Selecting Research Collections for Digitization: Applying the Harvard Model

KRISTINE R. BRANCOLINI

ABSTRACT
ONE OF THE MOST IMPORTANT CHALLENGES facing digital library planners is the selection of research collections for digitization. The costs associated with creating digital resources are significant. Planners must develop selection criteria and procedures in order to ensure that limited time and resources are committed to projects to digitize the most significant collections with the highest probability of successful completion. Librarians at many academic libraries have developed selection criteria for the creation of digital collections. These criteria consider many of the same factors that go into the decision to license or purchase information resources. However, there are additional considerations. Librarians at Harvard University have written the most comprehensive guide to selecting research collections for digitization. In this article, the author applies the Harvard Model to a digitization project at Indiana University in order to evaluate the appropriateness of the model for use at another institution and to adapt the model to local needs.

INTRODUCTION
Indiana University's Bloomington Libraries launched its first digital initiatives in the early 1990s, but it was not until November 1997 that a coalition of university partners created the Digital Library Program (http://www.dlib.indiana.edu). The Indiana University Digital Library Program is dedicated to the selection, production, and maintenance of a wide range of high-quality networked resources for scholars and students at

Kristine R. Brancolini, Digital Library Program, Indiana University Libraries, Main Library E170, 1320 E. Tenth Street, Bloomington, IN 47405

Indiana University and elsewhere. Building on a previous partnership with University Information Technology Services, it is a collaborative effort of the Indiana University Libraries, the Office of the Vice President for Information Technology, and the School of Library and Information Science. The goal of this collaboration is to capitalize on the institutional capabilities of this university, focusing university resources on digital library projects that support the teaching and research of IU faculty, support the learning and research of IU students, and foster research about digital libraries. Although one objective of the program is to support existing digital initiatives, such as the VARIATIONS Project in the Music Library, another is to encourage new digital initiatives, including projects to digitize portions of the research collections throughout the eight campuses of Indiana University. In the two years since the DLP was created, we have begun digitizing four research collections, two with internal funding and two with external funding, and are currently in the planning stages of a fifth project with partners from the Committee on Institutional Cooperation (CIC). We support digital operations to provide resources exclusively to affiliates of Indiana University, such as VARIATIONS and the DIDO Image Bank and digital collections offered via the Web.

All academic institutions that are planning and implementing digitization projects confront issues related to selecting collections for digitization. With limited time and resources, libraries can only undertake a limited number of digitization projects, based on wise and expeditious choices. A number of academic libraries have developed criteria and models for selecting research collections for digitization, including Columbia (1998), Harvard (1998), University of California (1997), and Oxford University (Lee, 1999). The most comprehensive model is the work of Dan Hazen, Jeffrey Horrell, and Jan Merrill-Oldham, of Harvard University, published in the CLIR monograph, *Selecting Research Collections for Digitization* (1998), referred to throughout this article as the Harvard Model. One of the attractive features of this monograph is that it includes a graphical matrix for decision making, summarizing the steps and questions outlined in the essay (see Figure 1). In order to evaluate this model, the author used it to reconsider the first DLP digitization project, an internally-funded project to digitize the Lilly Library's *Frank M. Hohenberger Photograph Collection*.

The Web site http://www.dlib.indiana.edu/collections/lilly/hohenberger/index.html, referred to throughout this article, contains a part of the Hohenberger Collection. The purpose of this evaluation was to answer the following questions: Would using the Harvard Model have led to the decision to digitize this collection? Does the Harvard Model include the major factors that were actually used to reach the decision? How might the model be customized to provide more reliable guidance to project planners in the DLP?
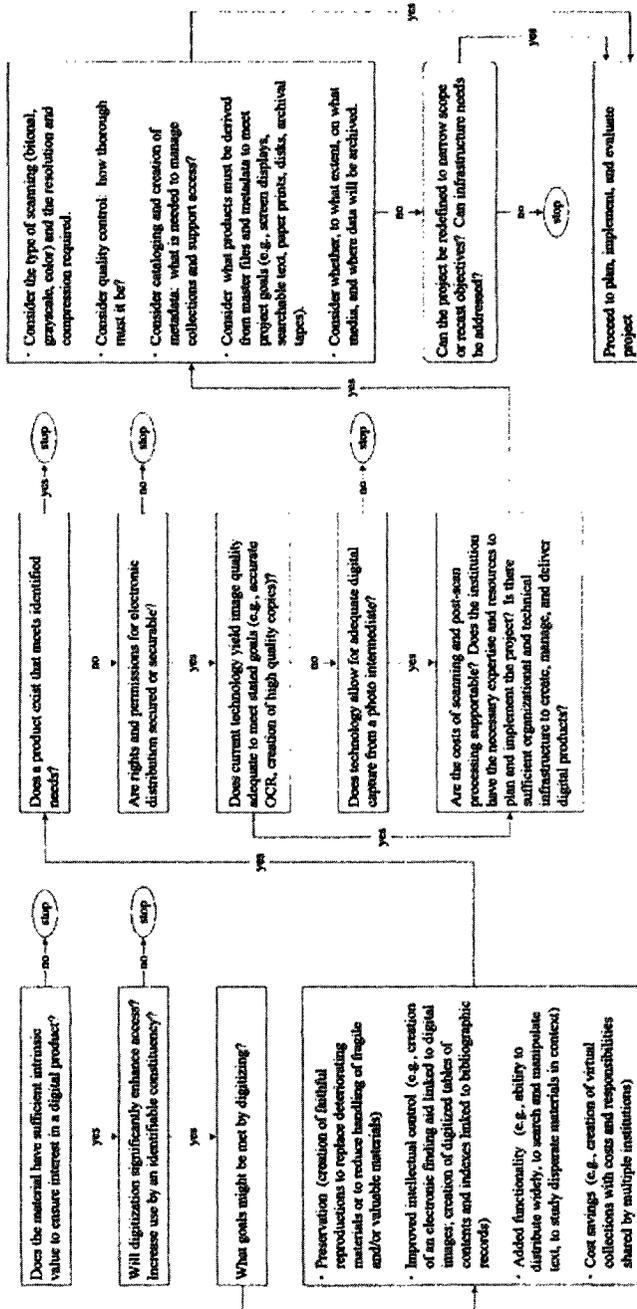
Figure 1. Selection for Digitizing: A Decision-Making Matrix.

COSTS OF DIGITIZATION

The decision to digitize a collection is similar to the decision to acquire an information resource through purchase or license agreement, but there are significant differences. For any selection, the selector evaluates the significance of the resource, its potential use, its relationship to other resources in the collection, its format, and its cost. For acquired resources, it is not necessary to include the costs of cataloging and overhead. However, these costs must be considered for any digitization project, and it is difficult to estimate the complete costs for a digitization project.

In a recent article in *RLG DigiNews*, Steve Puglia (1999) analyzes cost data from many digital projects. He found that the actual production costs differ significantly, depending on the category of material being digitized and whether or not textual material will be processed to become keyword searchable in addition to image scanned. He also found that, for all projects, there are hidden costs. For image projects, without text encoding, the overall average production costs of $17.65 per image breaks down as follows:

1/3 the cost is digital conversion (32 percent overall)—adjusted average $6.15 per image.

Slightly less than 1/3 the cost is in metadata creation, including cataloging, description, and indexing (29 percent overall)—adjusted average $7.00 per image.

Slightly more than 1/3 the cost is in other activities, such as administration and quality control (39 percent overall)—adjusted average $10.10 per image.

However, the most illuminating figures might be the range in projected production costs reported for individual projects: digitizing—$0.25-$16.65; metadata creation—$0.75-$17.25; other—$0.45-$28.15; total—$1.85-$42.45. These figures do not include the cost of maintaining digital resources. Few sources exist for these data and Puglia found that costs vary widely. Puglia notes: "Often major IT infrastructure costs are budgeted separately from digitizing projects, and therefore the network upgrade and database development costs were not factored into the estimates for long-term maintenance for the digital images." One estimate of the cost of maintaining master files and online access files was $1.70 to $4.70 per image for the first ten years or 14 percent to 38 percent of the initial cost per image.

On the basis of cost alone, it is essential that academic institutions choose digital projects carefully based on an established set of criteria within a well-planned procedure. Even for institutions with ample funding for the creation of digital resources, there are limitations of time, technology, and expertise.

HOHENBERGER PHOTOGRAPH COLLECTION

In discussions with librarians and archivists at Indiana University about the new Digital Library Program (DLP), many of the first questions dealt with digitization of research collections. DLP planners realized the need for selection criteria for digitization as well as a procedure for making the actual selection. Many libraries and archives within the institution have suitable collections, so the task was to prioritize projects. In order to frame the discussion, participants were asked to identify their most significant collections, preferably ones in the public domain or with Indiana University-held copyrights. DLP staff posed series of questions about the collections and their users. These questions focused on the copyright status of the collection; its size; its popularity; its use; its physical condition; the formats included in the collection—i.e., text, images, audio, film; the existence of electronic finding aids; and more. One outcome of these meetings was to begin writing project proposals for the most promising collections. The goal was to have some projects in development when suitable external funding opportunities became available. The staff used the LC/Ameritech Competition proposal outline in order to evaluate collections for digitization (Library of Congress, 1999).

As a result of these discussions and further evaluation, DLP project planners selected a collection for an internally funded project, images from the *Frank M. Hohenberger Photograph Collection.* Frank Hohenberger was a nationally-recognized photographer who lived and worked in the small town of Nashville, Indiana, from 1917 until his death in 1963. Upon his death, Hohenberger bequeathed his entire photograph collection, totaling more than 9,000 images and personal papers, to the Lilly Library. Based on past use, a Lilly Library curator selected 400 photographs, then an additional 100, for digitization. The digitization project began in spring 1998, and there are now 500 photographs on the site. The current plan is to outsource the digitization of the remainder of the photographs. The site also includes biographical information about Hohenberger and a 1933 article published about him in *American Magazine.*

THE HARVARD MODEL

The model described in *Selecting Research Collections for Digitization* grew out of Hazen, Horrell, and Merrill-Oldham's work at Harvard University. A task force was appointed in late 1995, charged with drafting a white paper to help Harvard's librarians and curators plan digital projects. The author of this article was impressed by the comprehensiveness of their model and their placement of the selection process into "the larger framework of collection building by focusing first on the nature of the collections and their use, and second, on the realities of the institutional context in which these decisions are made" (Hazen, Horrell, & Merrill-Oldham, 1998). Hazen, Horrell, and Merrill-Oldham note that the decision to

digitize a collection is similar to the decision to purchase, microfilm, and withdraw library materials. However, this process is further complicated by a range of procedures and technologies with widely varying implications and costs. They concluded that the judgments made in selecting collections for digitization involve the following factors: "the intellectual and physical nature of the source materials; the number and location of current and potential users; the current and potential nature of use; the format and nature of the proposed digital product and how it will be described, delivered, and archived; how the proposed product relates to other digitization efforts; and projections of costs in relation to benefits."

Overall, the Harvard Model seems to be more a planning model than a selection model, but this is by design. The authors note in their conclusion: "The process of deciding what to digitize anticipates all the major stages of implementation." By asking difficult questions about the collection and the proposed digitization project, librarians minimize the chances of making a costly error in judgment.

APPLYING THE HARVARD MODEL TO THE HOHENBERGER PROJECT

Hazen, Horrell, and Merrill-Oldham (1998) pose a series of questions in their essay. These questions are represented in their model, "Selection for Digitizing: A Decision-Making Matrix." The questions posed in the essay lay the groundwork for a plan of work should the collection under consideration be selected for digitization. The matrix includes nine broad questions that require a "yes" or "no" answer. Answering "no" to any question should stop the evaluation process for a given collection. Patently unsuitable collections will be eliminated early in the process, saving the effort of answering all other questions. Answering these questions requires considerable thought and investigation. One would want to apply it to a collection only after careful pre-selection.

One important issue is copyright. Formally, it lies outside the model, yet the authors call it "the place to begin." Following their advice, the author began with copyright, then used the questions posed in the essay to assess the decision to digitize the Hohenberger Collection.

Copyright

It was easy to resolve questions of copyright with regard to the Hohenberger Collection. The majority of the photographs are still covered by copyright, having been made between 1906 and the early 1960s. However, when Hohenberger bequeathed his collection to the Lilly Library, he assigned the copyrights to the Indiana University Foundation, which subsequently transferred them to the university. Thus, Indiana University could digitize the photographs and offer them on the Web without seeking permission. Of course, DLP staff now want to protect the

university's interests, which is accomplished by offering only relatively low-resolution images on the Web and providing access to the high-resolution TIF files only at the Lilly Library. Instead of worrying about seeking permissions, Lilly Library is concerned about infringements.

The Intellectual Nature of the Source Materials

The model begins with questions requiring highly subjective judgments, proceeds to more measurable judgments, then, with cost benefits, it becomes more subjective again. The first factor in the decision-making process requires a subjective judgment concerning the intellectual value of the collection under consideration. In making this determination, the authors pose a series of questions. Throughout this article, Hazen, Horrell, and Merrill-Oldham's specific questions will be given in italics.

Does the intellectual quality of the source material warrant the level of access made possible by digitizing? For the Hohenberger Collection project, planners relied on the judgment of the curators at Lilly Library. At the first meeting of DLP staff and Lilly staff, this was one of three collections presented for consideration. Frank Hohenberger and his work as both a photographer and a journalist have been the subject of scholarly publications including articles in the *Journal of Indiana History* and two books published by the Indiana University Press. In addition to the photographs, Hohenberger kept work-related diaries, recording the subjects of his photographs and anecdotes about them. These anecdotes formed the basis of character studies that he published in a column in the Sunday *Indianapolis Star* called "Down in the Hills 'o Brown County" from 1923 to 1932 and from 1936 to 1954. For this work, he was posthumously voted into the Indiana Journalism Hall of Fame in 1976.

Will digitization enhance the intellectual value of the material? The Web site with the digital images of the photographs allows users to scan the thumbnails or view higher-resolution images. It also provides captions with the images instead of separating them in a finding aid. The TIF files allow users to zoom in on the image, revealing details that are impossible to detect in the photographic prints without intense magnification.

Will electronic access to a body of information add significantly to its potential to enlighten, or are the original books, manuscripts, photographs, or paintings sufficient to the task? The first part of the Hohenberger Project was to mount an EAD-encoded finding aid. Previously, users were limited to using a printed finding aid. Even having the keyword-searchable finding aid without the digital images (and the site does not offer a significant percentage of the collection at this time) provides users with a valuable research tool. The collection is heavily used by genealogical researchers looking for photographs of their ancestors or ancestral homes. The Web site allows searching by name and date, providing dramatically improved access to the collection over the paper inventory. Furthermore, catalogers are enhancing

the inventory by adding Library of Congress subject headings to each photograph, thereby improving access to the collection.

To what extent will the combination or aggregation of original sources increase their value? Although Indiana University owns and plans to digitize Hohenberger's entire body of photographic work for this Web site, other images may be located in other collections. However, the DLP also plans to digitize other complementary photograph collections. The DLP is currently digitizing a collection of U.S. Steel photographs from the Calumet Regional Archive at Indiana University Northwest in Gary, Indiana. *Steelmaker-Steeltown: U.S. Steel Photograph Collection, 1906-1941* consists of 1,900 photographs made between 1906 and 1941, documenting the early years of Gary, Indiana, and the steelworks there. The DLP hopes to work with libraries and archives throughout the state, such as the Indiana Historical Society, to digitize other complementary collections and provide access to federated digital collections of images depicting people, places, and events across Indiana.

Current and Potential Users

Indiana University purchases resources to meet the needs of scholars across a broad range of disciplines. Bibliographers rarely make purchase decisions based on the number of potential users. However, the creation of digital resources varies in this regard. It is imperative to focus resources on collections that have the largest potential audience.

Are scholars now consulting the proposed source materials? Are the materials being used as much as they might be? The Hohenberger photographs are among the most heavily used resources in the Lilly Library. The curators identified this collection as a possible candidate for digitization because it is the subject of both scholarly and popular use. Many members of the general public request access to the collection. As a public university, providing information resources to the residents of Indiana is an important part of our mission. Project planners were convinced that providing digital access to the collection would increase its popularity.

Is current access to the proposed materials so difficult that digitization will create a new audience? The Lilly Library provides service to all users, but there are several obstacles to accessing the Hohenberger Collection. First, the library operates a limited number of hours and users must visit the library to see the Hohenberger photographs. Online access allows browsing, searching, and viewing digital surrogates twenty-four hours a day, seven days a week. Second, before creation of the online finding aid, users had to visit the library to use the finding aid, which provides inadequate and inconsistent access to individual images anyway. The online finding aid is more accessible and more useful. Third, users had to know that the collection exists and that it is at the Lilly Library. Through the Web, many people have found the photographs who were completely unaware of their

existence. Since September 1998, the site has received an average of 1,600 hits per month from all over the world. Fourth, the site is reaching young people, including students of Indiana history, who would never have considered visiting the Lilly Library even if they knew that the Hohenberger photographs are available there.

Does the physical condition of the original materials limit their use? Because the collection is so popular, many of the photographic prints had begun to show signs of wear. The Lilly Library has negatives for most of the photographs, so it is possible to make new prints. However, the digital surrogates have reduced handling of the prints. They are still available for examination, but most users are content with the Web versions.

Are related materials so widely dispersed that they cannot be studied in context? Lilly Library curators believe that the library holds the most significant Hohenberger materials.

Will the proposed digital files be of manageable size and format? There are numerous standards for the digitization and delivery of photographs on the Web. Project technical staff consulted the guidelines that the Library of Congress published for the American Memory Project and quickly concluded that the Hohenberger photographs would present no storage problems.

Will digitization address the needs of local students and scholars? The Lilly librarians assured DLP planners that the Hohenberger photographs would be of interest to students and faculty of Indiana University as well as the general public. However, it is probable that the digital collection is of greater interest to non-IU affiliates. For a public university, that may be a positive factor.

Actual and Anticipated Nature of Use

How do scholars use the existing source materials? What approach to digitization will facilitate their work? Project planners considered this project after viewing numerous photograph collections on the Web. Technical staff were convinced that the Web, coupled with the high-resolution images available via the campus network, would provide high enough fidelity to the originals to satisfy the needs of most users. Moreover, the thumbnails provide users with a quick way to browse the available digital images, while the complete finding aid provides information about the other photographs available at the Lilly Library. One service option might be digitizing on demand for remote users who would like to see an image that has not yet been digitized.

Will digitization increase the utility of the source materials? Will it enable new kinds of teaching or research? Do scholars agree that the proposed product will be useful? One of the goals of the project is to increase use of the collection in teaching. An intern from the School of Library and Information Science has created learning activities designed for grades 4-8 using the

Hohenberger photographs to teach visual literacy. These activities will be tested with students in 2000. Project planners also considered the possibility of working with a photography professor on campus to use the images with his beginning photography students. However, project planners did not consult with teachers or faculty while considering the collection for digitization. Perhaps this should have been a factor in the selection process. Instead, planners relied on the proven popularity of the collection.

Are there other scholars, librarians, and archivists who can collaborate to create a useful product? Given the broad appeal of the collection, scholars, librarians, and archivists throughout the state and perhaps regionally could contribute to the project.

The Format and Nature of the Digital Product

What critical features of the source material must be captured in the digital product? Are very high resolution copies, accurate rendition of color, a seamless combination of images and text, or other qualities considered essential? Project planners did not anticipate that the Hohenberger photographs would present special digitization problems. The DLP Visual Resources Specialist confirmed this initial assessment.

If the original sources are to be retained, can they withstand the digitization process? This was not an issue in the Hohenberger project. Upon the recommendation of the Lilly staff, project members planned to digitize duplicate negatives that are in good condition.

What type of hardware should be used for conversion? All photographs in the collection are black and white. There are established standards for digitizing black and white photographs. The Visual Resources Specialist identified possible scanners during the early planning stages.

Will a digitized sample meet users' needs? If so, how should the sample be constructed? The goal of the Hohenberger project is to digitize the entire collection of photographs and allow users complete access. However, a Lilly Library curator selected 500 images for a pilot project. This group constitutes highlights from Hohenberger's body of work, including representative and popular images.

Will the information resources upon which the project is based continue to grow? No, the collection is believed to be complete.

How will users navigate within and among digital collections? The finding aid is based on Hohenberger's organization scheme, which has value as well as drawbacks. Hohenberger organized his photographs for retrieval by size. This scheme is not particularly informative to users, so the finding aid offers keyword searching as well as browsing by Hohenberger's categories. Some of the categories are topical but most are not, so project planners proposed subject enhancements to the finding aid. It has been a challenge to retain important information from the artist while providing additional information that will improve access to individual images.

Describing, Delivering, and Retaining the Digital Product

How will users know the digital file exists? The metadata specialist on the Digital Library Program team created a plan for providing access to the collection. She created a collection-level MARC record for our online catalog with PURLs for the collection home page and the finding aid.

How can the digital product best be delivered to users? The project was designed for Web delivery from the beginning. Project planners never considered CD-ROM, near-line, or off-line storage. However, once the entire collection has been digitized, project staff may learn that only a small percentage of the images are actually being used regularly; this issue could be reexamined.

Who will be authorized to use the digital resource and under what circumstances? The university always intended to provide access to this resource free of charge on the Web. However, the copyright statement specifies how the images may be used. They are not intended for publication, but the university relies on their resolution to prevent misuse. The Lilly Library sells reprints of the images for private use and licenses them for publication. This information is on the Web site accompanied by an order form.

How will the integrity of the digitized data be ensured? Project planners did not consider this factor while considering the collection for digitization or during project planning.

Particularly for digital products created to meet local demand, is the existing technology infrastructure adequate? This was not an issue with the Hohenberger Project, since the Digital Library Program has the necessary infrastructure to support Web delivery of this collection.

What are the long-term intentions for the digital file? Indiana University intends to retain the files indefinitely and will create a plan for the longevity of all Digital Library Program collections.

Is the long-term preservation of deteriorated materials a project goal? Protecting the prints and negatives is a goal of the project. It was anticipated that the digital files would reduce handling of the originals and that most users would not require access to them for their information needs.

Relationships to Other Digital Efforts

Have the materials proposed for digitization already been converted to electronic form? No one had digitized a significant number of Hohenberger photographs.

Can cooperative digitization efforts bring together a cohesive body of material that would otherwise remain disassociated? Not in this case but, as noted above, DLP planners are interested in creating complementary collections of photographs, working in cooperation with other institutions.

Costs and Benefits

Project planners did not conduct an extensive cost/benefit analysis for the Hohenberger Project. However, technical staff knew from

preliminary investigations that the cost of storage for the grayscale images at the proposed resolutions would not be especially high.

Who will benefit from the proposed digital product? One of the important beneficiaries of the proposed digital project would be Lilly Library staff. At times they have been overwhelmed with requests for access to the Hohenberger photographs. Project planners were interested in providing relief for the staff as well as improving access to users and protecting the photographic prints.

Is the intellectual value of the proposed product commensurate with the expense? This is a difficult judgment to make. Project evaluators are asked to decide whether this project will have the same impact that another project might have. It is impossible to know whether this is the best possible project, so evaluators must be satisfied with the judgment that it will have a significant impact on users and the condition of an important collection. In some cases, alternative methods of reformatting would be considered here. With regard to the Hohenberger Collection, digitization is the only logical reformatting option for delivery via the Web.

Could an acceptable product be created at lower cost? The digitization plan is as inexpensive as possible given the nature of the collection. Because the project plan relies on the prints and negatives for back-up, image technicians are scanning for access only. For another project to digitize the university's Hoagy Carmichael collection, technicians are scanning twice, once for archival purposes with standard settings and once for access, manipulating the image to improve its appearance on the Web if necessary. To reduce costs, planners eliminated the archival scan from the Hohenberger Project.

How will the proposed project address the long-term costs associated with digital files? This question asks project evaluators to predict whether a digitization project might result in cost savings in another part of the institution. With regard to the Hohenberger Project, this is difficult to predict. However, planners anticipate that a digital collection might save staff costs in the Lilly Library and produce revenues by publicizing the collection. The Lilly Library sells photographic prints from the collection and licenses them for publication. The digital collection might lead to increased orders for reprints or reproduction and the digital files enable those prints to be made more easily and more cost-effectively. Staff can send the digital file(s) to Photographic Services, saving time as well as wear to the negative.

Can external funding be secured to support the proposed project? Although some digitization projects are so large that they require external funding, the Hohenberger Project was designed for internal funding. The concept was to select a small percentage of the photographs in order to learn the process of digitizing photographs for Web delivery, then consider a grant-funded project to digitize the remainder of the collection. Options for

grant opportunities figured into the selection decision. If the collection had no potential for outside funding, it is doubtful that the DLP would have undertaken the pilot project.

ACTUAL DECISION-MAKING PROCESS

To evaluate the Hohenberger Photograph Collection for digitization, project planners actually used a process similar to the Harvard Model based on the proposal outline for the Library of Congress/Ameritech Competition (http://memory.loc.gov/ammem/award/index.html). The DLP staff had recently submitted a proposal to the competition and were familiar with its content. The three-year program ended with the 1998/99 competition, but the LC/Ameritech Competition used a two-part evaluation process that provides a framework for planning a digitization project, especially one with preservation goals. For the competition, the first group of evaluators assessed content issues:

"Significance of the collection's content for understanding United States history and culture, as well as its breadth of interest and utility to students and the general public."
"Availability and usability of aids to intellectual access that can be integrated into the American Memory resource."

The second group of evaluators assessed technical issues:

"Technical and administrative viability of the project's plan of work in relation to the scope of the project."

Although project planners did not intend to submit a proposal to the LC/Ameritech competition to digitize the Hohenberger photographs, using that outline for the content evaluation would provide a framework for the collection selection and preliminary project planning. In addition to collection factors, project planners also considered intellectual access and institutional factors. With regard to intellectual access, the primary considerations were that we had a finding aid in electronic form (a word processing file) and that we have expertise in creating EAD-encoded finding aids. Although it would have been possible to create the finding aid from a paper inventory, this step would have slowed the project timeline considerably. Speed was a factor in selecting a collection for digitization. Of course, the additional work would have also resulted in additional costs.

The factors that project planners considered in selecting the Hohenberger photographs for digitization can be summarized as follows:

Collection Factors
- Description—How large is the collection? What is its content?
- Significance—Why is it important?
- Audience and Users—Who uses it now? Who might use it on the Web?

- Preparation and Preservation of the Collection—What preparation would the collection require for digitization? What are the preservation issues?
- Ownership, Privacy, and Copyright Issues—Who holds the copyright? How does the university want to protect its copyright?
- Intellectual Access—What does the finding aid include? How difficult would it be to convert to an EAD finding aid?

Institutional Factors—Added to the LC/Ameritech outline
- *Collection Format—Photographs*—The Digital Library Program wanted to digitize all or a selection from an archival photograph collection to gain the experience and develop expertise.
- *Indiana Connection*—The collection is the work of a local artist with a national reputation. Indiana University owns the collection and controls the copyrights. We are the logical creators of a digital collection of Hohenberger's work.
- *Benefits to Lilly Library Staff*—The online finding aid and access to the digital files would significantly ease the workload of Lilly Library staff in providing access to the collection.
- *Context*—Ample supplemental material is available to add context to the Web site. Text from secondary sources could be added to the site, or additional archival material, such as selections from Hohenberger's diaries and his newspaper column, contributing richness to the resource.

The LC/Ameritech proposal outline worked well for both the selection process and the preliminary planning. The Hohenberger Project proceeded as planned and has met or exceeded expectations with regard to user response.

EVALUATING THE HARVARD MODEL

The Harvard Model would have resulted in the DLP staff making the decision to digitize the Hohenberger Collection. The answers to the nine major questions in the model matrix were "yes." The more detailed questions posed in the essay would have produced more detailed information and required more careful thought.

The Harvard Model proved to be a valuable evaluative tool. However, the author encountered two problems in using it. First, *Selecting Research Collections for Digitization* (Hazen, Horrell, & Merril-Oldham, 1998) concludes with a one-page graphical summary of the process, "Selection for Digitization: A Decision-Making Matrix." This matrix presents the nine major questions and decision points, but it is impossible to align the issues and questions in the text of the essay with this graphical representation. The steps were in a different order and the questions in the matrix did not correspond to the questions in the text. The author

began working with the matrix but soon abandoned it for the questions in the text. Part of the difficulty may be that the matrix attempts to pose "yes" and "no" questions that would stop the project at any step. In reality, the questions are more likely to result in "maybe" rather than "yes" or "no" responses.

The second problem relates to the level of detail in the model. For many projects, the actual selection process would probably stop with an analysis of the content and finding aid issues. It would be necessary to proceed to the technical planning of a complex project in order to determine the cost-effectiveness of the proposed project. However, the Hohenberger Project did not require this level of analysis. The questions related to copyright, the intellectual nature of the source materials, current and potential users, and actual and anticipated nature of use would have provided project planners with enough information to make a decision. However, with a more complex project, selectors might have needed the additional input that only technical inquiry can provide. One recommendation might be to develop a two-tiered decision-making process, one for simpler projects that involve formats and technologies with which the institution has experience, and one for more complex projects that involve multiple formats and unfamiliar technologies. However, for any project, it is difficult to make a final decision without undertaking preliminary planning.

CONCLUSION

The vast number of worthy collections that should be considered for digitization require that digital library program planners establish criteria and procedures for selecting research collections for digitization. Each institution must develop its own criteria for selecting collections for digitization based on a standard set of criteria with adjustments and additions based on local needs. The Harvard Model provides an excellent foundation for creating a local adaptation. It is comprehensive, yet flexible. At Indiana University there may exist a need for a simpler version and a more complex version with a graphical flowchart representation. There may be decision points along the way, with a major decision point coming before preliminary technical planning and one coming afterward. A collection may pass the content test and fail the technical test. Additional selection criteria would emphasize Indiana history and culture; local technical expertise, such as the digitization of music; and local language expertise, such as Russian and Tibetan. By customizing an existing model for selecting research collections for digitization, institutions can maximize the probabilities of spending resources wisely, preserving valuable collections, and making them more accessible to users at home and throughout the world.

REFERENCES

Columbia University. (1998). *Selection criteria for digital imaging projects.* Retrieved January 31, 2000 from the World Wide Web: http://www.columbia.edu/cu/libraries/digital/criteria.htm.

Harvard University. (1998). *Selection for digitization.* Retrieved January 31, 2000 from the World Wide Web: http://preserve.harvard.edu/resources/digitization/selection.html.

Hazen, D.; Horrell, J.; & Merrill-Oldham, J. (1998). *Selecting research collections for digitization.* Washington, DC: Council on Library and Information Resources. Retrieved January 31, 2000 from the World Wide Web: http://www.clir.org/pubs/reports/hazen/pub74.html.

Lee, S. D. (1999). *Assessment criteria for digitization.* Oxford, England: Oxford University. Retrieved January 31, 2000 from the World Wide Web: http://www.bodley.ox.ac.uk/scoping/assessment.html.

Library of Congress. (1998). *The Library of Congress/Ameritech 1998/99 guidelines, National Digital Library competition.* Washington, DC: LC. Retrieved January 31, 2000 from the World Wide Web: http://memory.loc.gov/ammem/award/guide98.html.

Puglia, S. (1999). The costs of digital imaging projects. *RLG DigiNews, 3*(5). Retrieved January 31, 2000 from the World Wide Web: http://www.rlg.org/preserv/diginews/diginews3-5.html#feature.

University of California. (1997). *University of California selection criteria for digitization.* Retrieved January 31, 2000 from the World Wide Web: http://www.libnet.sh.cn/diglib/digselec.htm.