Nancy Kaiser. Diffusion of Innovations in Special Collections Libraries: The Motivations behind Adoption of Digitization. A Master's paper for the M.S. in L.S. degree. April, 2000. 39 pages. Advisor: Helen R. Tibbo.

This study explores the current rush to digitize in special collections libraries by examining case studies in the framework of Everett Rogers's diffusion of innovations theory. A content analysis of the case studies was performed to determine how each of Rogers's motivations—relative advantage, compatibility, trialability, observability, and complexity—shaped the context of the decision to digitize. Results suggest that digitizers are motivated by the goals of improved access and internet observability, and are not deterred by high equipment and staffing costs. Special collections libraries would benefit from pre-project assessments of the long term costs and commitments of digitization and its impact on institutional goals and the end user.

Headings:

Information systems – Special subjects – Virtual library Optical data processing Preservation of library materials – Automation Special collection

# DIFFUSION OF INNOVATIONS IN SPECIAL COLLECTIONS LIBRARIES: THE MOTIVATIONS BEHIND ADOPTION OF DIGITIZATION

by Nancy J. Kaiser

A Master's paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina

April, 2000

Approved by:

Advisor

# TABLE OF CONTENTS

Introduction	1
Literature Review	5
Methodology	13
Discussion	26
Conclusion	. 30
Appendix 1: Content Analysis Data Sources	32
Appendix 2: Content Analysis Data Collection Tool	. 35
Appendix 3: Digitization Project Planning Resources	36
Bibliography	. 37

# TABLE OF FIGURES

Figure 1: Motivations to Digitize	18
Figure 2: Access	20
Figure 3: Preservation	21
Figure 4: Compatibility Factors	23
Figure 5: Observability Factors	24
Figure 6: Complexity Factors	25

### Introduction

Few if any developments in the past century have captured the attention and pocketbooks of special collections caretakers as have digitization technologies during the 1990s. Given the monetary, staffing, and training costs associated with successful digitization projects, such interest and participation in this work is quite remarkable. As with most technologies, digitization offers several advantages, but these are accompanied by drawbacks as well. Among the attractions, digitization can: capture in electronic form materials that are too fragile for repeated handling by staff and patrons, protecting them from further damage in human hands; store millions of documents and images in virtual space, alleviating collection space shortages; expand the reach of an underutilized collection to a world wide audience; enhance classroom learning for students of every age; and, in some cases, serve as a preservation quality copy of the original (Sklar, 1998). The appeal of digitization is evident in the proliferation of digital imaging initiatives at special collections library web sites (McClung, 1996), museums, and archives.

While many of the features of digitization technology are desirable, disadvantages do exist. Special collections staff must stay current with the rapidly changing, unstable technology. Purchase and maintenance of new hardware and software can be very expensive in both direct and indirect costs. As technology evolves, older digitized formats may have to be migrated to newer formats so that the material is not lost to obsolescence. Each migration, in turn, may compromise the quality of the original digitized copy. Moreover, format migration may not always be possible, and in such

1

cases special collections staff must ensure that old and new equipment is kept in working order. Clearly a digital project is never a finished product for a special collections library (Lynn, 1998).

Careful guardianship would caution against jumping on the bandwagon of an unstable technology, but the reality is that there is a rush to digitize. For the typical budget-strapped special collections library, what motivates the adoption of this innovation? This is an important question because its answer affects the competition for labor and resources in the special collections library. While some digital projects initially are grant-funded, the subsequent migrations, hardware and software upgrades, and increased reference service responsibilities that accompany these projects may be competing for labor and resources with more traditional access services, such as the creation of MARC records and online finding aids, which arguably are more important for scholarly research than a sampling of digitized images from a collection.

Everett Rogers's theory of the diffusion of innovations provides a framework for examining why special collections may be moving into the digital world sooner and with more enthusiasm than expected given the extreme costs. Rogers describes the attributes of innovation as relative advantage, compatibility, trialability, observability, and complexity. Typically relative advantage, compatibility, trialability, and observability must be high and complexity low in order for an innovation to hold appeal for the innovation adopters (Rogers, 1995).

Relative advantage is "the degree to which an innovation is perceived as better than the idea it supercedes," and it is measured in terms of economics, social prestige, convenience, and/or satisfaction (Rogers, 15). To some extent, it represents the union of the four other motivations. In the case of digitization, access, preservation, and storage are all enhanced, if not improved, by this technology and thus provide a relative advantage over more traditional methods of caretaking in special collections; however, the relative advantage is tempered when economics and convenience are called into question. The cost and complexity of this technology and the accompanying labor can be prohibitive.

Compatibility is "the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters" (Rogers, 15). The race to commit scarce resources to an ever-changing technology does not seem compatible with traditional access and preservation strategies, but as more and more academic research libraries incorporate digital initiatives across their institutions, special collections are being asked to participate as well (Whaley, 1993). Compatibility thus is on the rise, although its presence still may be uneven from one library to the next.

Complexity is "the degree to which an innovation is perceived as difficult to understand and use" (Rogers, 16). There is no gray area here: complexity is almost always high on digitization projects because the technologies absolutely require acquisition of a knowledge base and skills previously not held by special collections staff. In addition, purchase of hardware and software frequently requires extensive consultation with systems administrators and vendors, and can be complicated further by the lack of standards for digitization of special collections materials.

Trialability is "the degree to which an innovation may be experimented with on a limited basis" (Rogers, 16). The digitization projects thus far seen among special collections never replace the originals, which makes them a low risk experiment for the

home institution. By this narrow definition trialability can be classified as an inducement to digitize. In terms of cost, however, trialability can be lower, especially if new equipment must be purchased and staff requires extensive training. Grant funding mitigates some of these expenses, but at the end of the project there are still costs to maintain the originals as well as the digital forms.

Lastly, observability is "the degree to which the results of an innovation are visible to others" (Rogers, 16). The availability of digital technologies in other sectors of the information marketplace has raised the expectations of users of special collections libraries as well. The need to please an audience may be the crux of the decision to digitize for most special collections libraries: in 2000 a special collection without a digital initiative accessible via the internet runs the risk of being labeled backward by its peers, patrons, and potential donors alike. Online observability has become too important for reputation among special collections to be underestimated in the role it plays in the adoption of digital innovations. Its importance may be compensating for questionable compatibility and relative advantage and decidedly high complexity in the digitization decision-making process.

Rogers's theory is a useful tool for assessing the motivations to digitize because it uncovers the pros and cons of this technology and encourages consideration of its impact on a collection, the staff that maintains it, and the end user. A review of some of the literature on digitization in special collections illustrates the challenges presented by this technology, and provides a guideline for understanding the context in which decisions to digitize are made.

#### **Literature Review**

Leaders in the field of digitization of special collections have written of their experiences gained at the helm of major projects, including Yale's Project Open Book and the many projects that fall under the Library of Congress's American Memory umbrella, or as members of best policy committees, such as the Council on Library and Information Resources. They share their insights to warn of the pitfalls likely to be encountered in digitization projects. Some of the recurring critical themes in the literature include the shortcomings of digitization as a preservation medium, the sticker price and hidden costs of specialized hardware, software, and dedicated labor, the importance of selecting materials for digitization that will have long term value, the dilemma of copyright protection for materials distributed on the web, and the role of the end in project design. These points are raised to give pause to smaller institutions that are considering adoption of digital technologies, but there is an underlying assumption that digitization is here to stay: cautious exploration rather than outright rejection is the approach of the day.

Paul Conway's "Digitizing Preservation" (1994) advises what, when, and how to digitize based on his experience at Yale's Project Open Book. The author emphasizes the importance of institutional support: digitization must be a long term commitment made by technical and administrative staff. The complexity issues that shape this commitment include expenses related to equipment acquisition, maintenance, and upgrade, subsequent migrations, and ongoing labor and training costs. Conway observes that while equipment expenses may decrease in price over the long run, the cost of human labor will not.

The author warns that the decision to invest in this technology is only the first of many hard decisions to come. To help anticipate problems and questions, Conway recommends hands-on exploration via pilot projects that bring together and solidify the technical and administrative infrastructure, and thereby cultivate compatibility between projects and institutional mission. Conway also highlights the importance of cooperative digital projects of groups such as the Digital Preservation Consortium, the Commission on Preservation and Access, and the Research Libraries Group. Sharing the tremendous costs across a network will help to reduce the risk of going digital. For this reason he suggests that smaller libraries may benefit from waiting for national standards for digital preservation systems to be set.

Conway's cautious approach emphasizes the trouble and expense of digital preservation, and the importance of long term commitment, but he does not question its place in the library, at least not in research libraries. The decision to digitize is a foregone conclusion.

E.C. Shoaf approaches digitization from the "preservation is access" perspective. That is, any effort to protect library materials is a step in the direction of "permanent" access. He points out in "Preservation and Digitization: Trends and Implications" (1996) the usual concerns: what types of technology are necessary to achieve a satisfactory level of preservation? How can the high costs be justified? Is optical character recognition a wise investment? Should SGML encoding be employed now or put off until later? Is integrity of the original more important than improved access? Shoaf contends that institutional commitment is the key factor in finding the answers to these questions. Shoaf reviews the major digitization projects to illustrate the "good news" of digital technologies. Lessons learned include the importance of getting standards in place so that others can follow; the slowness of institutions to meet the expense of installing and maintaining preservation systems; and the importance of consortiums to share the burden of funding projects. Once again, the decision to digitize is a foregone conclusion; at no point does the author consider that digitization is not always the correct answer.

R.W. Steward's "Does this project deserve the Erasmus Prize?" (1993), on the contrary, criticizes the awarding of the prestigious honor to the Archivo General de Indias for its digitization of New World discovery materials.<sup>1</sup> Steward argues that this technology is excessive because it does not serve the needs of the archive's principal users. He contends that rigorous (i.e. speedy) retrieval is unnecessary in archival settings; the format does not accommodate how real researchers actually use documents in archives, and furthermore, low resolution scanning adversely affects the reader's ability to read material once it is printed out; and lastly, surrogate copies are unnecessary because two patrons rarely want the same thing at the same time. Steward also points out that preservation can not be guaranteed with unstable mediums like disks, hardware, and software; moreover, no standards are in place to guide creation of permanent images. Much of the technology has changed since Steward wrote in 1993, but his criticisms are still valid in 2000: preservation still can not be guaranteed; and the (scholarly) end user's needs must play some role in decisions of what, how, and when to digitize.

<sup>&</sup>lt;sup>1</sup> The Archivo General de Indias has produced a cd-rom of text and images that is viewable at the archive in Seville, Spain, and at the Huntington Library in California. A write-up of the project is used as a data source in the content analysis section of this paper.

Abby Smith's "Why Digitize" (1999) is not quite as critical as Steward is; nonetheless she acknowledges that librarians have embraced digital technology without anticipating its impact on the rest of the library. She argues for evaluation of the costs and benefits of digitization prior to embarking on a project because "digitization often raises expectations of benefits, cost reductions, and efficiencies that can be illusory and, if not viewed realistically, have the potential to put at risk the collections and services libraries have provided for decades" (Smith, iv). Smith is particularly concerned by problems associated with preservation, access, and selection.

Smith sets the cautious tone of her article by reminding readers that preservation goals are not met by digitization: digital files are not eye-readable; and they require hardware and software that may become obsolete in short time or degrade over a longer period of time, and consequently may require frequent reformatting. Preservation-minded digitizers also must be concerned with loss of analog character in the initial reformatting because faithfulness to the original will be compromised. Further corruptions may occur in subsequent data migrations.

The author is more supportive of digitization that focuses on improving access. Surrogate copies often will satisfy researchers who need access to the intellectual content rather than the artifact itself. Digitization will provide these types of researchers with enhanced access at the item and collection level: hard-to-see images, for example, can be manipulated for closer examination, while dispersed items among various archives can be compared side by side on remote computer screens. Smith urges digital libraries to work collaboratively to create complementary rather than duplicative collections (Smith, 8-9). Beyond informational value, digitized copies add benefit by reducing handling of the original. In addition, digital forms that are part of a hybrid production of preservationstandard microfilm copies can fulfill dual access and preservation goals.

Smith's enthusiasm for access possibilities is qualified by concerns for selection strategies and implementation. She is quick to point out that mere potential for increased access does not justify digitization of an underutilized collection. Value must be added to the product by incorporating one or more of the following tools: enhanced intellectual control (improved access to finding aids, indexes, bibliographic records); increased searchability via image and text manipulation; improved quality of images; or virtual collections that bring widely dispersed items together. Smith insists that the allure of increased access to an underused collection must be weighed against the difficult evaluations of the collection's intellectual value and compatibility with other like digital resources.

Deciding what to digitize may not even be the hardest decision to make in the digitization process. Smith notes that the physical preparation of materials and intellectual labor involved in selecting materials for digitization can be significant because collections are rarely cataloged to item level. Once selected, targeted items may require conservation treatment and bibliographic research before they can be scanned. Linking bibliographic information to a digital object is especially important in the internet environment because users expect a higher level of functionality in the absence of immediate access to a reference specialist (Smith, 10). Smith, in fact, factors the end user into the decision to digitize. She calls for studies to assess how digitized collections are used, if they really are in the public interest, and if they do enhance education.

Despite her concerns for digitization in special collections, Smith encourages "a hopefulness tempered by patience and informed by experience" (Smith, iv). She recommends that digital technology should be thought of as just one tool in the tool kit rather than a replacement for tried and true techniques of preservation and access (Smith, 13).

Stuart D. Lee's *Assessment Criteria for Digitization* (1999) contends that assessment is the most important part of a digitization project because it helps define needs and determine feasibility before committing to a potentially lengthy, expensive, and demanding undertaking. The author recommends a study of how the proposed digital project fits into the home institution's agenda, and a study of how the collections are to be digitized. A review of digital projects at other institutions (principally their selection criteria) may provide guidance.

Determining institutional feasibility requires honest responses to the following questions: can the project be completed in a reasonable amount of time? Does budgetary and technical support exist to carry the project to completion? Will the project compete with other library priorities for limited resources? Can the project be completed to a satisfactory standard, i.e., are skilled staff and the appropriate technology available? Will the project be cost effective in the long run?

An assessment study also will help to define the need to digitize. Is the project needed because it will make content more widely available and add informational value to the collection, and thereby increase demand for the collection? Is it intended to fulfill preservation and/or conservation goals? Are institutional prestige and/or the attraction of grant funding needs of the library, or is it simply an extension of current programs?

Assessment of the collection targeted for digitization should include evaluation of current and potential formats. Would this collection be better served in microform? Do the remote access and search and analysis features outweigh the instability of the digital format? Decisions must be made about what to include: must the whole collection be digitized, or is it acceptable to "cherry-pick" for key items from multiple collections, and if cherry-picking is acceptable, will selection be based on items of high research interest, high demand, or in need of preservation or conservation treatment? Whatever the answers, assessment should point to items that are unique and in clear need of improved access. The author emphasizes that there are "no perfect answers" to the assessment questions digitizers must ask, but answers nonetheless are essential before proceeding.

Dan Hazen's *Selecting Research Collections for Digitization* (1998) also addresses the selection process, but he counsels beginning with copyright because so few source materials (relative to the holdings of the entire archive) can be reproduced in digital format. If the source materials are in the public domain, the items can be reproduced; if the source materials are copyright protected but permissions can be secured, the items can be reproduced; if permissions for copyrighted material can not be secured, the items can not be reproduced and the project may have to be redesigned (Hazen, 2). To complicate matters, copyright law can vary by country and by material type, and can change over time. The country in which the project is based and the country in which the source materials were created may have conflicting laws. In terms of material, different kinds of sources, such as recordings, sheet music, or photographs, may be governed by different rules. In short, copyright should not be underestimated as it can be as complex as the technical specifications of image capture. Stephen Chapman's and Anne R. Kenney's "Digital Conversion of Research Library Materials: A Case for Full Informational Capture" (1996) argues that the economic viability of digital conversion depends upon long term utility of resources; therefore, decisions should be made with long term functionality in mind.

The authors make the source documents, not the current end user, institutional objective, or technical capabilities, the focal point of the conversion process because it is the entity least likely to change. Chapman and Kenney advocate full information capture approach that identifies but does not exceed the scan settings necessary to capture all significant information in the source document. Attributes to be assessed in making scanning decisions include (but are not limited to) document size, text size, text characteristics, medium, illustrations, tones, and density and contrast. The value of a digital surrogate can be measured by how well it reflects the meaningful features of the source document.

Though this approach may be more expensive and time consuming, the authors contend that libraries will benefit in the long run from full information capture. Creation of a digital master from which multiple formats can be derived will be compatible with a range of computing capabilities, user needs, and institutional service missions, all of which are sure to evolve in a dynamic environment.

The literature reviewed above is important to understanding special collections and digitization because it outlines some of the challenges posed by adoption of this innovation. Even though the underlying assumption is that digitization is here to stay, the authors make clear that digitization is not a commitment to be taken lightly because of the sizeable investment of resources required. Judging by the proliferation of case study literature that describes digital initiatives at individual special collections libraries, low relative advantage and high complexity are not deterrents to adoption of digitization. What then is driving the rush to digitize? Much of the literature describes initiatives in terms of the successful creation of an observable, useable product that improves on the original, but does not necessarily address the context in which decisions to digitize were made. This study aims to assess motivations to digitize by examining case studies for articulated goals, problems, and achievements. Rogers's theory of the diffusion of innovations provides the backdrop for analysis because it is a standard assessment tool that can be used to help guide decisions before making major commitments. This approach may encourage special collections libraries to balance perceived expectations for a digital initiative against the realities of the process and its impact on institutional services.

#### Methodology

#### Data Collection

A content analysis was performed on a sampling of the case study literature that describes specific examples of academic and public library and research consortium special collection digitization projects. Library Literature (NCLIVE: OCLC FirstSearch) and Library and Information Science Abstracts cd-rom (available at the School of Information and Library Science library) were searched in January 2000 for articles that featured discussion and analyses of digitization projects that involved manuscript collections and/or archival materials. Searching on combinations of keywords, including (but not limited to) digital imaging, digitization, optical data processing, special collections, archives, and manuscripts, retrieved dozens of articles that addressed digitization from both theoretical and case study perspectives. Twenty-six case study articles were selected for content analysis based on immediate availability online or in the School of Information and Library Science library. The author did not limit the data set by currency in order that the full range of decision-making contexts would be represented in the study, from the first initiatives of the early 1990s to the proliferation of projects at the present. (See Appendix 1 for a complete list of selected articles.)

Content analysis was selected as a research method because it allowed the author to examine existing digitization projects, as described by their creators, for evidence of the motivations behind this innovation at the level of individual libraries, and to assess digitization trends in the aggregate. Notably, criticism of digitization is virtually nonexistent in the case studies. While some authors do report problems encountered with digitization, the end result is an endorsement of this technology. Readers will not encounter any case studies in which decisions to digitize are regretted. Given that the projects did not all enjoy the same degree of success, the lack of self-criticism perhaps may be attributable to a concern for the effect of negative publicity on a library's reputation.

Each case study article was coded by the author for the five motivations discussed by Rogers: relative advantage, compatibility, trialability, observability, and complexity. Because the authors in the dataset did not necessarily use Rogers's terminology, the author constructed a series of general and specific assessment questions to be applied to each article to determine the explicit and implicit existence or non-existence of each motivation in the project. General questions were designed to assess explicit discussion of motivations. Each motivation was assigned one general question that asked for a broad assessment of the level of that motivation in the project. Specific questions were designed to uncover implicit discussion of motivations. Each motivation was assigned up to four specific questions that provide examples of what might drive the motivation. The content of the general questions reflects the core terms of Rogers's diffusion of innovations theory. The content of the specific questions reflect the warnings and "best practice" suggestions from the literature review section.

The author performed all of the coding. To improve coder reliability the author eliminated overlap among specific questions in the motivation categories, and limited the range of scoring responses for all questions to two choices (Sommer, 1997). The five general questions required a "high" or "low" response; the remaining specific questions esponse.

The questions were as follows:

1. Was the relative advantage high or low? Was the project designed to improve access? If yes, was access improved by image or bibliographic enhancements? Was the project designed to meet preservation needs? If yes, was the goal permanent storage or creation of a surrogate copy? As the majority, if not all, of case studies are suggestive of some degree of success, the author decided that relative advantage could not be judged on the basis of these answers alone because all projects would appear to have a high relative advantage, regardless of the answers to subsequent questions. Because relative advantage can be interpreted as the average or sum of the other motivations, the rating was determined after review of the ratings of the other categories. Projects with "high"

ratings for compatibility, trialability, and observability were rated as "high" for relative advantage, regardless of the complexity rating. Projects with two "high" ratings from the set of compatibility, trialability, and observability (and without regard for the complexity rating) also were given a "high" relative advantage rating as a means of counterbalancing potential coder bias against relative advantage. This coding schema gives the benefit of the doubt to digitization decisions that have balanced pros and cons. Projects that received a "low" rating for relative advantage had two or more "low" ratings from the set of compatibility, trialability, and observability and a "high" rating for complexity.

- 2. Was compatibility high or low? Did the project involve collaboration with external institutions or agencies? Was the project supported by external funding? Is the project designed to work with demonstrated needs of specific users? If compatibility was not discussed explicitly, answering "yes" to one or more of the specific questions earned a "high" rating, and answering "no" to all of the
- 3. Was the trialability high or low? All digitization projects reviewed received a "high" rating because in no cases were the originals discarded after digitization. Trialability might also reflect an institution's level of commitment of resources to a project: did the digital initiative entail a pilot project, or did it commit immediately to a full scale project? Did grant funding contribute to trialability by reducing the direct costs carried by the institution itself? The author restricted data collection to the first definition because the presence or absence of a pilot project could not be determined in many of the articles coded, and lack of mention did

not necessarily mean that a pilot project had not been done. In addition, the author found assessment of the role of grant funding in these case studies to be uncertain based on the information provided. The case studies do not necessarily discuss how much funding came from grants, or what percentage of the project was funded by grant money; therefore, it would be impossible to assess objectively whether or not a grant contributed significantly to trialability.

- 4. Was observability high or low? Was the goal of the project to make the collection available on the internet? To be a leading user of cutting edge technology? To secure external funding? If observability was not discussed explicitly, answering "yes" to one or more of these questions earned a "high" rating, and answering "no" to all of the questions earned a "low" rating.
- 5. Was complexity high or low? Did the lack of standards contribute to complexity? Did equipment purchase and operation contribute to complexity? Did the need for staff training contribute to complexity? If complexity was not discussed explicitly, answering "yes" to one or more of these questions earned a "high" rating, and answering "no" to all of the questions earned a "low" rating.

Answers were recorded for every question. In the absence of a clear indication of the library's position, the coder assigned a favorable value. For example, if the question "Did the lack of standards contribute to complexity?" was not immediately discernible because the article did not discuss standards at all, the coder assigned the more favorable answer ("no" in this particular case, given Rogers's assumption that complexity must be low). This strategy was designed to compensate for potential coder bias by giving the benefit of the doubt to the library's decision.

All answers were entered into SPSS. Responses of "high" or "yes" were entered as "1." Responses of "low" or "no" were entered as "2." (See Appendix 2 for data recording tool.)

## Data Analysis

Because the dataset is small and not evenly distributed across the years of publication, only descriptive analyses—frequencies and cross tabulations—were performed to assess the general trends in digitization in special collections libraries. The following analyses were performed in SPSS:

1. What is the frequency of high/low ratings for each of the five motivations for innovation?



Figure 1 shows that 65.4 percent of the articles suggested high relative advantage; 73.1 percent self-described high compatibility; 100 percent were assigned high trialability; 96.2 percent self-described high observability; and 80.8 percent selfdescribed high complexity. These results suggest a few possible trends. First, a low relative advantage is not necessarily a deterrent to digitization. Keeping in mind that relative advantage is an average or sum of the other motivations, it should be surprising that roughly one third of the institutions adopted this innovation despite compatibility and complexity concerns. Certainly these institutions did not jump on the bandwagon of technology without concern for the future. Analyses 2, 3, and 7 explore the attractions of digital initiatives.

The second interesting trend in Analysis 1 is the high rate of observability. In some cases, the desire to be observable may outrank low relative advantage and compatibility and high complexity in the decision to digitize. Excluding trialability (in its narrowest definition), observability appears to be the greatest inducement to digitize at this time. The perceived benefits of observability include increased collection usage, publicity for other worthy collections, institutional prestige, and attractiveness to potential funding sources (Lee, 1999). Analysis 7 further explores observability as a motivation to digitize.

Lastly, as expected, complexity is high in many of the projects. The technology is still so new and the training is not yet standard in most library schools and workplaces. Only the institutions with the experience of one or more digital projects behind them, like the Bodleian, can cite low complexity in the decision to digitize. Analysis 8 explores some of the sources of complexity reported by the institutions.

An additional analysis of complexity ratings over time would be useful to determine if complexity is in fact decreasing. The dataset is too small to determine this, but it is hypothesized that high ratings will remain constant because of both the complex nature of unique historical materials, and the forward march of technology—"new and improved" innovations are introduced continuously and allow little time for grasping current technology before it becomes obsolete.

2. How frequently was improved access cited as a motivation for digitization? Was improved access achieved via bibliographic enhancement or image enhancement or both?



Figure 2 shows improved access was cited as a motivation in 100 percent of the cases reviewed. Bibliographic enhancements, like indexing, SGML encoding, and improved finding aids, were cited in 84.6 percent of the cases reviewed. Image enhancement was cited 42.3 percent of the time. Image enhancements and bibliographic enhancements were twin goals in 38.4 percent of the case studies.

The global citing of improved access as a motivation to digitize illustrates the "hopefulness" Abby Smith described in "Why Digitize" (1999), but it is not "tempered by experience." Despite the costs and complexity, faith in digitization's ability to

(and ergo effect scholarship?), remains strong. Certainly digital access is increased and different, but until studies demonstrate how scholars have implemented digital access tools it is difficult to judge how "improved" access really is.

The popularity of bibliographic enhancement was somewhat surprising because it often involves an increased amount of item level research and staff expertise over image capture alone. Enhancements introduce powerful searching capability to the intellectual content of source material, but again, whether or not end users find this capability desirable and use it is not clear. Smith's call for end user studies underscores the importance of matching the project specifications to how the end project will be used by patrons.

3. How frequently was preservation cited as a motivation for digitization? Was permanent storage or the creation of surrogate copies cited as the preservation goal?



Figure 3 shows that preservation was cited as a motivation in 65.4 percent of the cases reviewed. Not surprisingly, permanent storage was never a goal in any of the cases. The instability of the format makes permanent storage a risky proposition for caretakers of unique materials. Instead, the creation of surrogate copies was the goal in 65.4 percent of the cases. These results suggest that special collections caretakers have taken a conservative approach to preservation in the digital environment: surrogate copies that reduce wear and tear on the original will suffice for the present.

4. How frequently were digital projects initiated with the demonstrated needs of a specific group of users in mind?

Figure 4 shows that a demonstrated need was self-described in only 30.8 percent of the cases reviewed. The case studies that comprise this figure generally involved projects that brought together widely dispersed items in extremely frail condition, like the University of Michigan papyri database, or were created for specific educational purposes, like the Aberdeen Bestiary.

A low rating for demonstrated need suggests that special collections libraries may be developing projects without a clear conception of who the users will be. The notion of "if we build it they will come" is a dangerous assumption because potential users may not ever discover this great digitized resource on their own.

5. How frequently did digitization projects involve collaboration with outside institutions or agencies?

Figure 4 shows that Digitization projects involved collaboration with external institutions or agencies in 42.3 percent of the cases studied. The results suggest that collaboration is important but not necessarily a critical component of a digital initiative. An analysis of collaboration over time may identify an increasing amount of consortium activity as institutional strategies evolve from single collection experimentation to commitments to digitize critical masses of materials. Special collection libraries may want to participate in and develop similar consortiums that aim to build subject-based collections. Their subject strengths, for example, may benefit from virtual association with other like materials, thereby increasing observability for all involved. Projects like

*Studies in Scarlet*, a multi-institutional effort organized by the Research Libraries Group, are leading the way.<sup>2</sup>

6. How frequently were digitization projects supported by external funding agencies?

Figure 4 shows that external funding was reported in 50 percent of the cases. This result suggests that libraries recognize digitization as an essential component of collection development, and are funding their own projects. This is a good sign for compatibility with management objectives and the future of digital resources in an institution. It would be interesting to see if a study of funding over time would show that library administrators increasingly funded their own projects, or if special collections increasingly sought to expand digitizing initiatives beyond the current institutional scope by seeking external funding. If institutions are to create large digital libraries, like that of Yale Project Open Book or the University of North Carolina's Documenting the American South, they likely will have to look to external funding and consortiums to help support the level of commitment required.



<sup>&</sup>lt;sup>2</sup> Studies in Scarlet is included in the data set.

7. How frequently was the goal to be observable on the internet? To be observable in the use of cutting edge technology? To be observable in order to secure additional grant funding?



Figure 5 shows that the goal of internet accessibility was self-reported in 92.3 percent of the cases; the goal of being an innovator of cutting edge technology was self-reported in 26.9 percent of the cases; and the attraction of external funding was a self-reported goal in 26.9 percent of the cases.

The high rate of internet accessibility was both predictable and eye-raising. Certainly one of digitization's advantages is its compatibility with the internet environment, but at the same time the certainty among special collections caretakers that their materials need to be accessible worldwide is surprising. One explanation for the importance of internet observability is concern for special collection reputation. Special collection libraries do not want to be perceived as laggards, especially as user expectations for electronic resources are on the rise (Velgos, 1999). What libraries might rethink is what type of internet observability best serves their mission: is it more important to have finding aids or a digital image collection? The answer depends in part on whom the library defines as its primary patrons. Some special collection libraries are attempting to broaden their outreach beyond their traditional patronage—the scholarly community—to the general

public because they have been mandated by their funding sources to serve a broader constituency (Alden, 1996). In this case, digital images may better serve the patron community.

The relatively low rate of interest in being a leading user of cutting edge technology is ironic: by having digital projects special collections libraries in fact are on the cutting edge. The low rate of self-reporting technology leadership may be attributable to the historical mission of archives and manuscript collections. These places by definition preserve the past, not chart the course of the future.

The relatively low rate of interest in attracting external funding may be attributable to current availability of institutional funding, or perhaps reflects no immediate desire to initiate new digitization projects beyond the one described in the case study. One case study cited the amount of time devoted to grant-writing over and above assigned duties as a major drawback to the application process (Theyer, 1999).

8. How frequently was a lack of standards cited as a source of complexity? How frequently were hardware and software cited as sources of complexity? How frequently was the need to train staff/reallocate staff time cited as a source of complexity?



Figure 6 shows that a lack of standards was cited as a source of complexity in 19.2 percent of the cases. Purchase and operation of hardware and software were cited as sources of complexity in 61.5 percent of the cases. Developing a skill and knowledge base in staff was cited as a source of complexity in 57.7 percent of the cases.

The lack of standards rate benefits from the absence of explicit mention in the articles examined. Most articles in fact did not discuss standards at all. In these cases it was assumed that the lack of standards did not contribute to complexity. The lack of discussion in many of the articles by first time digitizers may reflect a belief that every project is different, requires unique treatment, and necessitates a slow learning curve by the innovators. Experienced digitizers, like the Heinz Library, were more likely to call for standards (Galloway, 1995, 1999).

Equipment and staff accounted for complexity in roughly two thirds of the cases examined, many of which described first time digitization projects. Failure to anticipate these problems may explain why this is not a deterrent to the decision to digitize; or perhaps these problems are assumed and consequently dismissed as unavoidable. Regardless, labor may become less of a problem in the long run, especially if special collections libraries are able to hire dedicated digitization staff, but the march of technology likely will prevent a simplification of equipment concerns.

#### Discussion

Looking at the results of the case study analyses in the context of the background literature suggests that decisions to digitize are not always based on careful pre-project assessments of start to finish commitments of time and resources. The decisions also may not reflect a clear understanding of what impact a project will have on users. Nonetheless, the relative advantage is perceived as high. Institutions seem to be basing this perception on high expectations for achievement of access and observability goals. Long term institutional costs may be underestimated in the glare of the aforementioned attractions. Blindness to or disregard for the realities of relative advantage will not reduce the risks associated with digitization. Instead, working actively to shape the decision to digitize with high ratings for compatibility, trialability, and observability and a low rating for complexity will improve the overall relative advantage by increasing the likelihood of a sound investment. Digitization will continue to be expensive in terms of time, money, and skill development for the foreseeable future, but the context in which digitization decisions are made still allows room for improvement in each of the motivation categories that contribute to relative advantage.

Compatibility can be improved by creating digital projects that are in consonance with institutional objectives, services, and budgets. First and foremost, emphasis should be placed on pre-project assessment of goals, expectations, and realities. Careful planning can compensate for some of the disincentives to digitize by defining institutional feasibility up front. Infeasible projects can be modified, scaled back if necessary, to improve the overall relative advantage.

Compatibility might also be improved by factoring end user needs into the selection process. The experts are mixed on what role end user viewing capability should play in technical decisions, but there is consensus on the need for selection of collections that respond to user needs (Chapman and Kenney, 1996; Hazen, 1998; Smith, 1999). Of course some items will be digitized for preservation reasons, but the majority will be

selected for content reasons. To avoid the dangerous assumption of "if we build it they will come," librarians might consider proactive "marketing" of a worthy digital collection to potential users. A specific group of users, be it scholarly researchers, graduate students, genealogists, teachers of undergraduates, high school or elementary school students, or all of the aforementioned, should be identified and encouraged to use the resources. End user assistance may entail construction of help pages to accompany digital image projects (Hazen, 1998). Certainly library staff should anticipate an increase in electronic reference (Tibbo, 1995). Post processing needs like these will increase the cost of the project, but signs of successful achievement of project goals might be worth the price.

Trialability, as defined for this content analysis study, can not be improved upon; however, if the definition is broadened to include level of commitment by an institution (see *Data Collection* question 3), then first time digitizers in special collections libraries ought to consider employing pilot projects to work out the pros and cons of digitization on a small scale (Conway, 1994).

In terms of observability, special collections librarians might reevaluate what impact it really has on the library. How much does internet observability affect daily operations? How much does it enhance prestige? While users increasingly are demanding electronic access to special collection holdings, this should not necessarily be the principle motivation in the decision to digitize. Creation of MARC records and online finding aids are still more important access tools for research than digitized collections (Library and Information Commission, 1998). Complexity may be resolved in part by learning from the mistakes and success stories of digitizers that have gone before: there's no need to reinvent the wheel. An additional solution might be to outsource to vendors the work that requires expensive specialized equipment and/or technical expertise (Gartner, 1997; Hazen, 1998; Research Libraries Group, 1998). Libraries will still pay significantly for the work but they will avoid having to hire dedicated staff, or to divert time of current staff from regular responsibilities to new projects that require extensive training and development. This, however, does not eliminate the planning and selection work or the need for post-processing by library staff. Materials processed by vendor technicians will have to be proofread and evaluated by library staff to make sure that they are consistent with quality standards and with archival methodology.

Finally, special collections libraries should be willing to use Rogers's motivations framework to make hard decisions, including the decision not to digitize, when the attractions are great but the costs outweigh the benefits. Dan Hazen reminds, "the limited resources available for digitization might have greater impact if they were directed at another project, or directed toward an entirely different approach to providing access..." (17). Technology should serve larger collection-related goals, not undermine them.

The above suggestions represent a few responses to the digitization concerns suggested by the content analysis study. More responses might be developed by expanding the study to include a larger data sample that permits studies over time and correlation studies, and that asks additional specific questions. The following questions should be addressed:

- 1. level and source of funding in special collections;
- 2. technical specifications;
- 3. how and why a specific group of materials was selected for digitization;
- 4. types of bibliographic enhancements (e.g. SGML, indexing, etc.) applied;
- 5. existence of other digital initiatives at the institution;
- 6. popularity of multi-institutional subject-based collections;
- 7. formal pre-project assessments;
- 8. pilot studies;
- 9. outsourcing to vendors;
- 10. post processing costs.

One drawback to asking more specific questions, however, is that the researcher may have to guess at more of the answers. Case studies will not uniformly provide answers for all of the questions asked, in which case, additional coders may be required to establish reliability. As an alternative, a researcher might consider combining the case study content analysis with survey or interview data collection techniques to acquire additional data about specific projects.

#### Conclusions

The motivations to adopt digitization in special collections libraries thus far have centered around the anticipation of successful achievement of improved access and internet observability. Given that achievement of these goals in most cases will not justify the extreme costs of equipment and staffing, it is important to evaluate where and how digitization technologies do not meet Rogers's criteria for the diffusion of innovations in order that the project might be reshaped for cost effectiveness. Examination of the case study literature at the individual and aggregate level reveals the importance of cultivating attractive ratings for all of the motivations: successful digital resource development is difficult to achieve without attainment of high relative advantage, which implies high compatibility, trialability, and observability, and well-managed if not low complexity. If these motivation ratings do not shape the institutional context of the decision to digitize, special collections libraries should consider alternatives to digitization.

#### **Appendix 1: Content Analysis Data Sources**

 Galloway, E. and Michalek, G. (1998). The Heinz electronic library interactive online system (HELIOS): an update. *Public-Access Computer Systems Review*, 9(1) [On-line]. Retrieved January 23, 2000: http://info.lib.uh.edu/pr/v9/n1/gall9n1.html

Galloway. and Michalek, G. (1995). The Heinz electronic library interactive on-line system (HELIOS): building a digital archive using imaging, OCR, and natural language processing technologies. *Public-Access Computer Systems Review*, 6(4) [On-line]. Retrieved January 23, 2000: http://info.lib.uh.edu/pr/v6/n4/gall6n4.html

- 2. Nelson, P. (1996). Digitizing archival photographs for access and preservation at Thomas Jefferson University: the JEDI project. *Microform & Imaging Review*, 25, 99-106.
- 3. Velgos, J. (1999). One bit at a time: bringing the state archives to the internet. *Texas Library Journal*, 75, 18-19.
- 4. Zidar, J. (1992). Optical scanning and text recognition: operating an in-house system. *Quarterly Bulletin of the International Association of Agricultural Information Specialists*, 37(1-2), 65-69.
- 5. Patkus, R. (1999). Creating a digital archive of the pre-Vatican II church. *Catholic Library World*, 69(3), 9-12.
- 6. Threatt, P. (1998). Someday, someone will hunt history in these very old notices: Cammie G. Henry Research Center. *LLA Bulletin*, *60*, 120-125.
- 7. Preservation project turns tattered papers into digitized documents. (1995). *Wilson Library Bulletin, 69*(May), 14.
- 8. Theyer, H. (1999). Planning the future of history: making a digital historical resource. *Computers in Libraries 19*(9), 16-18.
- Research Libraries Group. (1998, October). Studies in scarlet. Research Libraries Group Digital Collections Project. Retrieved March 20, 2000: http://www.rlg.org/rlgnews/news40.html
- 10. Krishnamurthy, R. and Mead, C. (1995). An overview of the project on the imaging and full-text retrieval of the Ava Helen and Linus Pauling papers at Oregon State University Libraries. *Microform Review*, 24(Winter), 12-15.

- 11. Gagos, T., Zachary, S., Loftus, A., & Daub, P. Scanning the past: a modern approach to ancient culture. *Library Hi Tech*, *14*(1), 11-22.
- 12. Commings, K. (1995). Local history comes alive in West Virginia. *Computers in Libraries*, 15(June), 20.
- 13. Alden, S. (1996). Digital imaging on a shoestring: a primer for librarians. *Information Technology and Libraries*, 15, 247-250.
- 14. Flanders, B. (1992). Optical imaging project links Spain to southern California. *Computers in Libraries, 12*(October), 27-28

Kaebnick, G. (1992). Rediscovering the age of discovery. *Inform*, 6(July/August), 36-38.

- 15. Fray, W. & Spar, L. (1998). The avalon project. The Electronic Library, 16, 243-244.
- 16. Fineberg, G. (1995). Whitman on the Web: four recovered notebooks to be digitized. *Library of Congress Information Bulletin, 54*, 139-144.
- 17. Chantiny, M. (1993). Incorporating digitized images in the UHCARL PAC online catalog. *Library Software Review*, *12*(Spring), 22-26
- 18. Gartner, R. (1997). Digitising the Bodleian revisited: linking word and image. *Audiovisual Librarian*, 23, 47-50.

Gartner, R. (1997). Conservation by numbers three years on: an update on digital imaging at Oxford. *Microform & Imaging Review*, 26, 147-151.

- 19. Beavan, I. & Arnott, M. (1997). The nature of the beast; or, the Aberdeen Bestiary on the World Wide Webb. *Library Hi Tech*, *15*(3-4), 50-55.
- 20. Kiernan, K. (1995). The Electronic BEOWULF. *Computers in Libraries*, 15(2), 14-15.
- Hopkin, D. (1997). Shifting the focus: digital imaging and photographic collections management at the National Railway Museum. *Microform & Imaging Review*, 26, 67-72.
- 22. Psohlavec, S. (1998). Digitization of manuscripts in the National Library of the Czech Republic. *Microform & Imaging Review*, 27, 21-26.
- 23. Ventress, A. (1996). State Library of New South Wales digitization projects. *Microform & Imaging Review*, 25, 150-155.

- 24. Whitfield, S. (1997). The International Dunhuang Project: a challenge for digitization. *Microform & Imaging Review*, 26, 15-21.
- 25. Bjornson, P. (1999). The early Canadiana online Project. Feliciter, 45, 152-158.
- 26. Herbener, E. & Stokes, B. (1993). The past visible. Inform, 7(October), 38-40.

	Relative Advantage	Access Enhancement	Image Enhancement	Searchability	Preservation	Permanent Storage	Surrogate Copy	Compatibility	User Need	Collaboration	Trialability	External Funding	Observability	Internet	Technology	Grant Funding	Complexity	Lack of Standards	Equipment	Staff Training
1																				
2																				
3																				

# **Appendix 3: Digitization Project Planning Resources**

## Assessment

- Studies in Scarlet: Decisions and Information Needed to Develop Project Plans and Cost Estimates: http://www.rlg.org/scarlet/prep.html
- 2. Studies in Scarlet: Technical Checklist: http://www.rlg.org/scarlet/tech.html

# Selection Criteria for Digitization Collection Development

- 1. Columbia University: http://www.columbia.edu/cu/libraries/digital/criteria.htm
- 2. Harvard University: http://preserve.harvard.edu/resources/digitization/selection.html
- 3. University of California: http://www.library.ucsb.edu/ucppp/digselec.html

# Copyright

1. IFLA Copyright and Intellectual Property Resources: http://www.ifla.org/II/cpyright.htm

#### **Bibliography**

- Arms, C. (1996, April). Historical collections for the National Digital Library: lessons and challenges at the Library of Congress (part 1). *D-Lib Magazine* [On-line].
  Retrieved March 20, 2000: http://www.dlib.org/dlib/april96/loc/04c-arms.html
- (1996, May). Historical collections for the National Digital Library: lessons and challenges at the Library of Congress (part 2). *D-Lib Magazine* [On-line]. Retrieved March 20, 2000: http://www.dlib.org/dlib/may96/loc/05c-arms.html
- Chapman, S. & Kenney, A. (1996, October). Digital conversion of research library materials: A case for full information capture. *D-Lib Magazine* [On-line].Retrieved March 20, 2000:

http://www.dlib.org/dlib/october96/cornell/10chapman.html

Conway, P. (1994). Digitizing preservation. *Library Journal*, 119, 42-45.

- Erway, R. (1998). Options for digitizing visual materials. *Collection Management*, 22(3-4), 123-132.
- Gartner, R. (1997). Digitizing the Bodleian revisited: Linking word and image. *Audiovisual Librarian*, 23, 47-50.
- Gwinn, N. (1998). Mapping the intersection of selection and funding. *Collection Management*, 22(3-4), 143-149.
- Hazen, D., Horrell, J. & Merill-Oldham, J. (1998). Selecting research collections for digitization. Washington DC: Council on Library and Information Resources.

- Lee, S. (1999, January). Scoping the future of the University of Oxford's digital library collections: Appendix G: Assessment criteria for digitization [On-line]. Oxford, England: University of Oxford. Retrieved March 20, 2000: http://www.bodley.ox.ac.uk/scoping/assessment.html
- Library and Information Commission. (1998, July 14). Virtually new—Creating the digital collection: A review of digitisation projects in local authorities, libraries, and archives [On-line]. Retrieved April 10, 2000:

http://www.ukoln.ac.uk/services/lic/digitization/intro.html

- Lynn, M. (1998). Digital preservation and access: liberals and conservatives. *Collection Management*, 22(3-4), 55-63.
- McClung, P. (1996). Digital collections inventory report. *Microform & Imaging Review*, 25, 110-125.
- Ogden, B. (1998). The preservation perspective. *Collection Management*, 22(3-4), 213-216.
- Research Libraries Group. (1998, October). *Studies in scarlet: Starting points* [On-line]. Research Libraries Group Digital Collections Project. Retrieved March 20, 2000: http://www.rlg.org/scarlet/sis.html

Rogers, E. (1995). Diffusion of innovations. (4th ed.). New York: Free Press.

- Shoaf, E. (1996). Preservation and digitization: Trends and implications. *Advances in Librarianship*, 20, 223-239.
- Sklar, H. (1998). Why make images available online: User perspectives. Collection Management, 22(3-4), 113-122.

Smith, A. (1999, April 26). Why digitize? *Council on Library and Information Resources*, 80 [On-line]. Retrieved March 20, 2000:

http://www.clir.org/pubs/reports/

- Sommer, B. (1997). *A practical guide to behavioral research*. (4<sup>th</sup> ed.). New York: Oxford University Press.
- Steward, R. (1993). Does this project deserve the Erasmus prize? *Conservation Administration News*, 54(July), 4-5, 33-35.
- Tibbo, H. (1995). Interviewing techniques for remote reference: Electronic versus traditional environments. *American Archivist*, 58, 294-310.
- Velgos, J. (1999). One bit at a time: Bringing the state archives to the internet. *Texas Library Journal*, 75, 18-19.

Whaley, J. (1993). Digitizing history. American Archivist, 57, 660-672.