

# MASS DIGITIZATION

## PROGRAM EVALUATION

SEPTEMBER 2011



**Background**

In the spring of 2006, the CIC Library Directors adopted a set of principles for “Federating Mass Digital Efforts in the CIC Libraries.” The 2006 document states that the CIC libraries are committed to federating their digital resources, noting that “a collectively owned, comprehensive digital collection affords the consortium unprecedented potential to pursue coordinated decision-making and new financial models relative to the long-term retention, storage, and sharing of print collections.”

In 2007, the CIC partnered with Google to digitize as many as 10 million of the 79 million volumes held across all CIC library systems. This project represents one of the largest cooperative ventures of its kind in higher education, one that would enable CIC institutions to preserve a vast realm of legacy content and make material available worldwide. The University of Michigan and the University of Wisconsin-Madison already held separate contracts with Google for digitization prior to the CIC agreement.

These legacy collections are known to scholars worldwide, reflecting decades of careful investment and curation to build exceptional resources for research. The Google partnership will open up these resources to a much broader audience, ensuring that they remain accessible and discoverable in a digital age.

Beyond the scope and speed of digitization made possible by the Google agreement, the libraries have invested in a shared digital repository to house public domain materials. The HathiTrust repository allows faculty, students and the broader public to immediately access the full content of all member universities' rich array of public domain holdings digitized by Google while ensuring long term preservation of the material for future research.

Currently seven of the ten libraries are actively sending material to Google for scanning: Chicago, Illinois, Indiana, Minnesota, Northwestern, Penn State, and Purdue.

**Mass Digitization Program Evaluation**

CIC programs are regularly evaluated in order to gauge program effectiveness, measure progress and impact, and provide opportunities for adjustment and improvement. Each CIC Program Evaluation is guided by a standardized list of questions intended to elicit the benefits of the program to member universities relative to their investments and potentially other complementary or competing projects. This evaluation was based on the following questions:

- How was this program initiated and to serve what need? Does the same need exist today, or how is it different?
- Is this program unique to CIC institutions?
- Are aspects of this program duplicated on one or more campuses, or by another outside entity?

- What are the program mission, goals, and objectives? Are these goals being met today? How have they changed over the years?
- Describe how this program advances the CIC mission, vision and strategic directions.
- Which CIC schools participate in this program? Has that changed over time, and if so, how?
- How do CIC staff and CIC institution stakeholders regularly track the success of this program- what data are collected? How is this information incorporated into the ongoing implementation of this program?
- What resources are provided by CIC, member institutions, and other sources to support this program? How has this changed over time? Is this likely to be different in the future?
- What have been the unintended consequences of the program, if any?

This program evaluation has been led by:

University Librarians: Joan Giesecke (University of Nebraska-Lincoln); Paula Kaufman (University of Illinois); Judith Nadler (University of Chicago)

Digitization Survey Team: Sherry Byrne (Chicago); Gary Charbonneau (Indiana); Nicki Saylor (Iowa); Shawn Nicholson (Michigan State); Claire Stewart (Northwestern); Kim Armstrong (CIC).

### **Library Survey on Digitization Activities**

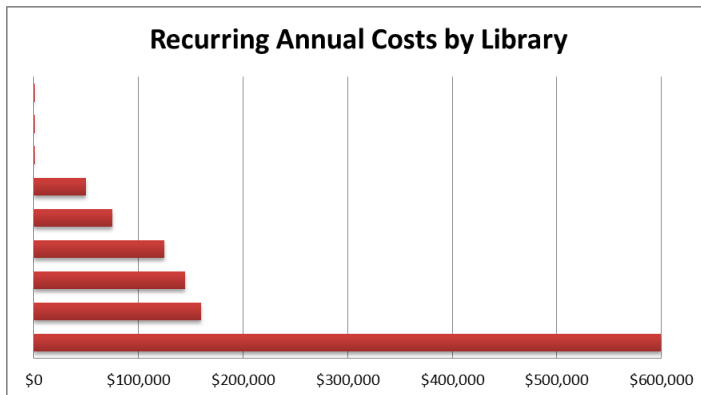
In August 2011, a survey was sent to all thirteen CIC member libraries with the purpose to:

- Determine the investments, in financial and human resources, made by individual libraries in the Google digitization program
- Articulate the value of the Google digitization program to member libraries and their users
- Document other digitization partnerships and business relationships in which CIC libraries are active
- Consider future CIC digitization activities

Results (N=11) are presented in two sections. Section I of the report focuses squarely on Google digitization activities. Members were asked to report on the costs and benefits of the Google scanning initiative, including how staff and other resources have been directed in light of the Google initiative. A particular emphasis of the survey was to gain insight into the impact the Google project has had on management of print collections. Section II reports on library digitization efforts beyond the Google initiative.

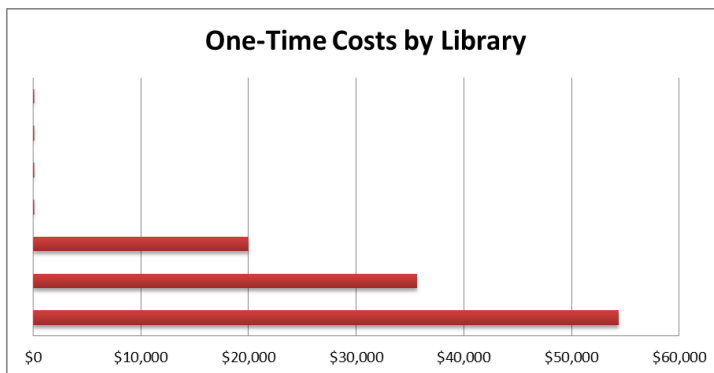
**SURVEY RESPONSES PART I: Google Cooperative Digitization**

## Annual Budget for Recurring Costs Related to Google Digitization



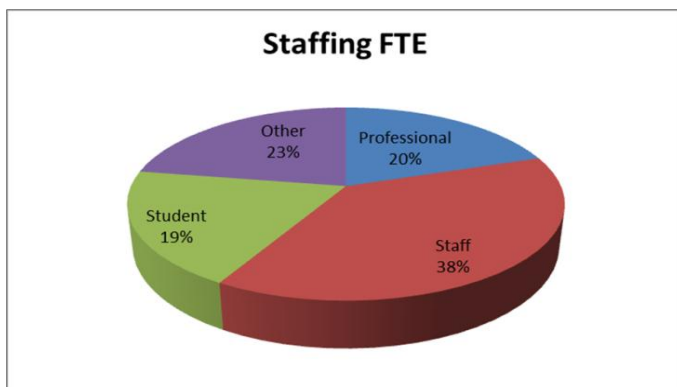
Currently, libraries budget annually between \$0 and \$599,000 for Google-related costs. Of 9 responses, 5 libraries reported annual budgets between \$50,000 and \$599,000. Five libraries reported no recurring budget. These costs are primarily staffing costs and do not include HathiTrust dues of \$50,000 annually.

## One-Time Costs to Prepare Collections or Space for Google Digitization



Three libraries reported that they expended between \$0 and \$55,000 in one-time start-up costs to prepare books for Google scanning. These costs are attributed to: equipment, supplies, binding, barcoding, and space remodeling. Other libraries had apparently not tracked these start-up costs.

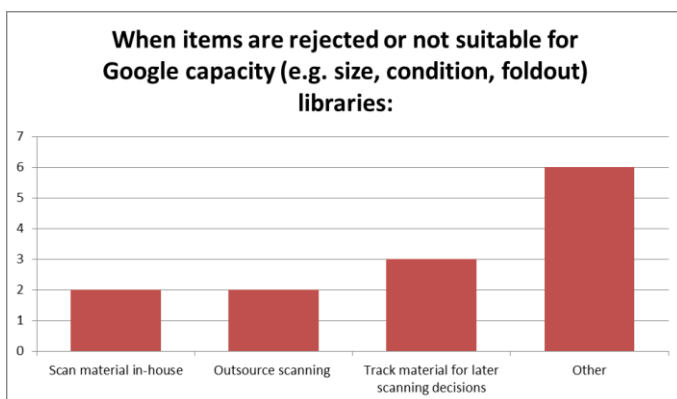
## Staffing Allocations for Libraries Currently Sending Material to Google



Libraries reported the following ranges of personnel dedicated to Google digitization activities:

Professional	0-6 FTE (though some not full-time on the project)
Staff	0-5FTE
Student	0-5 FTE

## Handling of Materials Not Suitable for Google Digitization



Each responding library indicated that they track books that are either rejected or unsuitable for Google scanning. Only a minority of libraries, however, have processes in place to digitize rejected items, either by scanning the material in-house or outsourcing the work to a digitization vendor.

### Adjustments to Individual Library Digitization Priorities as a Result of the Google Project

Because the Google Project is digitizing a large body of general collections material that is duplicated across many academic libraries, CIC respondents indicated that they have redirected local scanning resources away from widely circulating material to specialty and rare items. Non-Google scanning priorities reported by individual CIC libraries include: items unique to a single library collection, institution-published materials, subject collections and materials of priority to faculty researchers.

“We prioritize items that are unique to our collection or that we are fairly confident would otherwise not be digitized.” We routinely check Google/Hathi to see if an item has already been scanned.”

### How has the Google project benefitted your library?

Libraries reported positive impact from participation in the Google project both in terms of management of their print collections and in the opportunities the project provides for user communities at their universities. With the availability of public domain digital content both on the Google site and in the HathiTrust Shared Digital Repository, libraries have been able to repurpose key spaces in their libraries and utilize local and shared print storage capacity. For materials that were in a brittle or deteriorating state, the scale and pace of the Google project allowed these materials to be digitized before the content was lost.

Students and researchers at CIC universities have substantially improved access to library collections as a result of the Google partnership. Librarians have encouraged the use of Google Books materials to advanced scholars and undergraduates. The images are also used to fulfill interlibrary loan requests from non-CIC universities, saving our libraries operational expenses and protecting fragile print materials. The fulltext availability of public domain materials adds to the growing body of scholarly material available to users at anytime and anyplace they have access to the internet.

“Our participation has solidified some partner relationships with other institutions as a result of being a full participant in an important collaborative project.”

**SURVEY RESPONSES PART II: Non-Google Digitization Efforts**

Section two of the survey focused on questions about non-Google digitization activities. In particular, libraries were asked about other digitization partnerships, in-house capacity for digitization activities, outsourcing contracts, current challenges to digitizing materials, and digitization activities that would benefit from CIC-wide coordination. The survey also asked for a description of the collections that are currently being scanned, as well as collections that are priorities for future scanning projects.

The challenges libraries reported that prevent them from meeting their digitization goals include the need for additional funding and staffing. In addition, digitization units also expressed a need for specialized equipment, support for rights clearance, and development of digital access infrastructure. In short, digitization units encompass a wide variety of work tasks that range from simple scanning of material to determining how the material will be delivered, displayed, and ultimately preserved for the future.

Libraries were asked to list 3-5 activities in digitization that would benefit from coordination across CIC member libraries. There was significant alignment in the responses, suggesting that libraries would benefit from additional collective action or investment among their CIC partners to build on the successes of the Google project. Potential future areas of collaboration include:

- Building capacity for special formats (maps, microfilm, audio, video, film)
- Investing in rights clearance/copyright management
- Developing a coordinated contract for outsourcing digitization

“It would be incredibly helpful if the CIC were able to take some role to develop expertise and best practices which member institutions could draw upon when needed. Centralizing this activity would certainly help lessen duplication of effort across all of our institutions.”

**Current Status of Google Digitization and HathiTrust Shared Digital Repository**

The CIC continues to participate in the Google digitization project under the original terms of the 2007 agreement. Eight schools are actively supplying content for digitization, as are The Universities of Michigan and Wisconsin-Madison, who participate under their own agreements. In July 2011, Google indicated that they intended to significantly reduce their rate of scanning with library partners. Each CIC library has adjusted to shipping a smaller volume of materials at more protracted intervals. In addition, CIC libraries are primarily supplying out of copyright, public domain material that is fully viewable in Google Book Search and HathiTrust. As of

February 2011, the CIC schools had scanned approximately 1,000,000 items with Google and the public domain material that flows back to the HathiTrust exceeded 200,000 volumes. In-copyright material is held in escrow by Google per the terms of the 2007 agreement, although there is current discussion underway to give source schools the option of requesting that their in-copyright files be released to HathiTrust.

In addition to activities covered by our 2007 agreement, the CIC has partnered with Google in a separate scanning workflow to digitize a comprehensive set of U.S Federal documents. It is believed this collection will comprise between 1 and 1.5 million volumes. Digital facsimiles of successfully scanned Federal Documents from CIC institutions are fully viewable in the HathiTrust Digital Repository, as well as being indexed in and accessible through Google Book Search.

The CIC libraries are Founding members of the HathiTrust Digital Repository and make an annual investment of \$50,000 per library to support the repository's development and ongoing operations. HathiTrust is hosted at the University of Michigan and has a mirror site at Indiana University. Currently HathiTrust has over 60 members. Although the reduction in the Google scanning rate will impact the volume of material moving into the HathiTrust, CIC libraries have the option of contributing locally scanned files or files scanned with partners other than Google (e.g., the Internet Archive). Today, HathiTrust contains 9,613,343 volumes, of which 27%--2,624,698 volumes--are in the public domain and fully viewable.

### **Conclusion**

CIC universities, libraries, faculty, researchers, and students have derived benefit from the Google mass digitization project. Google has accelerated the preservation and discoverability of print book materials, with substantially higher throughput and lower costs than could be achieved by libraries, even if they had acted collectively to digitize this material. Despite the changes in Google's digitization plans over the life of the project, CIC libraries remain committed to scanning with Google.

As a result of the changes at Google, CIC libraries might now investigate new avenues to expand digitization collaboration and to build on the digitized body of material available on Google Book Search and in the HathiTrust Shared Digital Repository. Potential areas for additional co-investment or collaboration might include some or all of the following:

- building capacity to digitize non-print formats
- coordinated rights/copyright management
- a negotiated central contract for outsourcing additional digitization activities.

In addition to collaborating on ways to increase the body of available digitized content, CIC libraries also have the opportunity to build creative services and discoverability tools that offer scholars new avenues for accessing and exploiting the corpus of secure digital copies stored in



HathiTrust. Our libraries have enough confidence in the archiving strategies of HathiTrust, and the quality and utility of the public domain files being stored, that Hathi content is proving to be a catalyst for collaboratively managing print collections.

As one library responded in the survey, “Google-driven digitization has spurred innovation, collaboration, the development of technological skills, and provided insight into the digital realm of scholarly communication.”