

Feasibility of a Primarily Digital Research Library

Geneva L. Henry
Rice University
P. O. Box 1892 MS 280
Houston, TX 77251-1892
ghenry@rice.edu

Lisa M. Spiro
Rice University
P. O. Box 1892 MS 280
Houston, TX 77251-1892
lspiro@rice.edu

ABSTRACT

This position paper explores the issues related to the feasibility of having a primarily digital research library support the teaching and research needs of a university. The Asian University for Women (AUW), a new university in Chittagong, Bangladesh, will open in September 2009. It must make a decision regarding the investment to be made in research resources to support the university. Mass digitization efforts now make it possible to consider establishing a research library that consists primarily of digital resources rather than print. There are, however, many issues that make this consideration quite complex and far from certain. In this paper we explore the issues at a preliminary level. We focus on four broad perspectives in order to begin addressing the complex interactions that must be considered in transitioning to a primarily digital research environment: technical, economic, policy and social issues. The purpose of this paper is to begin to explore a research agenda for transitioning from a model for libraries where resources are primarily print to one that is predominantly digital. Our research in this area is just beginning, so our purpose is to raise the issues rather than offer firm conclusions.

Categories and Subject Descriptors

H.3.7 [Digital Libraries]: Collection, dissemination, systems issues, user issues

General Terms

Management, Measurement, Documentation, Design, Economics, Reliability, Experimentation, Human Factors, Legal Aspects.

Keywords

Digital libraries, mass digitization, digital research, libraries.

1. INTRODUCTION

Recent mass digitization efforts such as Google Books [1] and the Open Content Alliance [2] are creating online collections that can rival those of many traditional research libraries, particularly for public domain books in English. While some protest this effort, the fact remains that digital content is becoming increasingly

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

BooksOnline '08, October 30, 2008, Napa Valley, California, USA.
Copyright 2008 ACM 978-1-60558-249-8/08/10...\$5.00.

available and will be used by scholars worldwide. The migration to digital-only content has already progressed significantly with scholarly journals. While a relatively small number of journals are published solely in print, most now provide only electronic access, or electronic plus print. This has shifted the way research is done in the scientific community, where researchers rely primarily on journal articles and very little on monographs. As long as the researcher's institution subscribes to a journal, access is available anywhere through the Internet. With the increasing availability of monographs online as a result of mass digitization efforts, it is now possible to begin assessing the possibility of relying only on digital access for research in all disciplines, not just those who rely primarily on journals.

Use of electronic texts produced through mass digitization efforts will require the development of new tools, licenses, usage policies, presentation approaches, preservation strategies and methods of validation. The resulting impact to scholarship is likely to be significant, with new research methods emerging from a primarily digital research environment. Cultural, economic and technical changes will continue to emerge as more traditional resources migrate to digital.

How, then, does one begin planning for this new landscape? Established research institutions that have served the research and teaching needs of their communities with well-honed processes and large print collections can look at these changes in an experimental way, assuming a gradual shift over some unknown period of time. If, however, you were starting a new university and had to make decisions today about how to build your resources to support your teaching and research, could you assume a primarily digital environment? What are the economic implications? How is the curriculum affected? What are your critical infrastructure needs and dependencies? How is research now accommodated if faculty and students are not required to seek resources in a common location staffed with people to help with resource discovery? This is a critical decision that requires looking to the future for what will likely exist, but also meeting immediate needs for establishing a respected research institution. The Asian University for Women (AUW) [3] is facing this decision at the present, thus providing the motivation for this research.

2. BACKGROUND OF AUW

AUW is intended to serve the community of rural, economically disadvantaged, and/or refugee women in South and Southeast Asia and the Middle East. It will open as a university in September 2009, fully intending to be a leading institution of higher education that provides a world-class education. AUW will occupy a unique niche in the higher education community,

offering both graduate and undergraduate curricula to this previously underserved population.

AUW is located in Chittagong, Bangladesh. Unlike well-established research universities in developed nations, there are limited scholarly resources readily available to AUW students and faculty to support their work. Building traditional library collections to support the liberal arts and sciences would be costly, but not having those collections will likely mean AUW will not be positioned to succeed in fulfilling its mission. Building a sufficient infrastructure to support digital resources will also be costly, but perhaps more forward-looking. A better understanding of the costs and benefits, considering both the tangible and intangible aspects of each, is needed for a well-informed and prudent decision.

3. CORE ISSUES FOR CONSIDERATION

We have identified a number of issues that arise when considering the question, “Is it feasible to consider establishing a primarily digital research library to fulfill the teaching and research mission of a new university?” The research is just now beginning; we are in the process of identifying the key issues that will drive our investigation. We can examine the issues from four broad perspectives: technical, economic, policy and social. It is very clear, however, that this multifaceted research agenda has a very broad scope, which will result in ongoing research over several years. As we begin identifying the core issues, we will carefully navigate each at a less-than-exhaustive level, synthesizing research undertaken by others into specific issues such as the usability of e-books and licensing considerations.

3.1 Technical Issues

From a technical perspective, the following preliminary issues present themselves:

Providing access to electronic materials. How should AUW provide access to electronic resources—through an ILS (Integrated Library System) or some other means? How will AUW ensure the long-term availability of electronic resources?

Suitability of discovery tools for searching and browsing. What work is being done in this area with mass digitization resources in mind? What level of sophistication can be expected in the short term, medium term and longer range? What are the key impediments to significant advancements?

Availability of research tools to support use of digital texts. To what extent will annotation tools support tagging, highlighting, creation of marginal notes, etc.? Do the annotations persist over time and can they be easily shared? Are citation tools readily available and easily integrated into publishing/word processing packages? Can text mining and visualization tools be implemented to enable sophisticated analysis of large textual collections?

Availability of e-readers that can accommodate research needs. Are online books as useful as their print versions? Can researchers “flip” through multiple books while doing research? How easily can multiple monographs be compared to each other? Are displays eye-friendly? Can e-readers store a sufficient quantity of resources to meet the scholar’s needs? Is navigation sufficient? All tools identified above integrated?

Quantity of digitized resources. Are enough scholarly resources available in digital formats to support scholarship—particularly

books not yet in the public domain? Can institutions get access to these collections through subscriptions or other means?

Quality of digitized resources. Has the resource been captured at sufficient quality for scholarly use? Is it complete? If it is text, is it fully searchable? What is the quality of its associated metadata?

Digital preservation. Is there a preservation plan in place with the host of the resource? Is there a persistent identifier associated with the resource that ensures it will always be available? Will the format migrate as technology changes? Are there duplicate copies of the resource at distributed locations? Is the integrity of the digital copy verifiable?

Validation of digitized resources. Can a scholar easily validate the resource so that it can be used as a trusted surrogate of the original printed source?

Infrastructure. Can the power infrastructure sustain all-digital resources? Can portable devices (e.g. e-readers) support prolonged use? Can network connectivity be guaranteed 24/7? Are power and networking available in the countries where access to the resources depends on them? Will faculty, students and staff have the hardware they need to use e-resources?

The list of technical issues could fill many pages; they are the primary drivers for the current research now emerging from the mass digitization activities. We now turn our attention to the other equally critical and significant issues that are not being as aggressively pursued at present, yet pose equal if not greater challenges to truly enabling a primarily all-digital research library.

3.2 Economic Issues

If we could establish that it is technically feasible to establish a primarily digital research library, the practical concerns of affordability would likely be a strong determinant of how dependent on digital resources a library would decide to be. Subscription rates for scientific journals, which are now mostly all electronic, provide an example of the impact of cost in building rich collections. Few research libraries can afford to subscribe to all the scientific titles their faculty may wish to access. The following issues are a starting point for considering economic costs and benefits:

Support for a primarily digital basic infrastructure. What is the cost of building a campus that will support access to digital resources? Is power available and affordable? Is networking available and affordable? What will the increased cooling requirements be? What are the savings of not building a library that will hold a substantial print collection to support the research of the university? What are the costs associated with servers and storage that must be refreshed on a 3 – 5 year lifecycle? What library facilities (such as service points, training classrooms, or collaborative areas) will need to be provided? Can institutions get access to print materials not available electronically through Interlibrary Loan?

Staffing considerations. How do staffing costs compare for supporting the greatly increased IT resources over print resources? What new functions will need to be supported? Is it possible to outsource services when a central library of traditional resources is no longer provided? Do the personnel costs for digital skills significantly outweigh the personnel costs of traditional librarian personnel? Is the expertise needed to support a primarily all-digital environment available?

Licensing fees. How will the cost of licensing access to digital resources compare to the cost of acquiring those same resources in print form? Will licensing fees over time be cost-prohibitive? Will print resources change their pricing models with increased digital resources?

End-user cost burden. What is the cost to faculty and students for acquiring devices and software necessary to access and use digital resources? What additional training costs will be incurred in learning how to use new technologies required for doing research in a primarily all-digital environment? What additional costs are incurred with print materials when students and faculty must acquire resources not available through their institution?

Preservation costs. How do the costs of digital preservation and traditional preservation compare?

3.3 Policy Issues

Policy issues are the greatest unknowns in assessing the feasibility of a primarily all-digital research library. The mass digitization projects have yet to address these key concerns. Technology and economics aside, without the appropriate permissions and policies in place to address necessary access, the option of fully digital resource dependency will not even be possible. The following are basic policy considerations:

Licensing options. Will it be possible to license works that have been digitized as part of the mass digitization efforts by commercial vendors that are not available via open access? Will educational institutions be given any special consideration in licensing? Will licensing be reliable for providing long-term access?

Liability. What will institutions, faculty and students be held liable for with respect to use of digital resources? What measures will be in place to protect uses of derivative works and partial uses of materials for teaching and research purposes?

Library services in support of digital resources. What new services will be needed to support a primarily digital research library? Will services be provided to address curation and preservation concerns of scientific datasets? What in-house services are needed/not needed when the resources are primarily digital?

Copyright. Will newer copyright models (e.g. Creative Commons licenses) create greater or more restrictive access to digital resources? How will faculty and students be educated about rights associated with digital resources?

Policy issues are likely to garner increased attention as the mass digitization efforts reach maturity.

3.4 Social Issues

Social and cultural issues are considered together under what we are addressing as the social concerns associated with moving to a primarily all-digital research library. There is currently a significant difference in acceptance of digital resources for scholarly use among disciplines. Humanities faculty, where reliance on monographs for research is strongest, have been, on the whole, the most reluctant to adopt digital resources for use in research. The following issues highlight social considerations that will impact the decision of how feasible a primarily digital research library is to achieve:

Trust. Will faculty and students trust digital surrogates as well as born digital scholarly resources? Will validation technologies be

sufficiently robust to meet the demands of the scholarly community?

Recruitment and retention. Will there be an impact on the ability to recruit and retain prominent faculty and top students if the library is primarily digital? Will certain disciplines raise more objections?

Developing students' skills and credibility as researchers. Would the preparation of students for research positions in the future be compromised if the research library was/was not primarily digital? How will the reliance on digital resources affect how students learn and what kind of research they can undertake?

Collaborative research. Will a primarily digital research library impact the level of collaboration with colleagues working on similar research worldwide?

Methods of conducting research. If resources are primarily digital, does the way research is conducted change? How does this impact the integration of research into the curriculum? Will the reliance on digital resources lead researchers to produce more digital scholarship?

Many of the social and cultural issues will be addressed in the long term as the products of mass digitization work their way into research and classroom environments.

3.5 Assumptions

In listing this preliminary set of issues, we are making the following assumptions about what does or will exist:

- Students and faculty will possess computers, e-book readers and other technologies that enable access to digital resources.
- The institution is responsible for providing a reliable infrastructure.
- Access to most digital resources will require licenses.
- E-book publications will continue to grow in coming years.
- Mass digitization efforts will not cease.

4. OBSERVATIONS TO DATE

Our research builds on preliminary studies conducted by the co-authors. In a study of "The Impact of Digital Resources on Humanities Research," which focused on scholars of American literature and culture, Spiro and Segal found that humanities researchers generally value electronic collections for offering more convenient access to research materials, but that they are reluctant to cite e-resources and often regard print as being more authoritative [4]. To get a rough sense how many research materials in the humanities are available in a digital format, Spiro searched for electronic versions of the nearly 300 works--primary and secondary, monographic and journal--cited in her 2002 dissertation on American literature [5]. She found that while 98% of secondary (post-1923) monographs have been digitized (and are typically made available through Google Books as limited preview, snippet view, or no preview), only 24% are available in full-text, most commonly through subscription services such as NetLibrary and Questia. In contrast, 76% of primary monographs and over 88% of primary and secondary periodicals are available as full-text.

Spiro also evaluated the quality of a sample of digitized texts cited in her dissertation, examining the quality of the scanning, OCR and metadata as well as the terms of use, convenience and

reputation [6]. Her preliminary conclusions indicate that while subscription-based thematic research collections such as Early American Fiction typically offer better image quality and conversion of text, the Open Content Alliance provides more comprehensive access to public domain research materials and is probably of sufficient quality for most scholarly uses. Geneva Henry identifies some of the challenges and opportunities for the twenty-first publishing industry in [7].

5. SUMMARY

Our position paper presents the broad set of issues impacting the transition from print resources to primarily digital resources for a research library. The investigation into these issues is just now beginning. We do not intend to address all of these concerns as we embark upon this research, but we do want to ensure that a sufficiently comprehensive research agenda is established to guide the critical, ongoing transition from print to digital. The Asian University for Women is faced now with the decision about how to create a research library from scratch. Identifying the issues for consideration is a first step in helping to find the right balance needed to establish a strong reputation and meet the teaching and research needs of the AUW faculty and students. We look forward to participating in the Books Online workshop to engage in discussion with others who share similar interests in understanding and meeting future educational needs, as mass digitization efforts make available a broad array of digital resources.

6. ACKNOWLEDGMENTS

Our thanks to Dr. Charles Henry of CLIR for initiating this research and to Dr. Nancy Dye, Vice-Chancellor and President of AUW, for launching the stimulating exploratory conversations that make this research possible.

7. REFERENCES

- [1] <http://books.google.com/googlebooks/library.html>.
- [2] <http://www.opencontentalliance.org/>.
- [3] <http://www.asian-university.org/>.
- [4] L. Spiro and J. Segal, *The Impact of Digital Resources on Humanities Research*, Rice University, 2007; http://library.rice.edu/services/digital_media_center/projects/the-impact-of-digital-resources-on-humanities-research. An essay based on this research is forthcoming in *The American Literature Scholar in the Digital Age*, to be published by the University of Michigan Press as part of its digitalculturebooks imprint.
- [5] L. Spiro, "How many texts have been digitized?" Digital Scholarship in the Humanities, May 5, 2008; <http://digitalscholarship.wordpress.com/2008/05/05/how-many-texts-have-been-digitized/>.
- [6] L. Spiro, "Evaluating the quality of electronic texts," Digital Scholarship in the Humanities, May 9 2008; <http://digitalscholarship.wordpress.com/2008/05/09/evaluating-the-quality-of-electronic-texts/>.
- [7] Henry, G.: On-line Publishing in the 21st Century. *D-Lib Magazine* (2003) 9(10); <http://www.dlib.org/dlib/october03/henry/10henry.html>.