

More Bytes, Less Bite:

Cutting Corners in Digitization

Joshua Ranger, UW Oshkosh

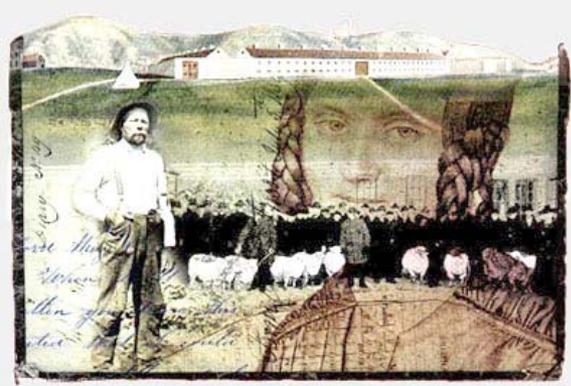


University of Wisconsin Digital Collections



Promoting the Wisconsin Idea by providing professional leadership in the creation of quality digital resources from libraries and archives for faculty, staff and students, citizens of the state and scholars at large.

*I intended to have written you
but have been so situated*



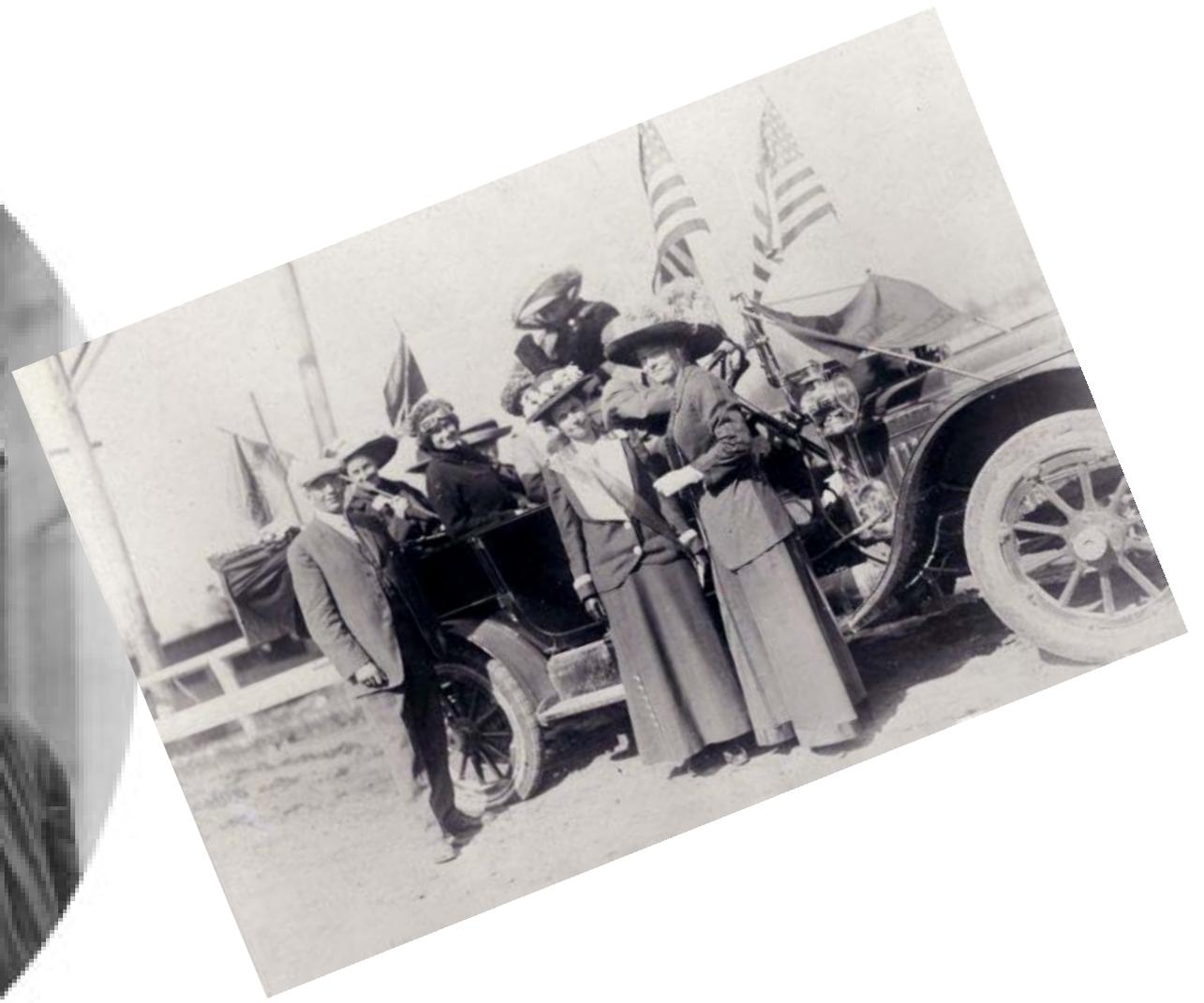
Wisconsin's Pioneer Experience

2, 245 pages in 23 collections - 24.8 minutes per page (includes over 150 hours of transcription work). Amounts to \$7.32 per page



Wisconsin Goes to War

2492 pages in 28 collections - 8.87 minutes per page or \$1.60 per page.



Ada Lois James, 1876 - 1952

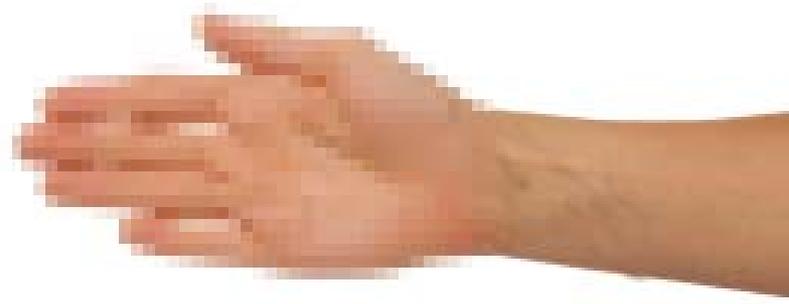


The Experiment: The Pinto

- Scan from Photocopies
- No item-level metadata

Boxes 18-20 Ada James
Incoming Correspondence





The Control: The Cadillac

- Full color scans
- Item (document-level) metadata



Folders 3-4 of Box 17 of Ada
James Incoming Correspondence



Reformatting

Includes: Scanning, image rotation/correction, creation of derivatives, uploading files

Experimental Approach:

- Made photocopies of originals, passing these through a high speed scanner.

Experimental



.86 minutes (51.6
sec.) per page

Control



5 minutes per page

Metadata

Includes: Creation of metadata, some quality control and corrections



Experimental Approach:

- Moved “issue level” from item (letter) to folder.
Abandons item level cataloging, making the folder the lowest level described, in par with existing (and LINKED) EAD finding aid.
- Folder is navigated via imposed “page numbers” similar to several Library of Congress approaches.

Experimental



0.6 minutes (36 sec)
per page

Control



3.12 minutes per
page

Control Model



The State: Ada James papers, correspondence, 1912, Nov. 8-Dec. 23: Contents - Mozilla Firefox

File Edit View History Bookmarks Tools Help del.icio.us

http://digital.library.wisc.edu/cgi-bin/WI/WI-idx?type=header&id=WI.JamesB17F3

http://www.rootswe... Getting Started Latest Headlines Midwest Archives Co... The Wisconsin Histori... National Weather Ser... FileMaker Instant We... menu Rosanna url

Search
Browse
Copyright
Help
Reprints
UWDC

Description

Ada James papers, correspondence, 1912, Nov. 8-Dec. 23

Source:

James, Ada Lois, 1876-1952
Ada James papers, correspondence, 1912, Nov. 8-Dec. 23
Wis Mss OP, Box 17, Folder 3
[unpublished]

URL to cite for this work: <http://digital.library.wisc.edu/1711.d1/WI.JamesB17F3>

Search for word or phrase within this work: **Search**

Contents

[\[Letter from Olympia Brown to Miss A. L. James: Racine, Wis. Nov. 8, 1912\]](#)
[\[Copy: letter from Olympia Brown to Miss A. L. James: Racine, Wis. Nov. 8, 1912; Copy: letter from H. M. Holton to Ada L. James and Crystal Eastman Benedict: Milwaukee, Nov. 12, 1912\]](#)
[\[Letter from Mrs. W. G. Adams: Merrill, Wis., Nov. 8, 1912\]](#)
[\[Letter from Katharine M. Steele to Mrs. Benedict: Whitewater, Wis., Nov. 8, 1912\]](#)
[\[Letter from Elizabeth Bass to Miss James: Chicago, Nov. 8th, 1912\]](#)
[\[Letter from Catharine Waugh McCulloch to Miss Ada James: Chicago, Ill., Nov. 8, 1912\]](#)
[\[Letter from Mary G. Hoyt to Miss James: Madison, Wisconsin, November 9, 1912\]](#)
[\[Letter from Mrs. J. W. Laughlin to Miss James: Janesville, Wisconsin, Nov. 9, 1912\]](#)
[\[Letter from Mrs. G. B. Twitchell to Miss James: Cincinnati, O., Nov. 9, 1912\]](#)
[\[Letter from Elizabeth Low Watson to Miss Ada L. James: Saratoga Cal, Nov. 10th, 1912\]](#)
[\[Letter from Belle C. La Follette to Miss James: Maple Bluff Farm, Madison, Wisconsin, November 10, 1912\]](#)
[\[Letter from Katherine Runkel to Miss A. L. James: Prairie du Chien, Wis., Nov. 11, 1912\]](#)
[\[Letter from Miss Effie E. Alt to Miss Ada James: Neillsville, Wis., Nov. 11, 1912\]](#)
[\[Letter from Anna E. Blount to Miss Ada James: Chicago, November 11, 1912\]](#)
[\[Letter to Miss Shaw: Nov. 11, 1912\]](#)

“Issue” level metadata describes folder

“Item” level metadata describes letter

Experimental Model



The State: Ada James papers, correspondence, 1913, Jan.: Contents - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://digioll.library.wisc.edu/cgi-bin/WI/WI-idx?id=W1.JamesB18F1

The State: Ada James papers, corre... Digital Resources

The State of Wisconsin Collection UW DC

◀ Volume ▶ Go to page

Description

Ada James papers, correspondence, 1913, Jan.

Source:

James, Ada Lois, 1876-1952
Ada James papers, correspondence, 1913, Jan.
Wis Mss OP, Box 18, Folder 1
[unpublished]

URL to cite for this work: <http://digital.library.wisc.edu/1711.dl/W1.JamesB18F1>

Search for word or phrase within this work: Search

Contents

[Ada James papers, correspondence, 1913, Jan., pp. \[1\]-\[196\]](#)

▲ [Top of Page](#)

◀ Volume ▶ Go to page

This material may be protected by copyright law (Title 17, US Code).| For information on re-use see:
<http://digital.library.wisc.edu/1711.dl/Copyright>

[Technical Assistance](#) | [Content/Navigation Questions](#) | [University of Wisconsin Digital Collections](#)

“Item” level metadata no longer distinguishes each document.



Administration

Includes Inventorying materials, setting up metadata for student entry, final quality control (click through), portal design and HTML programming planning/meetings.

Experimental Approach:

- Abandon subject approach to building collections
- Use of EAD finding aid as primary access point
- Minimal metadata makes quality control process easier

Experimental



Control



0.34 minutes (20.4 sec.)
per page

(17.4 sec/page without
HTML)

0.56 minutes (33.6
sec.) per page

Bottom



Line

Experimental



Control



1.8 minutes
per page

Approximately
\$0.43 per page

8.68 minutes
per page

Approximately
\$1.53 per page



User Satisfaction



Assessment Methodology

- Study of seven undergraduate history majors, seven library science graduate students.
- Participants completed six tasks using both Control and Experimental Model sections of the Ada James collection.
- Participants were then interviewed about their preferences and expectations.

General Observations

- Participants did not perform well with tasks related to experimental section, many failed or gave up.
- Desire more metadata, not less, about individual letters. 11 of out of 14 students suggested some sort of calendaring or subject cataloging.
- Have high expectations of how primary sources should behave on-line. Wish to conduct “Google-like” searching or, at the very least, to have the resource behave like a journal database.

General Observations

- Students will assume that everything has been digitized. EAD Finding Aid key in indicating how much a collection has been digitized.
- Motivation is very important in the potential for a student to use the experimental model in future research.
- Interviews suggest a big difference in how long students will spend researching archival documents in person versus the amount of time they'd spend with the same material on-line.

Ease of Use

Asked about ease of use of the two models.
Scale - 1 (very difficult) to 5 (easy)

Undergraduates		
	CONTROL	3.5
	Experimental Model	2.7
Graduates		
	CONTROL	4.28
	Experimental Model	2

Browsing

In order to effectively BROWSE papers, students desire more metadata, not less, about individual letters. Even students who report that browsing is their favorite way to use digital collections do NOT like the browsing of the experimental model.

Some comments:

“Waste of time”

“You’d lose me at ten pages”

“Painful”

“I expect to find it as detailed as if I went to the archives. I expect all that descriptive information in detail”

Searching

Most students report a desire to search over browsing. Wish to conduct “Google-like” searching, with simple understood search results. Would prefer full text searching but would accept searching across abstract metadata. Few students seem to consider the mechanics of the search mechanism.

Some comments:

“Searching was frustrating”

“I assumed everything was searchable”

“Put up a disclaimer”

Navigation and Searching

Asked about to rate the navigation and searching. Scale - 1 (very difficult) to 5 (easy)

Undergraduates		
	CONTROL	3.4
	Experimental Model	2.7
Graduates		
	CONTROL	4.14
	Experimental Model	1.7

The Money Question

When the comparative costs were explained (four/five times more stuff for same amount of money), **ALL** respondents stated that the Experimental Model model **WAS TO SOME DEGREE** acceptable.

Some comments:

“Better than not having it at all.”

“Better than driving an hour away”

“This (Experimental Model approach) may turn people off from using primary sources.”

Likelihood of Use Again

When asked about how **likely** they would consider using the source for future research.
Scale - 1 (least likely) to 5 (most likely)

Undergraduates		
	CONTROL	4.28
	Experimental Model	3.57
Graduates		
	CONTROL	4.57
	Experimental Model	1.86

Desire for Mediation

In the future would you:

- a) Drive to the nearest archives and use the collection on-site
- b) Use the collection online if its provided online in its entirety and try to figure it out on your own
- c) Browse the collection on -line and use chat, email or phone assistance if such help was provided as part of the collection.

Undergrads		
Drive	Figure it out	Use help
2	4	1
Graduates		
Drive	Figure it out	Use help
2	1	5

Conclusions: Where to go from here?



- Realize what we have lost
- Improve Browsing
- Improve Searching?
- Accept a variable approach?

Conclusions: What we have lost

We have lost the ability to assess certain qualities of a folder of material:

- Amount of material, pure bulk plus a sense of how many documents are in the folder.
- Types of documents in a folder (reports, correspondence, drafts of documents, contracts, etc.)
- Color (not needed), type of paper used

Conclusions: What we have lost

We have lost the ability to browse quickly:

- No browser will ever replace the ability to quickly review a folder in person.
- Most students suggest that this is a reasonable trade off for 24 hour, distant access.

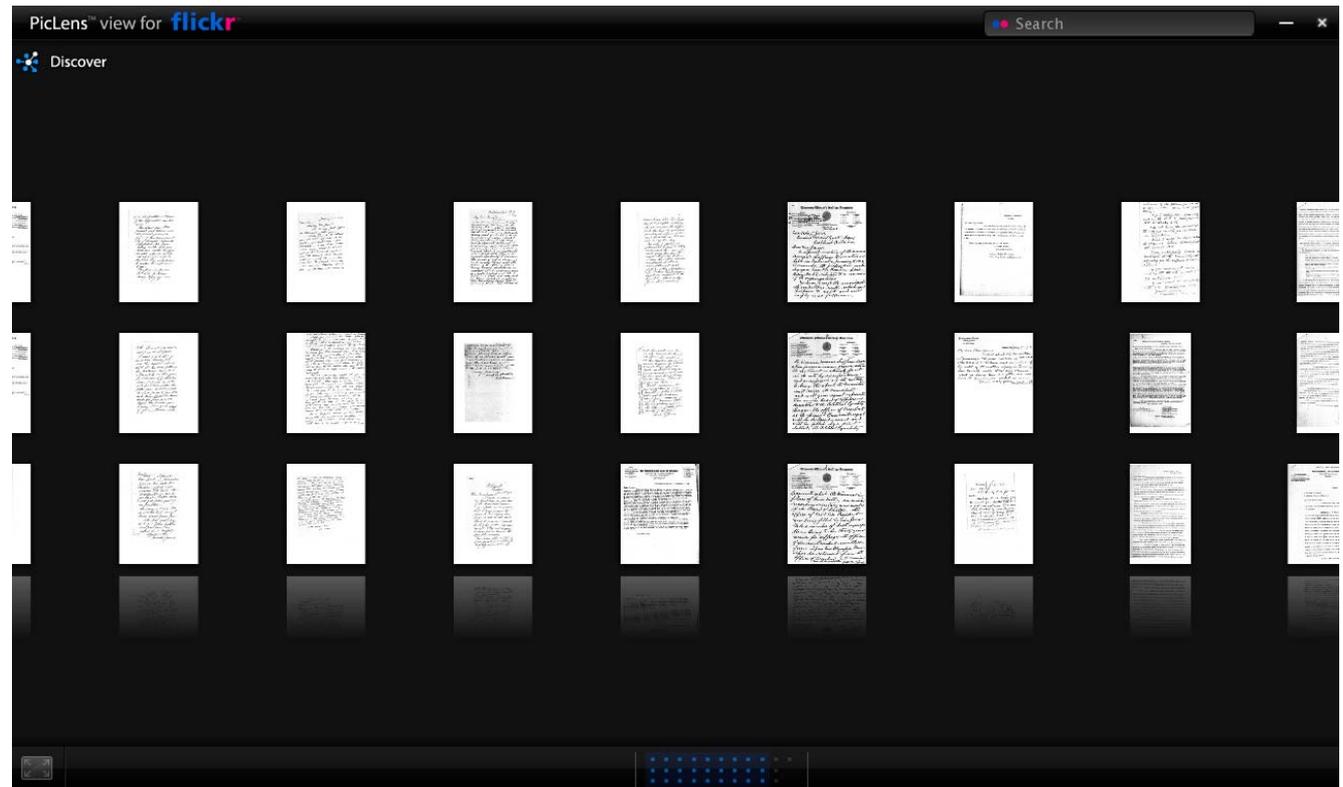
Conclusions: Improve Browsing

Would a graphical interface help students?

Can the *isolation* of individual documents be automated?

Conclusions: Improve Browsing

Graphical Interface

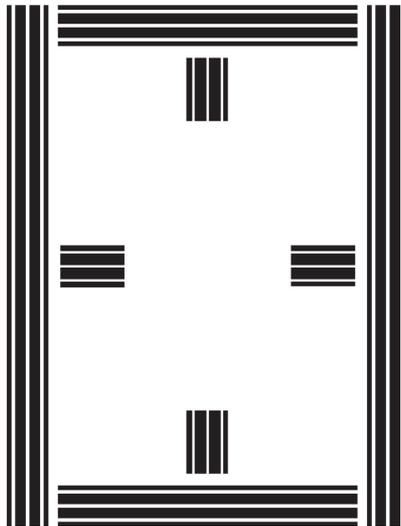


Conclusions: Improve Browsing

Isolation

Reported need for students to know how many documents are in a folder.

Use “patch codes” among photocopies to signal the software to isolate document.



Kodak Patch Code

Conclusions: Improve Searching

The presence of a search box makes almost everyone assume that all documents in the collection are searchable. Do we keep it?

Do the photocopies effect efficacy of OCR?

Are there different products that can more accurately work with older documents?

User supplied transcriptions?

Conclusion: Variable approach

Horses for Courses

Might this approach be best used in collections NOT organized chronologically?

Allow detailed finding aids provide metadata needed for discovery.

More Bytes, Less Bite:

Cutting Corners in Digitization

