The Google Books Project at Indiana University

Gary Charbonneau
Systems Librarian
Indiana University Libraries
May 21, 2009
What It Is: Part of a CIC Project

- University of Michigan and University of Wisconsin were previous participants under previous contracts.

- Indiana University was first participant under CIC contract.

- Contract negotiations on behalf of CIC led by president of IU.

- New contract being negotiated to cover changes brought about by settlement of lawsuit by publishers and authors.
What I Can’t Tell You

- Specific numbers (but “very large”).

- Specific technologies employed by Google, including:
  - Scanners (other than that technology is non-destructive)
  - Carts
What Is and Isn’t Being Digitized

• Not everything is going: we sent Google an extract from our catalog, they told us what they wanted via a “candidates list” (they are now trying to avoid duplicate digitization).

• Mostly print monographs, but some print serials.

• Books from Bloomington campus only.

• Rare books (Lilly Library) excluded.
What Is and Isn’t Being Digitized (cont’d)

- Both public domain and in-copyright (~95% is actually in-copyright or not known to be in public domain).

- Some books excluded due to size or condition
What We’ll Get from Google

• Digitized works will be added to Google Book Search with “attribution” to Indiana University.

• In GBS, fully indexed, but may or may not be full text depending on copyright status.

• Under current CIC contract: We will get digitized files for titles in public domain.

• Under revised contract: Digitized files for all titles?
How It Works

On IU’s end, primarily a logistics effort. We:

• Pull books from the stacks from pick list.

• Charge books out to Google in our ILS (for tracking and file preparation).
How It Works (cont’d)

• Books are shipped to Google scan center and returned at regular intervals (currently every three weeks).

• Google digitizes books and returns them to IU.

• We check books back in and re-shelve.
The End Product
A Typical User Scenario
A Typical User Scenario (cont’d)
A Typical User Scenario (cont’d)
A Typical User Scenario (cont’d)
A New Wrinkle: “Collections of Distinction”

• A concept introduced under the CIC contract.

• First collection of distinction: IU folklore collection. Being done now, will probably finish in October.

• Google will digitize entire collection (including duplicates from previous library partners).
Collections of Distinction (cont’d)

• Google might make the collection separately searchable.

• Not sure if there will be other “collections of distinction,” or if this will be the only one.
Folklore Collection (cont’d)

• Google may build on collection by adding folklore books that IU does not own.

• IU planning to complete its part of the collection by digitizing works (and individual pages) that Google cannot do.

• IU’s digital files will be shared with Google and become part of GBS
Another New Wrinkle: HathiTrust

• HathiTrust: Digital repository originally set up by/for CIC. University of California recently signed on. http://www.hathitrust.org.

• Most documents currently in HathiTrust are digital files received from Google by University of Michigan.

• Also a smaller number of files from University of Wisconsin.

• Files are stored on mirrored sites at University of Michigan and Indiana University
HathiTrust (cont’d)

- University of Michigan has begun downloading IU’s digital files from Google into HathiTrust; IU will not have to manage them itself.

- Other types of content will be added to HathiTrust eventually (e.g., archived versions of digital content purchased from commercial sources).
What Is IU Doing in Its Catalog?

• NOT adding bib records for books digitized by Google, nor are we adding 856 field links to the Google content to our existing print records.

• Instead, we have added the Google Book Search API to our OPAC. API provides link from the print record to Google.

• Have recently added the HathiTrust API as well.
Example from IU OPAC
Yet Another New Wrinkle: The CIC Government Documents Project

• Will digitize federal government documents.

• Participants will send duplicate copies from their collections.

• This digitization will be “destructive” – items will be dis-bound (and generally discarded).
The CIC Government Documents Project (cont’d)

- Google will use sheet-fed scanners to obtain better quality images.

- University of Minnesota was pilot site.

- Indiana will be third site and expects to start in late fall.
When Will We Be Done?

• At current rate, maybe some time around 2013?

• IU would like to increase the rate at which we are sending books to Google, which would of course mean that we would be done earlier.

• We did increase from once a month to once every three weeks.

• Depends on Google’s ability and willingness to accept more frequent shipments at their scan center.
QUESTIONS?