Large-Scale Digitization:
Academics and Librarians Confront the Issue

A Report on *Scholarship and Libraries in Transition:*
A Dialogue about the Impacts of Mass Digitization

"We are at the threshold of an era unlike any other."

A Symposium Presented by
University of Michigan University Library
U.S. National Commission on Libraries and Information Science (NCLIS)

by Guy St. Clair

The impacts of large-scale digitization have yet to be completely grasped by all the players. Many professionals – in a wide variety of fields – are wondering just how this tremendous change will affect them and their work. More than 500 academic and research librarians, scholars, students, and representatives from government and publishing met in Ann Arbor, MI on March 10-11, 2006 to discuss the subject. By and large, the consensus was that large-scale digitization is good for scholarship, and continued efforts in this direction are encouraged. Nevertheless, there are dissenting views, and current litigation efforts were recognized and acknowledged as necessary for providing a legal operational framework to ensure that society at large benefits from the development and future realization of mass digitization efforts in libraries.

With the announcement in December, 2004 that Google would digitize the collections of the University of Michigan Library and selected titles from Harvard, Stanford, and Oxford and the New York Public Library, the subject of mass digitization moved from the domain of "wouldn't-it-be-nice-if…" to the world of practical reality. Few subjects have so affected librarianship (it is not uncommon to hear the idea of mass digitization compared to the development of the printed book), and there is no question but that all areas of librarianship and, indeed, of scholarship in general will be affected by this important change in the management of knowledge. While digitization programs have been around for some time, it is with the involvement of the commercial sector that these huge efforts can be accepted and sought after as an information, knowledge, and strategic learning management practice.

Supported in part by the John K. Goodrich Discretionary Library Fund, the Ann Arbor symposium was called by the University of Michigan Library and the U.S. National Commission on Libraries and Information Science (NCLIS)[*] to provide an opportunity for interested parties to air the many questions and concerns that have come up since Google made its announcement. Air them they did, and the many presentations, together with well thought-out and sometimes spirited remarks from many of the attendees, gave the two-day meeting an impressive framework in which to operate. No one, it seems, left the meeting without being intellectually stimulated (or provoked) in one way or another.

---

[*] [Disclosure: The author has served as a strategy planning consultant for NCLIS.]

The implications for knowledge services are clear. The value of mass digitization (at the symposium, the preferred term seems to have migrated to "*large-scale* digitization") for the delivery of knowledge services in all organizations – academic or otherwise – is pretty much a given. The easier it is for those searching for information, knowledge, and strategic learning to find what they require, the better it is for all concerned. Indeed, this is a goal not just for individual searchers but for organizations as well, as organizational leadership seeks to deal with the massive amounts of information required for the management of the organization. If digitized information is how most people want their information delivered (which is the case, as noted by several symposium participants), then large-scale digitization definitely has its advantages. And a natural result of the process is that the great digitized resource includes not only the captured knowledge of the past. The very concept of the library as a stand-alone warehouse of knowledge becomes secondary to its role as a repository for mission-critical information, a point not lost on attendees connected with specialized librarianship. While there was no specific reference to specialized librarianship and its requirement (as the discipline's founder John Cotton Dana put it) to provide "practical and utilitarian" information, there is no question but that practical and utilitarian information serving the cause of the commercial research sector benefits from large-scale digitization as well as that supporting the non-profit and academic sectors.

Representatives of the sponsoring organizations opened the proceedings. Speaking on behalf of the U.S. National Commission on Libraries and Information Science, Commission Chairperson Beth Fitzsimmons welcomed the attendees and assured attendees that the purpose for the symposium, to delve into the specific impacts of large-scale digitization, would not be lost on the Commission and its efforts. With a statutory mission to advise the President and the Congress on library and information science issues as they affect the American public, NCLIS is positioned to bring the discussions of the symposium to the attention of the nation's political leaders. In discussions with Commissioners (most of whom were in attendance at the symposium), it became clear that NCLIS expects to do exactly that.

Brenda Johnson and John Wilkin, Associate University Librarians at the University of Michigan, also welcomed attendees and described their professional enthusiasm for the large-scale digitization effort being undertaken, and with University of Michigan President Mary Sue Coleman established a very positive and friendly ambiance for the meeting. As the co-organizer of the event (with NCLIS), the University of Michigan University Library went to great lengths to ensure a good symposium, and they were remarkably successful in their effort, as noted by U.S. Government Printing Office CEO Bruce James in his presentation, when he spoke of the symposium as being one of the best-organized conferences he had attended.

President Coleman herself, in her Opening Remarks, described how the University Library's seven million volumes, also open to the public (an unusual arrangement in the academic library community), is a great public treasure. With the Google announcement, a "world library is on its way," she said. Emphasizing digitization as a preservation tool, President Coleman made it clear that this step is an important effort in "protecting the written word for all time." Obviously there are different points of view about technology (which can be seen as disruptive), but the ultimate goal of the effort is the "social good of sharing knowledge," a point well taken by symposium attendees. As for the controversy about works not in the public domain, President Coleman noted that, as far as the University of Michigan University Library is concerned, these works would remain "dark" until copyright is removed.[*]

Tim O'Reilly, Founder & CEO, O'Reilly Media, was the symposium's keynote speaker. Noting that books exist for three purposes, for reference, to teach, and to entertain, O'Reilly picked up on President Coleman's assertion about the preservation role of digitization. Commenting that a library is not just about the dissemination of information "but for the preservation of information," O'Reilly made a persuasive case for building a digital economy and recognizing that the emphasis in the future will not be on artifacts but on ongoing services, which digitized collections enable.

---

[*] On February 6, 2006 President Coleman spoke to the Association of American Publishers about the Google Library Project: http://www.umich.edu/pres/speeches/060206google.html

Five major areas for discussion were chosen by the symposium's organizers, and each focused on a particular subject or issue impacted by the digitization of library collections:

Libraries

Panel Moderator: Josie Parker, Director, Ann Arbor District Library
Panelists: Barbara Allen (Director, Committee on Institutional Cooperation), Michael Keller (University Librarian, Stanford University), Karin Wittenborg (University Librarian, University of Virginia).

"Mass digitization projects offer the opportunity to advance the Library's mission as part of a great public university by benefiting both our direct user population and the broader user community. … the library's role … as a center of the campus and a connection for users with the scholarship and learning they are seeking, remains stronger than ever…." [From the printed "Symposium Schedule"]

Highlights:

- Allen: sees the digitization project as a window of opportunity for experimentation in partnerships and the collaborative management of knowledge
- Keller: recommends changing the terms of reference – as a society we think of the library as a building – we need to think of the "library of the mind," the "library of the scholar" – there is a need to convince policy-makers that digitization projects are for the public good
- Wittenborn: "absolutely thrilled" about the project and recognizes the major importance of the effort – it changes the status quo and "that's a good thing"

Research, Teaching & Learning

Moderator: John King, Dean, School of Information, University of Michigan
Panelists: Jean-Claude Guédon (Professor, University of Montreal), Ed Tenner (Professor & Author, Princeton University), Ann Wolpert (Director of the Libraries, MIT)

"People from all walks of life have embraced Google as an indispensable tool for managing their personal lives. This has led to the development of new patterns of intellectual interaction, and scholarly communication. The question remains, however, as to whether Google has approached the degree of ubiquity as a scholarly research tool that it has as a tool for general discovery and organization…."

Highlights:

- Tanner: recognizes that Google and other search engines provide information that is "good enough" but does not necessarily provide the *best* information – asks what we can do as librarians and authors to raise the level of literacy
- Guédon: compares mass digitization with mass printing, noting the differences with digitized publishing – Wikipedia, for example, is not a thing – it is a process, a collaborative encyclopedia, aggregated to create a community
- Wolpert: describes user survey and results – library customers want a single interface to search across a variety of information sources, expanded online content, more access to all library material via a commercial search engine, and a "wizard" to help choose the best tools for a topic

Publishing

Moderator: Mark Sandler, Collection Development Officer, University of Michigan
Panelists: Suzanne BeDell (Vice President, ProQuest Information and Learning), Daniel Greenstein (University Librarian and Executive Director, California Digital Library), Alicia Wise (Chief Executive, Publishers Licensing Society)

"The future of publishing in an era of mass digitization, electronic publishing, and open access raises issues for copyright, scholarly communication, sustainable economic models, role of libraries, and the role of commercial and university press publishers."

Highlights:

- BeDell: states a negative reaction to the Google Library Project and recommends TCP (Text Creation Partnership) model
- Greenstein: notes that information is becoming part of the commodity infrastructure and describes the Open Book Alliance – comments about the importance of giving some attention to the digitization of materials other than books (records, archives, etc.)
- Wise: indicates no concerns about the program in principle but asserts that the Google Library Project is not well understood – legal framework is problematic

At the close of the first day, Adam Smith, Senior Business Product Manager at Google with primary responsibility for the Google Print and Google Scholar products, spoke about the company's goals in bringing these products to the public, and its commitment to full cooperation with the scholarly and academic communities.

The symposium continued the following day.

Economics

Moderator: Ron Milne, Acting Director of University Library Services & Bodley's Librarian, Oxford
Panelists: Paul Courant (Professor, University of Michigan), Karl Pohrt (Owner, Shaman Drum Bookshop), Hal Varian (Professor, University of California, Berkeley)

"There are many approaches to thinking about the economic impact of digitization initiatives…. the local economy … the State… a broad impact on national and even international economy…. "

Highlights:

- Courant: notes that with the digitization project, "we have everything we need to make what we do easier" – speaking very generally, comments that if the material isn't available in digital format, it won't be used
- Pohrt: notes that with universal access, information becomes available throughout the global community
- Varian: Nntes that in choosing between the opt-in and the opt-out models, opt-out is "the way to go," since transaction costs are too prohibitive – the Google Library Project model is the same as the Web and fair use – opt-out provides a valuable user service

Public Policy

Moderator: Nancy Davenport, President, Council on Library and Information Resources
Panelists: James Hilton (Associate Provost for Academic, Information and Instructional Technology Affairs and Interim University Librarian, University of Michigan), Bruce James (Chief Executive Officer, U.S. Government Printing Office), Brian Kahin (Senior Fellow, Computer & Communications Industry Association & Adjunct Professor, University of Michigan)

"…raises significant issues that will ultimately have impact on public policy. Copyright questions, open access, and the digital divide…. The role of the government in a partnership between a private and public institution will also be a point of much debate. We are at the threshold of an era unlike any other. Decisions we make now will lay the foundation for how future generations will be able to access and use online

content. We must take the time to make these decisions with the greatest of thought and care. …an opportunity to begin this discussion."

Highlights:

- Hilton: comments that the publishing community seems to see copyright as a licensing arrangement, and they are two different things – librarians will be more involved in the digital age but recommends some caution – "don't let technology transfer wag the academic dog"
- James: notes his enthusiasm about what Google is doing and suggests that if mistakes are made, we'll learn from them – his job is to save government information in perpetuity – notes that the government will partner with private sector to take advantage of private sector enterprise/expertise but "not going to trust a private company to protect government information in perpetuity"
- Kahin: reviews the four factors of fair use, and notes that the subject boils down to one issue: who is in the best position to make policy?

At the conclusion of the panel discussions, the conference ended with wrap-up comments and Closing Remarks from Clifford Lynch, Director, Coalition for Networked Information. Lynch noted that while many important subjects had been discussed in the symposium, there were other topics that were not touched upon. Delivery, for example, was not much discussed, and there needs to be conversation about how to build up a large-scale public domain library. Lynch also commented that the present situation is not so much a legal problem as a public policy problem, and that issues such as copyright need to be looked at in terms of a broader picture (not just books). Lynch also commented that digitization and large-scale replication can be considered as one of the best forms of insurance, not in terms of providing a replacement, of course, but in a content management/protection perspective, since digitization serves a very good stewardship purpose. He noted that the "biggest wild card" has to do with the changing use of texts, as we have moved considerably beyond the individual, personal use of text. What we are seeing is the rise of collaborative scholarship.

Full symposium information can be found at http://www.lib.umich.edu/mdp/symposium/index.html.

About the Author: Guy St. Clair is President and Consulting Specialist for Knowledge Services at SMR International, a management consulting practice with headquarters in New York, NY (www.smr-knowledge.com), and Special Consultant for EOS International, the library automation and knowledge management software development company in Carlsbad, CA (www.eosintl.com). He can be contacted at guystclair@smr-knowledge.com. The opinions expressed in this report are personal opinions of the author and do not necessarily reflect the corporate position of either SMR International or EOS International.