Scanning Key Content
of Text-Based Material at Point of Accessions or Cataloging

Final Report
11/16/2012

Project Summary
This project proposed to insert scanning into the acquisitions and/or cataloging process for the key content of some of Harvard's hidden printed book collections. Key content might include elements such as title pages, table of contents, bibliographies, indexes, inscriptions, etc. Our goal was and is to expand capabilities within our current infrastructure, challenge existing methodologies, speed up the process of making our collections available to our users, and lay groundwork for a much more dynamic and flexible future online catalog.

Accomplishments
As we found our original proposal had an ambitious scope for our timeline, we reconstructed the project into phases. We have completed the first phase:

- Researching scanners and image capture systems suitable for all project stages
- Consulting with colleagues both within and outside of Harvard and with vendors
- Producing digital images and depositing them into the test digital repository
- Demonstrating that the images can be displayed in the test HOLLIS catalog (see Figure 1 at end)

Challenges
A few of the challenges we faced were:

- The condensed schedule (from 10 to 5.5 months) resulted in more intensive time commitment than anticipated, made scheduling time with our consultants more difficult, and did not allow us sufficient time to master and evaluate our equipment
- The sheer amount of information and choices related to digitization technology can be almost confounding
- The catalogers in our group had little previous experience with the technologies associated with digitization projects which necessitated more foundational learning activities
- Efficient file naming, image processing, depositing, and linking continue to be a great challenge in our experimental workflow
- Understanding the process for purchasing, invoicing, and reimbursement would have been helpful

Next Steps
We would like to continue work on this project and will formally request an extension. We anticipate the next part of our project will include:

- OCR processing of the digital images and demonstrating a way to make the text searchable in the catalog (or elsewhere)
- Establishing routine workflows and evaluating them to inform feasibility of expanding this project from pilot to production; undertaking time studies, qualitative review of the images and OCR, and proposing a way to evaluate how/if users interact with the new content
- Sending a sample of material to Imaging Services for cost, quality and efficiency comparisons
- Beginning to explore the possibility of using OCR text for streamlined data entry by catalogers
- Partnering with the Automatic Subject Heading Extraction project to explore how applying their methodology to key content compares with extracting data from full text; their project will provide an evaluation of the quality of our OCR results
- Partnering with the Link-o-Matic project to facilitate at least part of the process

Budget
Total expenditures were $33,449.73 with the majority of expenditures on equipment. Major purchases included a Zeutschel Zeta scanner (at Houghton), an Atiz BookDrive Mini (at Divinity), and an array of smaller scanners and camera setups (at Schlesinger).

Publicity
Sharing our ideas and enthusiasm with others was by far the best way of publicizing our project. The Showcase was the perfect culmination of that effort.

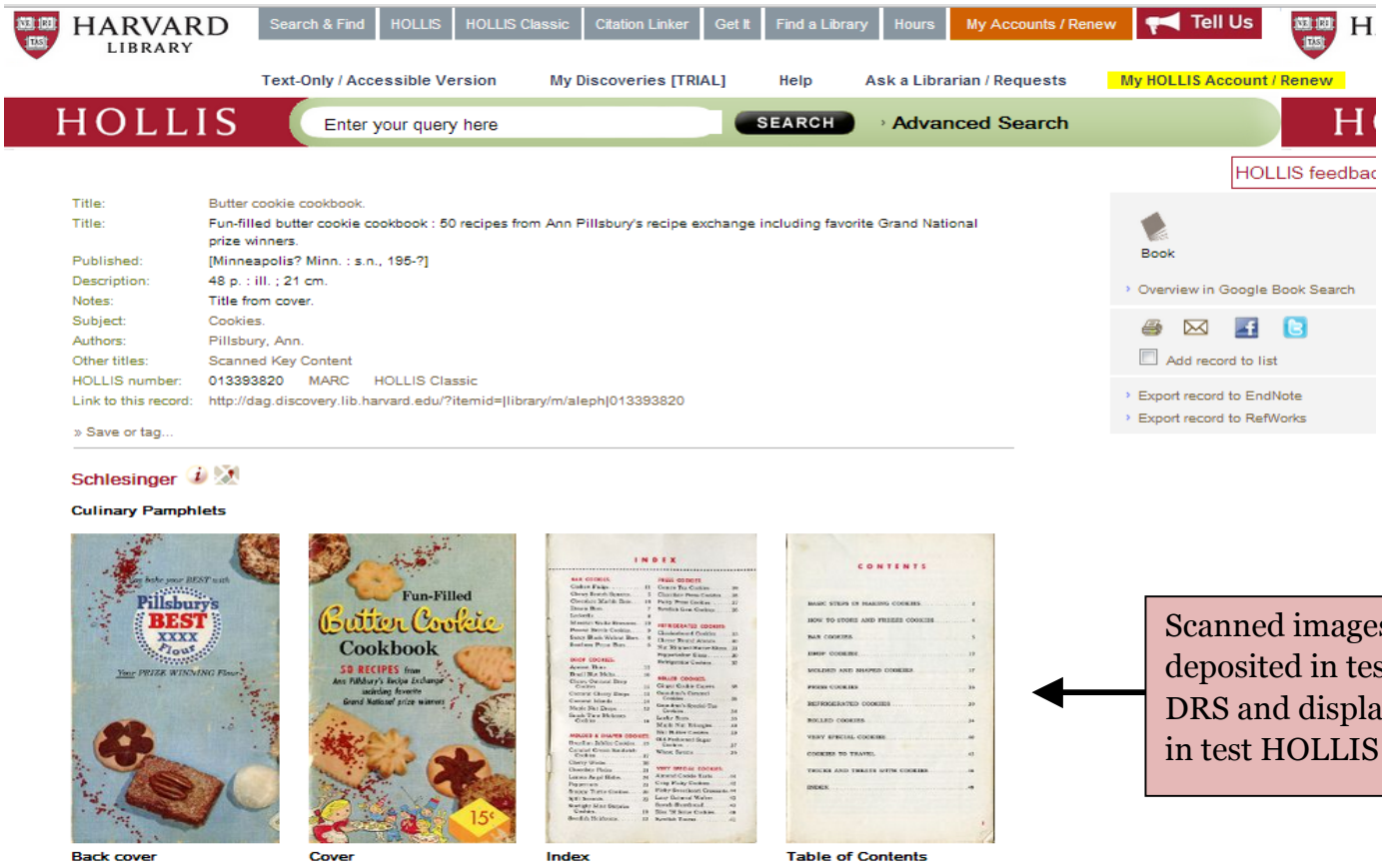
Presentations
- Library Lab Lightning Round
- Schlesinger Library Staff Retreat
- To the Houghton staff (planned)
- To the staff of Andover-Harvard Theological Library (planned)

Submitted by:

Amy Benson
Nell Carlson
Debbie Funkhouser
Karen Nipps

Additional project details (detailed budget, lists of resources, spreadsheets with equipment comparisons, etc.) available upon request.

Figure 1.



Scanned images deposited in test DRS and displayed in test HOLLIS