Google Books & UMN

19th Annual Minitex ILL Conference

Wendy Pradt Lougee
May 2010

University of Minnesota
LIBRARIES

Driven to Discover®
AGENDA
• Google Book History
• Library agreements
• Hathi Trust
• Project launch @UMn
In the Beginning...

Google™

Google™ 5
Google Books

• Partnership program for publishers and libraries
  – 31 libraries world-wide
• Scanned pages (images), full text
• Provides search, display, delivery of books (depending on copyright status)
• Additional functionality, integration with other Google programs

http://books.google.com/
The View(s)

- Full View
  - Public Domain

- Publisher Partner Preview Program

- “Snippet” View
Key words and phrases
pediatricians, resident staff, pediatric, ward rounds, counties, basins, child visits, dentists, isolated

Academy’s program for the improvement of child health. The second and far more difficult step is to translate this study into terms of action. The Academy has already demonstrated its determination that this study should not end as a study, but that it should be, as it was originally intended to be.

INTRODUCTION

The medical profession is being challenged as never before to assure that
Library Agreements
A Motivational Study: Committee on Institutional Cooperation

CIC Institutions (more than football...)

- University of Chicago
- University of Illinois
- Indiana University
- University of Iowa
- University of Michigan
- Michigan State University
- University of Minnesota
- Northwestern University
- Ohio State University
- Pennsylvania State University
- Purdue University
- University of Wisconsin-Madison

Provost-governed
Academic programs
Shared infrastructure
80+ million volumes
Timeline for Google Agreements

Oct. 2006
CIC Executes agreement

June 2008
Settlement terms released
CIC Negotiating Team named

Feb. 2009

June 2007
CIC Executes agreement

Oct. 2006
Wisconsin announces agreement

Aug. 2006
UC System announces agreement

May 2006
CIC approaches Google about a partnership

Dec. 2004
G5 including Michigan announce agreements

AP /AG Lawsuits

Oct 2005
CIC Framework

• Get as many out-of-copyright works digitized as possible ("preserves" decaying volumes)
• Digitize as much in-copyright works as possible to extend user access
• Seek ways to assure eventual return of in-copyright files
The CIC Agreement

• Up to 10 million volumes from CIC (Michigan and Wisconsin); ~1 million from UMN
• Focus on unique titles in CIC
• Capture “collections of distinction” in their entirety
• Create Shared Digital Repository (SDR)
• Out-of-copyright returned to SDR
• In-copyright works held in escrow with specified release conditions
Library Opportunities

• Preservation and access
  – “Discovery happens elsewhere”
  – Global user community

• Collection management
  – Shared print retention and storage
  – “Cloud Library” potential

• Shared/collective action for future services
Hathi Trust (aka SDR)

• HathiTrust is a bold idea with big plans
  – Digital repository for the nation’s great research libraries

• HathiTrust is a solution for G-partners
  – Trusted solution to archiving vast amounts of digital content

• Shared funding & governance
  -- Common infrastructure, distributed development, scholar-sensitive
Hathi Trust

• Shared Digital Repository
  – Launched 2008 by 25 institutions (now 26)
  – Non-profit context
  – Initial focus on digitized book and journal content
  – Expanding to non-book/non-journal, born digital

• Collaborative development

• Services
  – Preservation and access ("light archive")
  – Print disabilities
  – APIs for data access through other interfaces
Hathi Trust Growth Projections

Currently 5.7 million volumes

There’s an elephant in the library.
Hathi Content Distribution

As of May 2:
5,729,692 - Total
914,214 - Public Domain
PROJECT LAUNCH @ UMN
April 2010 Launch!
What is Being Digitized

• Books and bound serials
• General collections, most locations
  – Wilson (Arts/Humanities) 38%
  – Bio-Medical 9%
  – St. Paul (Biol/Agric/Envir) 12%
  – Walter (Physical Sci/Engineering) 4%
  – Law 11%
  – MLAC (UM holdings) 23%
What Is NOT Being Sent?

• Rare and special collections

• Items that don’t pass size or condition criteria
UM Libraries Process

• ~ 1 million volumes
• Work balanced across 3-4 libraries simultaneously
• Pull volumes from the collection (“pick lists”)
• Charge items out to Google in catalog
• Items are shipped to Google scan center
• Google digitizes books & returns to UMN (6-12 weeks)
• UMN checks volumes in, re-shelve
Google Digitization Project

The University Libraries, in partnership with Google and through a joint agreement with the Committee on Institutional Cooperation (CIC) of the University of Chicago, is sending monographs (books) and serials (periodicals) to be scanned and made searchable. This will be done with Google to digitize millions of volumes in their library collections. This is an important step to ensure access to the collective resources held in our libraries. The Google project, coupled with programs to coordinate access to library collections in the future.

Frequently Asked Questions

Q: What if the book I want now is checked out to Google?

If you are a current student, staff, or faculty member, and another copy is held in the University Libraries, you can place a request for the book through Interlibrary Loan (most interlibrary loan requests are filled within 5-10 days).

Q: How many works will be digitized?

The University Libraries will digitize and archive more than 1 million volumes (books and bound journals) from their collection at least two years.

Q: When will the University of Minnesota make these materials available online?

http://www.lib.umn.edu/google
Proposed Settlement Framework

• Authorizes Google to continue scanning, additional display uses/products
• Allows display of in-copyright/out-of-print
• Creates Registry for rights holders (non-profit)
• Creates revenue stream for rights holders
  – Revenues from product sales, advertising
  – Revenues split between rights-holders and Google
• Creates 2 “research corpus” sites for non-consumptive research uses
  – Data mining
Access Models

• Preview Uses (free, up to 20%)

• Online Consumer Access
  – Enables user to buy online access to a single work ($1.99-$29.99)
  – Default pricing set by Google using an algorithmic pricing model
  – Rightsholder able to set price if desired.

• Institutional Subscription
  – FTE pricing for higher education and corporate markets.

• Public Access Terminal
  – Free on-site “terminals” for all public and university libraries in US.

• Additional models:
  – Print on demand, custom publishing, consumer subscription, PDF downloads.
The Future

• Settlement opportunities
  – Revealing in-copyright works
  – Broader access
  – Serving print disabilities
  – “Non-consumptive” uses for research
• Collaborative development & management

Yet to be imagined…
Questions?