

Google and the University of Michigan

Perry Willett
University of Michigan
November 13, 2006

University of Michigan and Google

- What is the project?
- Why is the Univ of Michigan participating?
- Why is the Univ of Michigan creating a new access system?
- How does it work?
- Why does this matter?
- Demo
- What's next?

Michigan and Google

- Digitizing entire library
 - 7 million volumes
 - 6 years (from Dec 2004)
- Productivity
 - recent week, 30,000 volumes
 - exponential growth—doubling
- University of Michigan gets a copy of everything

About the files

- Returned to UM: package per volume, id'd by barcode
 - 600dpi TIFF ITU G4 (bitonal) for print
 - 300dpi JPEG2000 color/grayscale for illus.
 - File naming conventions corresponding to UM specs
 - Checksums
 - Production notes
 - OCR (UTF-8 text files)
 - Page metadata: page numbers, page features, chapter start
- Quality control
 - Ongoing improvement of hardware/engineering
 - Image quality good and improving
- What is secret and why?
 - Technology
 - Numbers

Technology



Why is Univ of Michigan creating a new access system?

- To serve our primary audiences: researchers, students
- Light archives are more reliable than dark archives
- To learn what's required to put a collection of this size online
- For Google



Architecture overview

Four basic pieces:

- Page image and metadata repository
- Mirlyn, the library catalog
- Rights and GeolP databases
- Pageturner

Repository

- Automatic validation:
 - Check barcode check digit using Luhn algorithm
 - Fixity check on JPG, TIFF, UTF8 using MD5
 - Well-formedness and embedded metadata check on JPG, TIFF, UTF8 using JHove
 - Various completeness cross-checks
 - Failures retried, will eventually refer to Google
- Periodic fixity checks using MD5

Repository storage

- Current: Local, single-server, multi-terabyte, proof-of-concept.
- Short-term: Local, mirrored, sustainable, scalable storage system designed for high storage.
- Longer-term: multi-institutional repository, mirrored at multiple, geographically remote institutions

Mirlyn

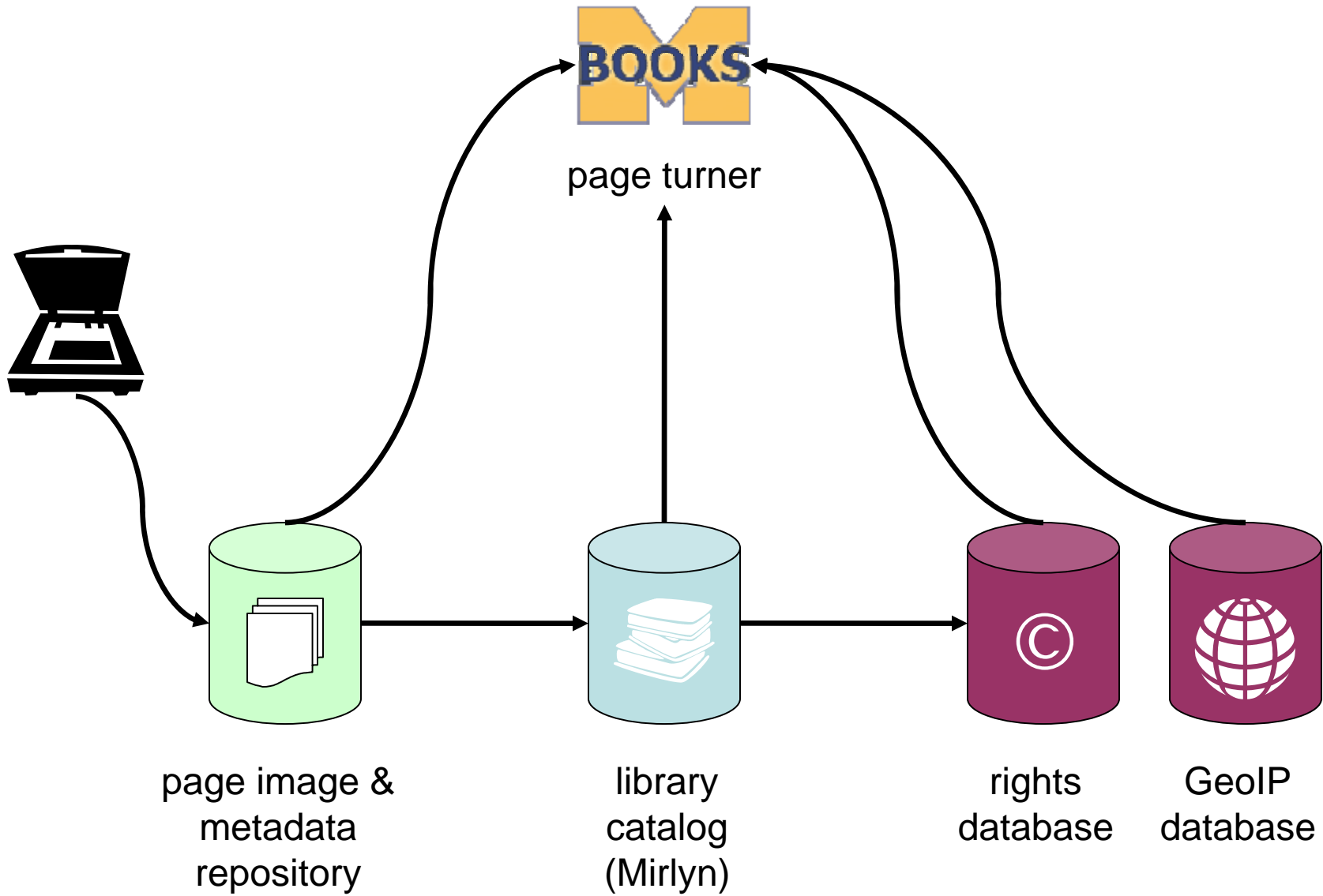
- ExLibris Aleph, includes X-Server
- Store information in item record
- Automatic link generation, not using 856 field
- Need to have accurate information in fixed fields, holdings records for rights determination

Rights Database

- What information to store?
 - Considered complexity and maintenance
 - Considered using MARC directly
 - Needed to accommodate both bib record-derived rights and manual overrides
- Approach: examine bib record, determine authoritative copyright status, store rights attribute, reason, and timestamp
- Stored in MySQL

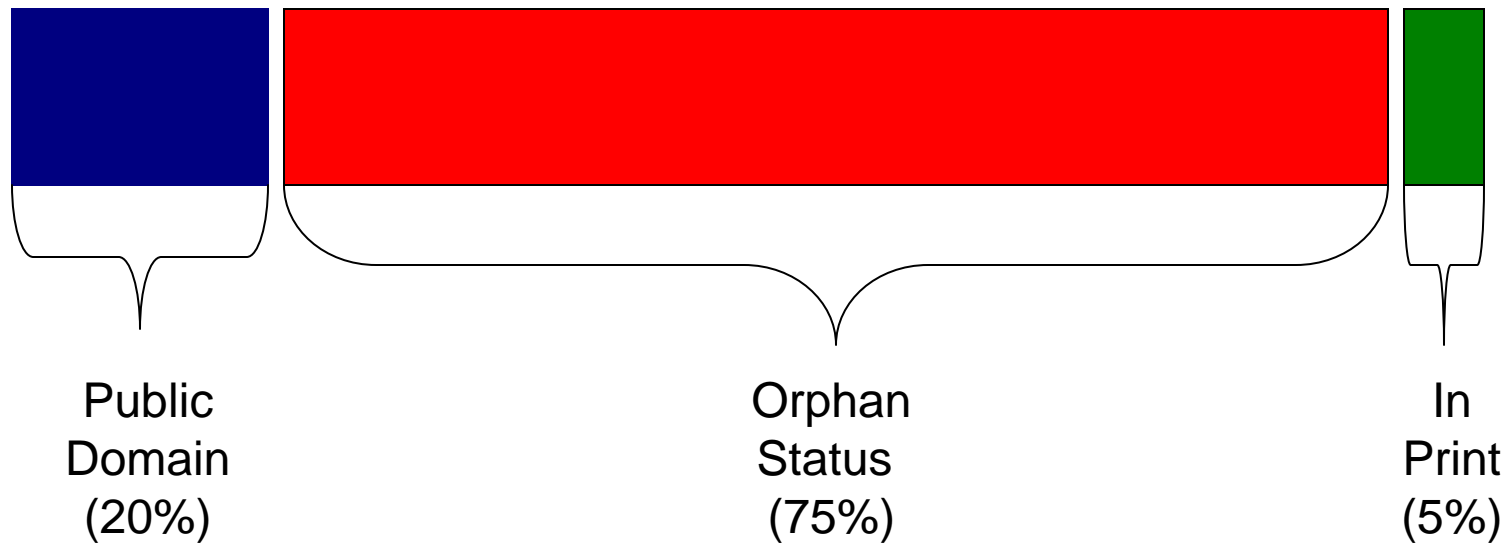
GeoIP Database

- Last minute request from University General Counsel: distinguish US and non-US users
- Several options; all use DNS registry information with tweaks for additional accuracy.
- We chose MaxMind GeoIP database:
 - 99% accurate
 - Update utility and several query APIs included
 - \$12/month



Why this matters, I

32 million unique titles in all US libraries



Source:

Tim O'Reilly <http://www.lib.umich.edu/mdp/symposium/presentations/oreilly.pdf>, slide 25

Why this matters, II

- Potential to change research, education
- For example:

Garratt, H.A.

The modern safety bicycle

London, New York: Whittaker & co., 1899.

<http://hdl.handle.net/2027/mdp.39015002947029>

(Held in 8 libraries according to WorldCat)

- Anecdotal evidence

Demo

- mirlyn.lib.umich.edu
- Search: The Journal of the Royal Geographical Society of London

Mirlyn - Ann Arbor - Holdings - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://dev.mirlyn.lib.umich.edu/F/DX7TIQGX9SDLVHGYBB3Q5X4CPRT6GU2N3KEXI3FRMDFMYLD6YH-26271?func=item-gl

Getting Started Latest Headlines CNN.com

FIND ARTICLES FIND DATABASES FIND OTHER LIBRARY CATALOGS

Library Catalog of the University of Michigan Ann Arbor

Mirlyn

Login
 My Account
 My Shelf
 My Saved Searches

Ask Us
 Help
 Course Reserves

Basic Search Advanced Search Command Language Browse Results List Previous Searches

Ann Arbor : HOLDINGS Back

Royal Geographical Society (Great Britain):
 The Journal of the Royal Geographical Society of London..
 London, : J. Murray..
 50 v. : ill., fold. maps. ; 22 cm..
[1831-1880 { Access restricted \(authentication may be required\): } http://www.jstor.org/journals/02666235.html](http://www.jstor.org/journals/02666235.html)

Owning Location	Buhr Shelving Facility - Ask at any library G 7 .R89
Public Note	Vols.1-39,41-50 in Buhr
Library has	1-50 1830/1831-1880
Owning Location	Electronic Resources See URL for access
Public Note	Electronic serial
Library has	1-50 1830/1831-1880

Select location Hide loaned items

Done






Start 2 Firefox Meeting... 2 Micr... Microsoft... Inbox - ... 1:38 PM

Mirlyn - Ann Arbor - Holdings - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://dev.mirlyn.lib.umich.edu/F/DX7TIQGX9SDLVHGVB3Q5X4CPR6GU2N3KEXI3FRMDFMYLD6YH-26271?func=item-gl

Getting Started Latest Headlines CNN.com

Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.8	U-M Online  Full Text Google Online	On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.9 1839		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.10 1841		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.11-12		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.13-14 1843		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.15-16 1845-1846		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.17 1847	U-M Online  Full Text Google Online	On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.19 1849		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.21 1851	U-M Online  Full Text Google Online	On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.22 1852		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.23 1853	U-M Online  Full Text Google Online	On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.24 1854		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.25 1855		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.26 1856	U-M Online  Full Text Google Online	On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.27 1857		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.28 1858		On Shelf
Expand	Buhr Shelving Facility	Ask at any library	G 7 .R89	v.29 1859		On Shelf

Done

Start | 2 Fi... | Meeti... | 2 M... | Micro... | Inbo... | navs... | EN | 1:39 PM

journal of the royal geographical society of london - Google Book Search - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://books.google.com/books?q=journal+of+the+royal+geographical+society+of+london&btnG=Search+Books&as_l

Getting Started Latest Headlines CNN.com

Sign in

Google **Book Search** BETA

Web Images Video ^{New!} News Maps more »


journal of the royal geographical society of lo Search Books


Advanced Book Search
Google Book Search Help


Search: All books Full view books


The following words are very common and were not included in your search: of the of. [\[details\]](#)

Book Search Books 1 - 10 with 40100 pages on journal of the royal geographical society of london. (0.00 seconds)

 [The Journal of the Royal Geographical Society of London - Page v](#)
by Royal Geographical Society (Great Britain), JSTOR (Organization)
THE ROYAL SOCIETY. Geographical work of the Royal Society 7 Captain Cook, Phipps, Bruce 18 CHAPTER V.—FOUNDATION OF THE ROYAL GEOGRAPHICAL SOCIETY. ...
Snippet view - [About this book](#)

 [The Journal of the Royal Geographical Society of London - Page 1](#)
by Royal Geographical Society (Great Britain), JSTOR (Organization)
Journal of a Tour to Karen-ni, for the purpose of opening a Trading- Road to the Shan Traders from ...
Snippet view - [About this book](#)

 [The Journal of the Royal Geographical Society of London - Page 40](#)
by Royal Geographical Society (Great Britain), JSTOR (Organization)
... Geographical Society ...
Snippet view - [About this book](#)

 [The Journal of the Royal Geographical Society ...](#)
by Norton Shaw, Hume Greenfield, Henry Walter Bates - 1831
212736 LONDON ... London ...
Full view - [Table of Contents](#) - [Index](#) - [About this book](#)

Find: panlexicon Find Next Find Previous Highlight all Match case Reached end of page, continued from top

Done

Start 3 Firef... Meeting... 2 Micr... Untitled ... navsa2... EN 9:22 AM

Journal of the Royal Geographical Society of London, vol. 8 (1838)

- <http://hdl.handle.net/2027/mdp.39015013755734>
- “our late most gracious Sovereign, William the Fourth”

our late most gracious Sovereign, William the Fourth" - Google Book Search - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://books.google.com/books?q=%22our+late+most+gracious+Sovereign%2C+William+the+Fourth%22&btnG=Sea

Getting Started Latest Headlines CNN.com

[Sign in](#)

Google [Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

Book Search **BETA** "our late most gracious Sovereign, William th" [Advanced Book Search](#)
[Google Book Search Help](#)

Search: All books Full view books

Book Search Books 1 - 1 with 1 pages on "our late most gracious Sovereign, William the Fourth". (0.00 seconds)

Did you mean: ["or later most gracious Sovereign, William the Fourth"](#)



[The Journal of the Royal Geographical Society ... - Page iii](#)
 by Norton Shaw, Hume Greenfield, Henry Walter Bates - 1831
 During the past year we have had to deplore the loss of **our late most gracious Sovereign, William the Fourth** ; a monarch whose name, associated as it is in ...
[Full view](#) - [Table of Contents](#) - [Index](#) - [About this book](#)

"our late most gracious Sovereign, William th"

[Google Book Search Help](#) | [Advanced Book Search](#)

[About Google Book Search](#) - [Information for Publishers](#) - [Google Home](#)

©2006 Google

Find: panlexicon Match case

Done

Start 3 Firef... Meeting... 2 Micr... Untitled ... navsa2... EN 9:25 AM

What's next?

- Wiki-like functionality for correcting OCR?
- Social tagging?
- Personal collections?
- Multi-institutional repository and union catalog?

More Information

- Documentation:
<<http://www.lib.umich.edu/mdp/>>
- FAQ :
<<http://mdp.lib.umich.edu/m/mdp/mdp-faq.htm>>
- Email: pwillett@umich.edu