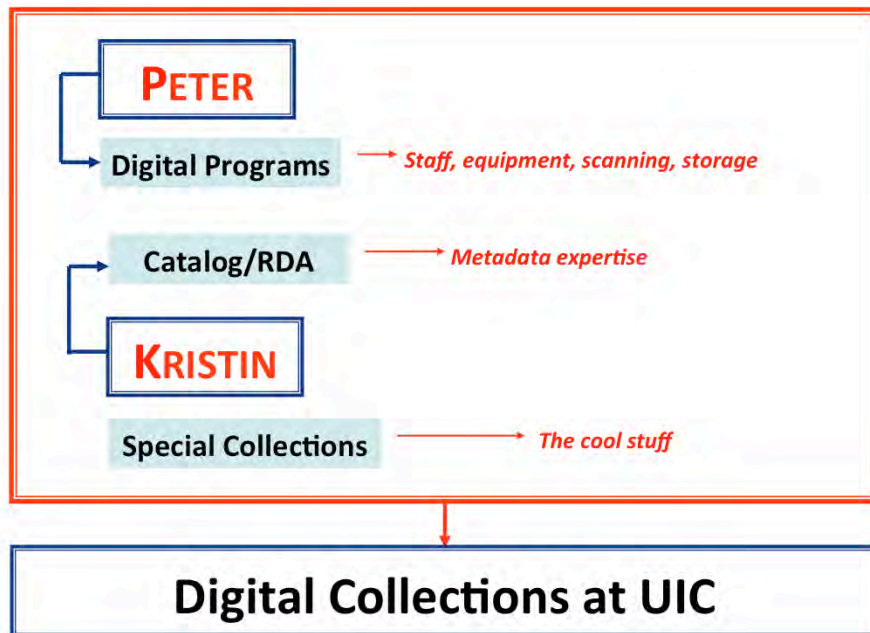
My past and current role in the organization and in digitization

- Circulation librarian
- Moved to Cataloging (working in digitization for a year given an extraordinary pool of money)
- Then moved to Systems (permanent position in digitization)
- Now head of my own department, Digital Programs – still reporting to AUL for Information Technology

Over the course of this move, I have been hands on with a number of things, but most of all with writing descriptive metadata.

All of this points to a couple of key things that influenced the development of the data dictionary:

- Workflow issues – who does what at what point?
- Ownership – who determines the metadata – the department doing the digital conversion and presentation, or the department responsible for the source material – we are moving closer to the latter, but as with many



Current set-up for digitization projects

The arrival of Kristin Martin has been a great catalyst for changing the digitization work dynamic at the Library.

In the past few years, we've been working on building a collaborative, not competitive relationship among the stakeholders.

- Establishing – and DOCUMENTING - workflows
- Involvement of multiple departments
- Shared information - use of a wiki
- Shared planning and development of an agreed-upon set of priorities



History of metadata at UIC Library

- UIC-Core
- No metadata specialist
- Involvement of multiple departments (but not metadata specialists)

Background on metadata work at the Library

UIC-Core pre-dates my time in digitization but not my time at the Library. It didn't gain traction.

When I started in digitization, it was a learn-as-I-go-along process, with no metadata specialist in Cataloging to turn to. Thrown into deep end: Dublin Core? Sure I've heard of it, but...

Multiple departments got involved in discussion of metadata: Systems, Special Collections, Cataloging, Photo Services. No department heads involved, no formal sanction, no power to implement outside of the group who informally met.

History of metadata at UIC Library

- **Development of guidelines**
- **Special Collections centric**
- **Not all metadata understood**

Background on metadata work at the Library (continued)

Early development of guidelines before Kristin's arrival at UIC

Special Collections was the gatekeeper. They would sometimes make changes to the metadata guidelines, but the guidelines were on a Blackboard (course management system) site with access only for that department.

Competing perspectives, and competing wrong perspectives. A group of us from stakeholder departments had some "lively" discussions about how to approach metadata. Some of our assumptions about how Dublin Core worked and about how ContentDM worked were very wrong. Also external considerations coming from CARLI, our consortium. As an example, Special Collections preferred not to assign titles to unnamed images, using the record number instead. CARLI does not permit numbers of titles.

Variable metadata needs

- Per collection and even within collections: geography, time period, source, and more
- Legacy collections
 - Example: CITY2000
- Dirty metadata - cleaning, re-writing needed before mounting

Start with understanding of primary stakeholders within Library

- Cataloging
- Special Collections
- University Archives
- Visual Resources Collections (inherited from Architecture and Art)

And others, minimally

Variable metadata needs

Not uniform by collection, nor even necessarily within collection given that some collections can and do have a mixture of types of materials

Issue of legacy collections – earlier efforts that pre-date the current data dictionary (or ANY dictionary in some cases) and don't adhere to the guidelines as well. Some of these we've cleaned up, but retrospective work is time consuming. Where we had to extract from a different presentation system, we've had headaches with mapping fields and funny characters added in. CITY 2000 is an example.

Workflow – who writes metadata?

- Who should?
- Fit with processing collections and digitization work
- Quality assurance?
- Example: CITY2000:
 - Special Collections responsible for collection but not metadata

Issue of ownership returns in considering the workflow for writing metadata

We seem to be moving toward a system where the department responsible for the original materials is also responsible for descriptive metadata. Digital Programs, by contrast, can share some of the metadata duties by filling in fields that can copy exactly from item to item, like the rights statement.

This now fits better with some of the processing work that feeds into digitization projects a little more directly.

Departments consult with Kristin Martin.

Quality assurance by both the department and Kristin

How this works for CITY 2000 – not in place for the legacy collection. Not the only example.

Comparison to newer IDOT images wherein when we tackle the collection, we have mapped out very carefully with Special Collections what they'll be responsible for in terms of description.

Considerations

Metadata must

- meet standards as interpreted by Catalog Department
- be grasped by librarians writing it
- *above all, serve user needs for accessing materials*

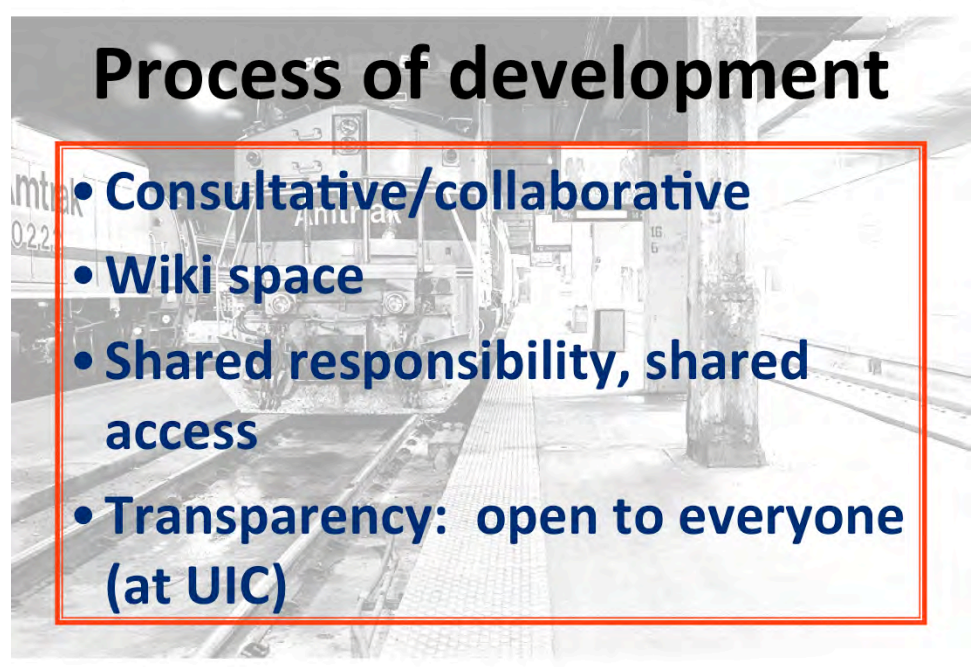
In the development of the dictionary and workflows, there have been additional considerations

Meeting standards of Catalog Department. UIC-Core may no longer be part of the conversation, but we do not proceed without collaboration with Cataloging to ensure metadata meets best standard possible.

Metadata must be grasped by the librarians who write it. We need to understand what a field is, its properties, before we can populate it. The guide has been IMMENSELY helpful with this – the documentation is extensive but clear and useful.

Metadata must serve users. I think we may have got away from that into what the librarians needed in order to do their work. No. We now focus on how the users might be able to find images and other content effectively. This is THE guiding principle above all others.

□



The dictionary was a collaborative effort again involving multiple departments: Cataloging, Special Collections, Digital Programs.

Use of wiki space for group editing. Kristin did a lot of the heavy lifting on this.

Everyone who works with metadata can edit – for example, add in special case examples when they pop up in new collections (like numbering/naming conventions). Edits are not frequent, but the tool enables us to easily accommodate unanticipated situations with new projects.

Transparent – the wiki is open to everyone at UIC. That means our colleagues can read, comment, edit. In practice, this hasn't proved to be the case often, but it is somehow reassuring.

Standards to Consider

- **Prior UIC work**
 - UIC Core
 - Digital Image Guidelines
- **CARLI**
- **Required metadata fields**

□

Standards to Consider

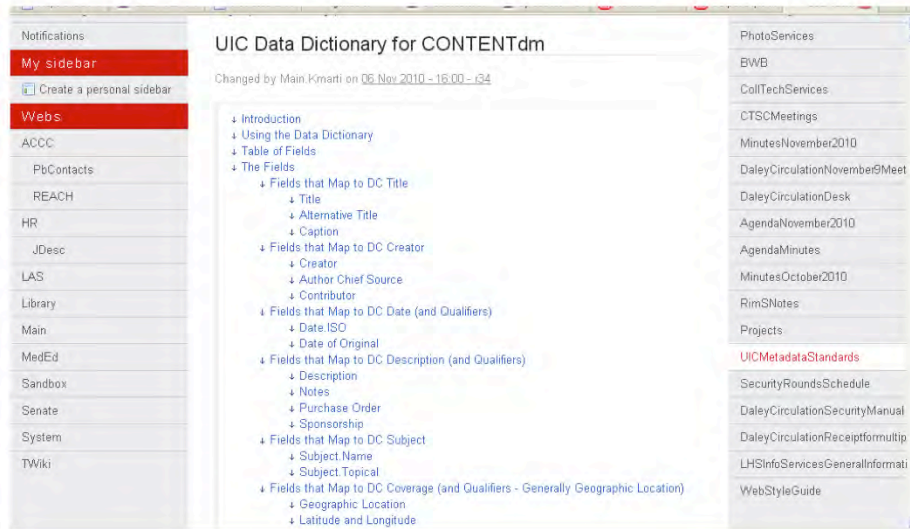
- **Best practices for descriptive metadata**
- **Other libraries**
 - UNC, Penn State, Collaborative Digitization Program, etc.
- **New: best practices for shareable metadata from OCLC**

□

Ongoing maintenance

- **Modifications as needed**
- **Project by project**
- **Development of Application Profiles for specific collections**
 - **Geographic collections**

<https://www.uic.edu/wiki/view/Library/UICMetadataStandards>



FosWiki wiki: The wiki is great for fast updating and versioning control, and now is a WYSIWYG editor. Library uptake of wiki is growing fast. Some caveats: Using multiple wiki software can create to confusion (different software from CARLI wiki) Not to mention timeouts with Bluestem. Also, FosWiki leaves a lot of blank space on the side of the screen, as I discovered when redoing the screenshots!

□

Table of All Fields

Dublin Core Field	CONTENTdm Field	CARLI Recommendation	UIC Recommendation
Title	Title	Required	Required
	Alternative Title	N/A	Optional
	Caption	N/A	Optional
Creator	Creator	Recommended	Required if applicable
	Author/Chief Source	N/A	Optional
	Contributor	Optional	Optional
Contributor	None	Optional	Maps to Creator
Date	Date ISO	Recommended	Required
	Date of Original	N/A	Required
Description	Description	Recommended	Optional
	Notes	N/A	Optional
	Purchase Order	N/A	Optional
	Sponsorship	N/A	Optional
Subject	Subject Name	Recommended (any Subject field)	Optional
	Subject Topical	Recommended (any Subject field)	Optional

Easy table format allows users to click straight to instructions for field

□

Creator
LABEL: Creator [or more specific label]
DUBLIN CORE: Creator
UIC --Required if applicable --Searchable --Viewable --Repeatable content --Controlled vocabulary
DESCRIPTION A individual, family, or organization primarily responsible for creating the intellectual content of the source object (for resources that are digitized) or the resource itself (for born digital objects).
INPUT GUIDELINES 1. Enter names in authoritative form, following rules from AACR2. <ul style="list-style-type: none">▪ Check the Library of Congress Name Authority File for already established names▪ If no authorized form exists, create one based on rules from AACR2.▪ Personal names should be entered, Last, First, with qualifications and/or dates following as appropriate.▪ Other authorized lists, such as ULAN (Union List of Artist Names) can be used as well (ULAN is now available as a preloaded controlled vocabulary in CONTENTdm).▪ For organizations, if names are in a hierarchy, list the parts from largest to smallest, separated by periods, following AACR2 for which portions of the hierarchy need to be included.
2. If a more specific label is appropriate, select a label that is appropriate for role of the creator

Information indicates:

- label name
- mapping to DC
- required/optional
- searchable/not searchable
- viewable/hidden (why would you want to hide? Example coming up with date fields)
- repeatable content (CDM does not allow the same field to have the same label, but you can repeat with different names)
- use of controlled vocabulary
- description
- instructions
- examples

□

2. If a more specific label is appropriate, select a label that is appropriate for role of the creator (e.g., Photographer, Author, Architecture Firm). More than one Creator field may be used if there are multiple creators who fill different roles in creating the intellectual content of the object. Multiple names can be input in separate fields with different labels, or in the same field, separated by semi-colons.

3. Creator can be "Anonymous" if that information is indicated on the object. If a creator is truly unknown, leave the field blank.

EXAMPLES

- Graw, Renata
- Lew, Michael S., 1965-
- Smith, John C. (John Charles), 1948-
- C. F. Myrick, Jr. Studio
- International Business Machines Corporation
- Illinois. Dept. of Adult, Vocational, and Technical Education. Professional Development and Curriculum Section

Author Chief Source

LABEL: Author Chief Source

DUBLIN CORE: Creator

UIC

- Optional
- Optionally Searchable
- Viewable or Hidden

DESCRIPTION

An exact transcription of creator name as it appears on the resource. This is the equivalent of

Creator example continued

□

Date ISO

LABEL: ISO Date

DUBLIN CORE: Date Created

UI:
--Required
--Searchable
--Viewable or HIDDEN

DESCRIPTION
The date of the creation of the original work, not the digital reproduction or the reprographic work (e.g., the date of a painting, not the date of the photograph of the painting or the scan of the photograph of the painting, unless the context of when the photograph of the object is what is relevant to the collection). This field must be used in conjunction with the [Date of Original](#) field. The ISO Date provides a standardized form for searching; the Date of Original provides a human-friendly version of the same information. Originally this date was hidden from public view, as it was only used for searching; however with the upgrade of CONTENTdm to 5.1, having the date be viewable means that it can show up in the faceted results after users perform a search. It also allows users to click on the date to bring up other objects with the same date. You should weigh whether displaying the date will provide better access or more confusion for users.

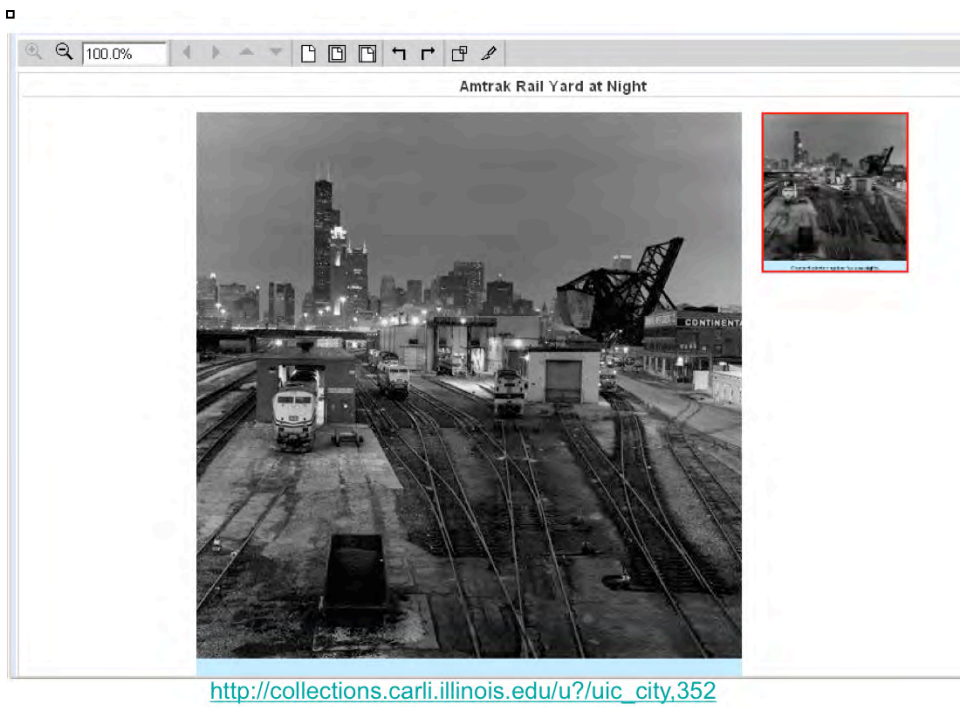
INPUT GUIDELINES
1. CONTENTdm has special guidelines about how to input dates, which vary, depending on whether you are going to be importing the metadata in a tab-delimited file, or entering it directly in the Acquisitions Station or Web Interface. Once the information is input, CONTENTdm will automatically flip the date to follow [ISO 8601](#) (e.g., yyyy-mm-dd). Note that ISO 8601 accepts either yyyymmdd or yyyy-mm-dd, but CONTENTdm requires the hyphens. More information from the [CONTENTdm help files](#):

- If using a tab-delimited file, the only acceptable date format is mm/dd/yyyy.
- When using the Media Editor, the Project Spreadsheet, or the Template Creator, acceptable formats for date entry are:
 - yyyy
 - yyyy-mm
 - yyyy-mm-dd

Date has both a viewable and hidden version.

Issues/New Directions

- **Contending with new materials**
- **Retrospective work**
- **Taking advantage of ContentDM 5 upgrade**
 - **Shareable vocabularies**
 - **Faceted browse options**
 - **Sharing metadata via WorldCat**



Example from CITY2000

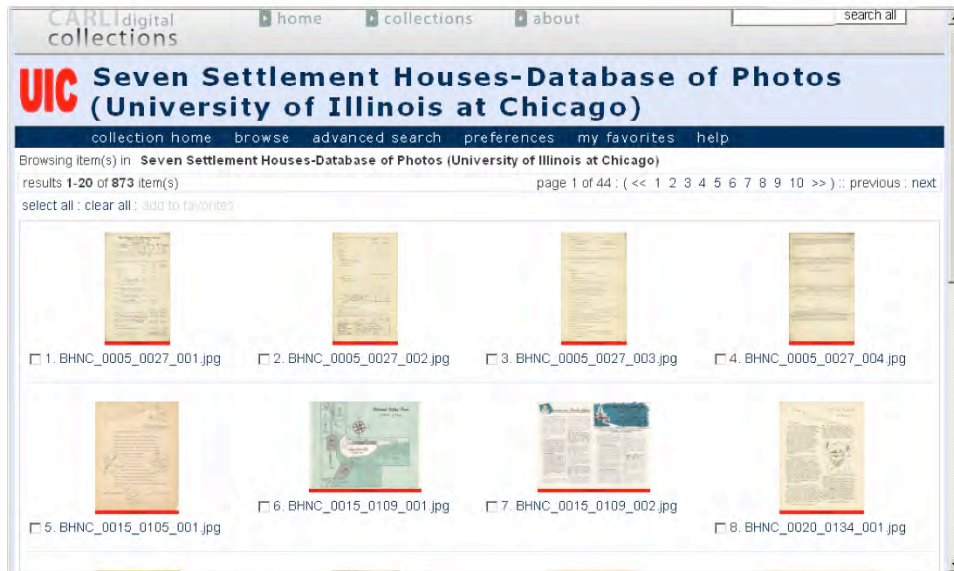
Title	Amtrak Rail Yard at Night
Caption	The Amtrak rail yard at the Chicago River's South Branch on New Year's Day
Photographer	Gordon, Ron
Date	January 1, 2000
Date.ISO	2000-01-01
Geographic Location	University Village, Chicago, IL
Project Number	104
Project Title	Destroyed Building Series
Subject	Amtrak Railroad yards New Year Night
Type	Image
Form	Photographs
Medium	Black and white, 120
Host	University of Illinois at Chicago. Library.
Collection	Cornet Archive of Chicago in the Year 2000 (University of Illinois at Chicago)
Repository	University of Illinois at Chicago. Library. Special Collections and University Archives.
Citation	Cite as [Creator]. [Title]. [Identifier]. Cornet Archive of Chicago in the Year 2000 (University of Illinois at Chicago).
Rights	This image may be used freely, with attribution, for research, study and educational purposes. For permission to publish, distribute, or use this image for any other purpose, please contact Special Collections and University Archives, University of Illinois at Chicago Library, 801 South Morgan St., Chicago, IL 60607. Phone: (312) 996-2742, email: lib-permissions@uic.edu.

Metadat for example. Discuss use of specific labels: advantages within CDM and controlled vocabulary for clicking, but disadvantage with sharing (all maps to description, meaning can get lost); note date



[Carberry Collection](#)

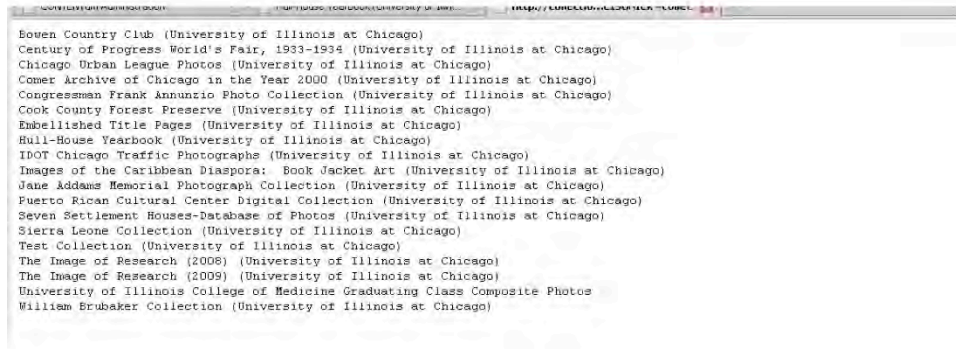
Date browsing from Carberry. Why making fields hidden now has a cost.



Seven Settlement Houses

Clean-up needed to have metadata conform to CARLI guidelines. Note that titles are identifiers, a no-no.

Shared vocabulary for all collections



Have a controlled vocabulary for collection name allows the user to easily retrieve all items in a specific collection (and gives them the nifty facets on the left)

□

Conclusions

Questions?



Closing comments

- Mistakes are not the end of the world
- Some findability is better than none
- Still, better to iron out issues beforehand rather than partway through
- Good to ensure that any plan that does come into place is flexible enough for exceptions or special circumstances
- Document, document, document