John Foster. Institutionalizing Success: The Growth of a Digital Strategy in the Cornell University Library System. A Masters Paper for the M. S. of L.S. degree. April, 2002. 51 pages. Advisor: Helen R. Tibbo.

This paper traces the growth of digital preservation and access programs, and the institutional structures created to preserve them, in the Cornell University Library system. The paper is a case study, tracing the transition from a series of individual projects to an institutional approach that has made Cornell a leader in the field. This approach provides for both the preservation of collections already created and for the establishment of standards to assist in the development of further work in this field.

The staff of the Cornell University Library system has made use of the opportunities presented by digital and network technology to augment the library's traditional role of preserving and providing access to scholarly materials. They have forged cooperative relationships among various segments of the university, as well as corporate and governmental partners. Finally, they have taken a proactive approach to the establishment of best practices for digitization programs.

Headings:

Preservation of library materials – Automation.

Virtual library.

Historical libraries and collections – New York.

Optical data processing

INSTITUTIONALIZING SUCCESS: THE GROWTH OF A DIGITAL STRATEGY IN THE CORNELL UNIVERISTY LIBRARY SYSTEM

by
John C. Foster

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, 2002

Approved by:

_____

Advisor

Table of Contents

1. Introduction

Viewed purely in terms of volume, the library system at Cornell University has established itself is a leader in the production and use of digitized scholarly content. Over the course of the last ten years, the various elements of the Cornell University Libraries (CUL) have undertaken more than 70 projects involving digitization, electronic preservation, and distribution of scholarly, historical, and cultural heritage materials. The contributions of the CUL, however, go far beyond these considerable achievements. Cornell has demonstrated a consistent commitment to aggressively pursuing the potentials of digital and communications technology for augmenting the mission of the library system. The CUL system has been especially assertive in its efforts to promote collaborative project models and intra-institutional cooperation. The establishment of the Cornell Institute for Digital Collections (CIDC) in 1997 is an example of this institutional approach to a sustainable digital strategy. Cornell has also been a leader in forging collaborative projects with industry, and in solicitation of funds from government agencies such as the National Endowment for the Humanities and the Institute of Museum and Library Services.

Finally, and perhaps most importantly, Cornell has been a leader in establishing benchmarks, best practices to facilitate the innovative use of new technologies and to create models whereby other projects can benefit from their experience. Examples of this include the very prominent publications by CUL staff members in the area of digital projects (Kenney and Chapman 1996; Kenney and Rieger 2000) and the workshops the CUL has conducted. Their participation in organizations relating digital preservation and

access, principally the Council on Library and Information Resources and the Research Libraries Group, and the expansion of the professional literature on topics related to digitization through the production of *RLG Diginews*, a leading source of information for librarians, archivists, and museum curators. The work on this journal has both extended the CUL's influence and provided exemplary service to the field. This paper is a case study and analysis of Cornell's rise to this position of leadership. It will attempt to show what were the crucial factors of this process that have led to the CUL's standing in the area of digital preservation and access, how they relate to the larger mission of the library system, and to what degree they might be models suitable for other institutions and programs to emulate.

There were two primary methods used in the completion of this study. First, there was an extensive consultation with the available literature relating directly to the work done at Cornell. Second, the author conducted a series of semi-structured interviews with personnel in various segments of the CUL (see appendices A and B for documentation on this process). The interviews were meant to fill in the informational gaps left by the published materials, as well as to gain personal perspectives on the significant elements of the work done at Cornell. The notes from the interviews were used to inform the larger structure of the paper.

The body of this paper will have two main parts. The first will consist of an account of key events, processes, and figures that led to Cornell's current position. It will have three subsections. The first will discuss *Making of America I* (MOA 1), its origins in the CLASS Project (a collaborative venture with Kodak to digitize and provide remote access to library materials), and its implications in terms of the adoption of technological

innovations. The purpose of the section will be to highlight the work done in the Department of Preservation and Conservation (DPC), the significance of which is less a matter of the projects completed than of the skills and technology acquired in the process. The second subsection will discuss the work done in the Mann Library to provide digital access to materials relating to agriculture. The focus will be on the process of growth of cooperation between segments of the CUL and the particular role of the Mann Library projects in cultivating new approaches to providing networked access. The third subsection will discuss the practical dimension of the growth of digital projects undertaken by other segments of the CUL, particularly *Utopia*, *Louis Agassiz Fuertes*, and the Project to Democratize Access to Scholarly Sources. These projects will be explored in the context of a larger discussion of the formation of the Digital Access Coalition (DAC) and its transformation in the Cornell Institute for Digital Collections (CIDC). Taken together, these parts will tell the story of the incremental growth of Cornell's strategy for providing digital access to scholarly materials. I will argue that this was a process of building on the work done in the Department of Preservation and Access, but also a move to a more user-driven institutional approach to digitization.

The progress of the CUL from early adopter of information technology to institutional leader has been characterized by a broad vision rather than by a discreet and extensive plan. That is to say that the process was incremental and driven by the particular circumstances rather than by a consistent, pre-existing scheme. On the other hand, Cornell's success has been fostered and enriched a well-considered perspective toward technology and institution building demonstrated by staff members, particularly Stuart Lynn, H. Thomas Hickerson, Anne R. Kenney, and Jan Olsen. This approach has had the virtue of

allowing the integration of network technologies into the traditional mission of the library, rather than making that mission subject to a fetishistic commitment to technology for its own sake.

The second main part of this paper will address the overall strategic approach of the CUL in terms of particular issues and their solutions. Topics addressed in this section will include acquisition of technological and fiscal resources, technology and data management, copyright, and the relation of technological methods to end-users. The purpose of this section will be to highlight the crucial developmental issues raised in the preceding sections, particularly addressing the role of leadership (both personal and institutional), and to analyze the relationship between digitization projects, their user communities, and the larger role of the CUL in promoting and facilitating scholarly work. The discussion will make reference to the framework elaborated by Anne R. Kenney and Oya Rieger in the section of *Moving Theory into Practice* entitled "Mainstreaming Digital Initiatives." Since Cornell has adopted a leading role in research, professional education, and the establishment of prescriptive values in this field, it seems appropriate to use these prescriptions as a tool to evaluate their own performance.

The rapid development of digital and communication technologies has resulted in libraries being confronted with both a variety of new opportunities and with a host of new and unexampled problems. The goal of librarians and archivists must be the measured application of technology in ways that promote the traditional values of the institutions, rather than the adoption of new technologies for their own sake. Cornell's fiscal position in terms of endowment (one of the top 15 universities in total endowment in the United States) and the relationship with vendors (primarily Sun Microsystems) that they have

forged have provided opportunities that may not be in play for other institutions. Nonetheless, I will argue that there are many aspects of Cornell's work in digital preservation and access that are exemplary and that provide potential models for other institutions to emulate.

2. Digital Projects in the Cornell University Library System

Efforts to employ rapidly developing communications technology to enhance access to the various scholarly resources at Cornell began in the early 1990s. Three main areas stand out in particular: the brittle books program in the DPC (which would eventually become MOA 1), the projects conducted in the Mann Library of the College of Agricultural and Life Sciences, and the projects conducted by the DAC with various institutional collaborators. The brittle books project was the first of these projects to begin (Anne Kenney announced the collaborative project with Xerox from which it grew at the ALA Midwinter Meeting in December 1991) and will be discussed first in order to illustrate the early stages of digital projects in the CUL. The next section will consist of a discussion of the collections created by the Mann Library with an eye to their significance in terms of institutional cooperation and user orientation. Finally, I will discuss the formation of the DAC and some of the projects conducted under its auspices.

Reading the copious prescriptive literature that Cornell has produced, one might get the impression that their road to leadership has been an orderly progression. This has not entirely been the case. They have certainly benefited from technologically forward thinking, but they have also shown the flexibility to be guided by the demands of circum-

stances. The purpose of this entire section will be to show some of piecemeal, incremental steps that they have taken to reach their current position.

2.1.1. The Class Project and the Diffusion of Innovation

The DPC, headquartered in the basement of the Olin Library, was a relatively recent addition to the library infrastructure. The department was founded in 1985 with a grant of $90,000 from the New York State Legislature.[1] Shortly thereafter, Anne R. Kenney was appointed as Conservation Reference/Liaison Librarian. This appointment was a crucial moment in the process, since the move toward applying digital technology to traditional problems of preservation and access was, in large part, driven by Kenney's vision. Further, as will be seen below, this influence crossed institutional boundaries and facilitated the growth of a broadened institutional approach to digitization. Kenney was a major force behind a collaborative venture with Xerox aimed at using digital technology both to improve the capacity of the department to preserve materials threatened by embrittlement, and also to provide better access to library materials in general. It is interesting to note that, although the primary implication of the project for the DPC was the potential to improve preservation practices, a more significant outcome of the work would be improving access to materials irrespective of their immediate need for preservation.

The preservation strategy currently employed in the DPC at Cornell has its roots in a collaborative project called College Library Access and Storage System (CLASS) begun in 1991. Cornell partnered with Xerox and the Commission on Preservation and

[1] Under the terms of the 1984 Library Omnibus legislation, funding was given to libraries at Columbia, Cornell, Rochester, Syracuse, New York Public Library, New York University, New York State Library, and State University of New York Albany, Binghamton, Buffalo, and Stony Brook for the establishment of programs to preserve and maintain materials of scholarly and cultural value.

Access in a study that had two primary aims: to evaluate the image quality that could be obtained in paper output from scanned images, and the potential for digitized images to facilitate preservation. The first part of the project involved creating 600 dots per inch bitonal TIFF images of approximately 1,000 volumes selected for scholarly value on the basis of recommendations by a faculty panel and by citation analysis. Half were drawn from the Mathematics collection and half from other disciplines. Other selection criteria included format (no larger than 8.5x11), content (no large or color illustrations), and circulation (heavy). The project was seen by Xerox as an opportunity to test prototypes of document scanning and printing equipment. According the Barbara Berger, this was "the first time that Xerox used its potential customer base to help create a product" (Berger 1996, 50). The collaborative work with Xerox laid the groundwork for the integration of digital and communication technology. It is also an example of a model of collaboration with external partners that would be repeatedly employed in the course of developing Cornell's digital and technological capacities. The salient elements of this process will now be discussed in terms of the adoption of a technological innovation.

The application of new technologies to old problems is a matter particular interest in the context of libraries and archives. The issue is often cast in terms of the challenges present (growth of new media), but there are also potential advantages to be gained. Libraries have not, traditionally, been leaders in applying new technology to their subject area,[2] This is one reason that Cornell's approach is worthy of remark. The CUL system has been singularly progressive in their approach to integrating new technologies into their mission. Before proceeding to the discussion of further projects undertaken at Cor-

[2] In a paper prepared by Elaine Engst and H. Thomas Hickerson in 1998, the authors pointed out that "archivists have sometimes acted like passive victims of technological change." (Engst and Hickerson, 1998). This point might be extended to librarians in general.

nell, is useful to discuss the process of innovation adoption in its earliest phases in order to obtain a clear perspective on the foundations of the process.

The investigation of the pattern of innovations and their adoption (or rejection) has become a major avenue of approach for sociologists seeking a clearer understanding of modern capitalism.[3] Everett M. Rogers' work in this field is preeminent. *Diffusion of Innovations* (Rogers 1962/1995), first published in 1962, is a broadly synthetic work. Taking examples from a wide range of political and cultural situations, Rogers analyzes adopters of innovation and the process in which adoption occurs in terms of ideal types. He defines an innovation as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (Rogers 1962/1995, 11). It will be argued below that the development of MOA 1 out of the brittle books preservation program at Cornell is an instance of the adoption of a new idea and its use in reconfiguring practice to meet new goals. A further aspect of Rogers' work that is of interest in the present context is his schema for explaining the rate of adoption. Rogers cites five factors that affect subject's readiness to adopt an innovation: relative advantage, compatibility, complexity, trialability, and observability. This schema will form the basis of the discussion of the adoption of network technologies by leading actors in the DPC at Cornell, and their resulting expansion into MOA 1.

---

[3] Cycles of innovation and obsolescence have been of interest to sociologists and economists since the beginning of the 20th century. The Austrian economist Joseph Schumpeter's analysis of the larger processes of development in the industrialized world were founded on a conception of the innovation-driven dynamism of modern capitalism (Schumpeter, 1942). Ironically, Schumpeter predicted that the world of modern industry would be characterized by the domination of increasingly large industrial units, a process that would result in a progressively more stable pattern of economic development. This has hardly been the case. Indeed, one of the most intractable problems facing the partisans of digitization in general, and libraries in particular, is the increasing rapidity of product cycles and the instability and unpredictability of the technological environment.

1. Relative advantage. Rogers defines relative advantage as "the degree to which an innovation is perceived as being better than the idea that it supersedes" (Rogers 1962/1995, 212). The application of digital and communications technology to the preservation of brittle books was indeed viewed in these terms. The staff in the DPC saw this technology as providing an opportunity to improve access to embrittled materials through the creation of print surrogates made from digital masters. Department staff viewed the creation of the print surrogates from a digital masters as superior to those created from photocopied masters, since digitization provided a significant improvement in output quality, as well as in ease of manipulation and use. There was less image degradation than with photocopying, and high quality duplicates could be produced, in virtually unlimited quantities, from a single master. More importantly, digitization and networking technology provided a means of access to materials for users to whom they might otherwise be unavailable.

2. Compatibility. This issue of compatibility relates to the degree to which a new technology fits with the current way of doing things. As mentioned above, from Cornell's perspective, the primary focus of the CLASS project was to investigate the potential for creating more accurate master copies of materials in need of preservation, copies which could then be easily replicated without loss of content. The processes used in creating the images for the CLASS project (and for subsequent projects undertaken by the CUL) were analogous to those used in creation of microfilm copies. This was a process with which the participants of the project were intimately familiar. At around the same time as the CLASS project, Kenney and her colleagues were investigating the possibility of using digital masters to create microfilm preservation copies (Kenney 1993). The Association

for Information and Image Management had recently developed a formula for comparing digital and photographic image resolution. They had collaborated with Image Graphics, Inc., a private company from Connecticut. Image Graphics had been doing work in the field of high-speed film recording using electron beam recorders. This work paralleled that being done in the CLASS project, and reflects the commitment in the DPC to a pro-active relationship to technological development. Further, it demonstrates a commitment to the promulgation of standards and their application in practice. For Kenney and her colleagues, the standards and processes elaborated in the CLASS project were, to a great extent, an outgrowth of ideas and concepts with which they were already familiar.

3. Complexity. The complexity of an innovation can have a negative impact on the prospects for adoption. In the case of the CLASS project, the complexity of the new processes was mitigated by the collaboration between the DPC and Xerox, and by an essential similarity between the new technology and the old. The staff at Cornell already possessed extensive experience in the use of microfilm for preserving materials. This base of experience formed a system of indigenous knowledge on which the purveyors of the new technology were able to build.

4. Trialability. A further factor affecting the probability and speed of adoption of innovations is trialability or the potential to try something out on a smaller scale before committing more resources. This is an issue of particular significance in the present context because the CLASS project was, to a great extent, a trial stage for approaches that were more extensively applied in the construction of MOA 1 (as well as in other subsequent projects). As Rogers notes: “[r]elatively earlier adopters of an innovation perceive trialability as more important than do later adopters” (Rogers 1962/1995, 243).

5. Observability. The final factor Rogers discusses is observability: the degree to which an innovation is visible to others. This is rather less important in the present context, although it does relate to issues in the larger world of libraries and their relationship to technology. In the case of the CLASS project, the dimension of corporate involvement led to a decrease in overall observability since the work was meant to be kept secret to protect the corporate profit interest. However, it is certainly the case that MOA 1, which grew out of these efforts, had a high degree of observability. The importance of this fact relates both to the spread of innovations, as well as to the potential prestige goods associated with digitization projects. In the case of the former, the success of MOA 1 and the work in benchmarking and the application of standards derived from microfilming provided an observable example for other potential adopters of these methods. As to the latter issue, the prestige goods realized from the successful completion of projects of the scale of MOA 1 contributed directly to the development of institutional structures within the library system that maintain current products and carry forward further projects of this sort. As such, the intra-institutional visibility of the prototypical CLASS project contributed to the adoption of digital approaches in the larger institutional environment. The extended visibility of MOA 1 yielded both high visibility and prestige goods.

2.1.2. *Making of America 1*

The *Making of America* (http://library5.library.cornell.edu/moa/) project began in 1994 in a collaboration between the CUL and Cornell Information Technologies (CIT) (the administrative department responsible for technological initiatives). It was facilitated financially by a grant of $324,000 from the Charles E. Culpepper Foundation and techno-

logically by collaboration with Sun Microsystems. In 1995 it developed into a joint undertaking with the University of Michigan, with funding from the Mellon Foundation. It was (and is) one of the most extensive projects of this kind ever conducted. It is important not only because it involved the extension of knowledge and approaches developed in the CLASS project, but also because it, in turn, became a model for similar projects in other segments of the CUL. Two important and interconnected considerations underpinned MOA 1. First, there was potential for improved preservation through the creation of digital images that could act as surrogates for the items themselves. Second, there was also the possibility of broadening access by using emergent networking technology. Also of importance were the dimensions of institutional collaboration and management of technological issues.[4]

Cornell's part of the project involved the scanning of roughly 900,000 pages, mostly from journals relating to American history and culture. The ostensible date range of the materials selected extended from 1850 to 1950, but the bulk of the works date from the early part of this range. There are several reasons for this. Probably the most important is the fact that the older publications fall outside of copyright protection and therefore did not involve the CUL in legal complications. Another factor is the brittle books orientation from which this project grew. The paper used to print books in the second half of the 19th century was exceptionally bad and the CUL was in possession of a large number of volumes that had become practically unusable through embrittlement. The work in digitization was seen as a means of addressing this problem. Since journals do not circulate in the same way that monographs do, it is difficult to track the frequency with which

[4] Since the purpose of this section is to discuss the larger implications of MOA 1 for Cornell's digital strategy, I will not attempt to provide a comprehensive discussion of the technical dimensions of the project. Rather, I will discuss technical aspects to the extent that they are relevant to the larger goal.

they are used. Selection was, therefore, informed by consultation with faculty providing at least some contact with user needs, although not to the level that might have been achieved if frequency of circulation has been an available criterion.

The mounting of the materials on the web presented a number of technological challenges, the solutions to which grew from the incremental acquisition of expertise that underpinned the practical dimension of Cornell's digital strategy. For example, this was one of the first projects to use outsourcing as a workflow solution. It was decided that the volume of images to be produced would be too large to be handled adequately in-house. Therefore, Northern Micrographics Inc, a Wisconsin based company performed the actual scanning of the documents. Storage of digitized content presented other challenges. In a decision that would later have unfortunate consequences, the images were stored on digital platters in an EPOCH "jukebox" digital server. By 1997 this technology was no longer supported and was failing rapidly, necessitating a rapid migration of the data (it now resides in a Sun Enterprise 3500 Server).

One significant problem that had to be resolved internally related to the software that users would employ to manipulate the collection. Cornell intended to use the Dienst/Hunter protocol, which was already in use locally. Dienst/Hunter was a page turning program used to search and manipulate portions of the collection. It essentially searched the metadata for individual volumes, then allowed the user to search within single titles. The University of Michigan Digital Library Production Service had developed its own page turning software. This package, often referred to as "Michigan middleware" allowed users to search among a number of volumes rather that within a single one. At first there was a certain degree of institutional resistance to the adoption of the Michigan package.

Dienst had been developed at Cornell and had already been used in other digital projects. Eventually it was decided to retain both side by side on the Cornell site, allowing users to decide which they were most comfortable. Although the Dienst/Hunter protocol retained some partisans at CUL, it seemed to result in more frequent visits by perplexed users to the reference desk at the Olin Library. In the end, the Michigan middleware package was adopted, albeit in a slightly modified form.

David Ruddy reconfigured the Michigan middleware to fit the specific needs of the Cornell collection. Ruddy had been brought in during 1998 to integrate the metadata for the project and to mark it up in a TEI compliant format. This was crucial to the success of the project. Provision of network access had been a consideration from the earliest origins of MOA 1, the incremental nature of its growth meant that metadata had not been systematically compiled.

A further access issue related to creating the capacity to search the full text of the journals. Originally, the University of Michigan had proposed an extensive program involving multiple runs of the scanned page images through OCR software and corrections. Cornell technical services experimented with running the images once through the Textbridge OCR application and found that obtained an acceptable level of accuracy without the added time and expense of correction. Project staff decided to leave the OCR products "dirty" because the page images would be presented to users and because the inaccuracy in the full text did not seem to hamper searching functionality in a significant way.

The experience of the MOA 1 is illustrative of a number of crucial aspects of the rise of digitization as a strategy in the CUL system. First, it involved inter-institutional collaboration (the cooperation with the University of Michigan and the Mellon Founda-

tion). Second, it extended the expertise obtained in a previous increment (the application of microfilming standards to digitization developed in the CLASS project), while also providing further expertise (the integration of the metadata, the "dirty OCR" approach). Third, it exhibits a high degree of observability (in Rogers' sense of the term). While there has not been a comprehensive user study to test the contribution of MOA 1 to the educational mission of the library, it should be noted that the University of Michigan claimed that MOA 1 was "averaging 1,000,000 true uses per month."[5] Finally, the successful completion of MOA 1 has acted as a model for technical and institutional organization of subsequent projects.

In this section I have tried to illuminate some of the significant moments in the process of development and innovation in the early stages of digitization in the CUL system. The work on the CLASS project and on MOA 1 were important for their new and innovative approach to providing access to users, but even more so in the respect that these projects facilitated the growth of technical experience and knowledge that would provide a foundation for further work. It is also important that the vision of the role of network technologies in access provision was able to build on an already existing institutional and technological basis in order to develop the critical mass necessary for the further growth of this approach. This, however, is only one part of the story. The next subsection will discuss some of the main features of the history of digital projects undertaken in the Albert R. Mann Library. The approach found here was somewhat more directly user oriented as particular features of their projects arose from a community of use that the Mann Library was tasked to serve. It is important to note, however, that the programs

[5] http://www.diglib.org/pubs/news03_01/Michigan.htm. We may, of course, wonder what precisely is meant by "true uses," but in any event 1,000,000 is a considerable number.

developed at Mann involved a process of building on the knowledge acquired in other parts of the institution.

2.2. Digital Projects in the Mann Library

The relationship of the Albert R. Mann Library, which is a part of the School of Agriculture and Life Sciences, to the rest of the CUL is an example of the peculiar institutional structure of Cornell University. Whereas most colleges and universities are either private or public, Cornell is both. Of the thirteen individual colleges that make up Cornell, four are statutory (i.e. part of the SUNY system), with the remaining nine funded by private endowment (one staff member characterized it as "thirteen schools united by a heating system"). In a further complexity, the various libraries are funded by the individual units that they serve. Thus, funding for the Mann Library comes from the College of Agricultural and Life Sciences (CALS), and its administrative elements are self-contained. This lack of a unified structure in the library system has been an important element in the impetus toward creation of a means of institutional communication and coordination. CALS is part of the statutory segment of the university, but the Mann Library is part of the organizational structure of the CUL. In this instance, the CUL functions as a coordinator among the various individual library systems (see chart in Appendix C). Funding and day to day administration are handled by the CUL in the case of the endowed libraries, and by the individual colleges in the statutory segments. There are some functions that are distributed across these boundaries. One such function is preservation. The DPC conducts some of its operations in the Mann Library, although it is actually based in the Olin Library.

Jan Olsen, who served as Mann Library Director from 1982 until 1997, was an avid proponent of digital libraries. Starting in the early 1990s under Olsen's leadership, the Mann Library undertook several digital projects and received the inaugural American Library Association/Meckler "Library of the Future Award" in 1993 in recognition for the work done there in providing digital access to scholarly materials. The next year, the Mann Library was the subject of a special issue of the journal *Library Hi-Tech* that declared it "the prototypical digital library."[6] The projects undertaken during Olsen's tenure at Mann were reflected a mission and a community of use rather different than those served by the Olin Library. The mission of the Mann Library is, in the first instance, to serve the College of Agricultural and Life Sciences, of which it is a part. Because CALS has a narrower focus than the university at large, it may be easier to specify goals and to develop digital projects to meet them at Mann than in a large university library such as Olin. Under Olsen's leadership, the collection and preservation policies, including those involving digitization, were literature-oriented rather than library-oriented. In an article published in the issue of *Library Hi Tech* mentioned above, Olsen noted:

> Selection for preservation invariably involves sophisticated bibliographic projects and rigorous scholarly review of titles to establish priorities. While the bibliographic methods used in each project are adapted to address the nature and needs of the specific literature, they are all variations on a basic theme. The focus in every case is on analyzing and evaluating the literature of an entire discipline or discrete subject area, *not* strictly on the collection of a particular library (e.g. the Mann Library). (Olsen 1994, 83)

The work done in the Mann Library demonstrates the readiness of Olsen and her staff to integrate technological and practical innovations arising in other segments of the system with their own approach. The commitment of the Mann Library staff to investigate avenues for employing digital technologies for access provision date from at least

[6] *Library Hi Tech* (14) 9, 1994.

1988[7] when discussions for what would become the Chemistry Online Retreival Experiment (CORE) began. The project, which actually got underway in 1990, was a joint venture between the Mann Library, the Chemical Abstracts Service of the American Chemical Society, Bellcore, and OCLC. The goal was to provide networked access to chemistry journals. The project provided a number of interesting challenges, both in presentation of the materials (i.e. the complex and peculiar set of symbols used in representing chemical processes) and in getting the potential user population to actually try out the service.

The project was the subject of an unusually extensive user study that sought to track not only patterns of use, but also the substantive uses to which the collection was put (Entlich, et. al. 1996). This was done by not only tracking use through transaction logs, but also through online questionnaires and face-to-face interviews. For our purposes the results of this study are less important than the fact that it was done at all. It reflects a practical and extensive commitment to users and an interest in ensuring that digital products functioned successfully within established patterns of scholarship. Further, it also evidences a degree of testability (in Rogers' terms). The study of use allowed the knowledge gained in the project to be operationalized. It gave the Mann Library staff, explicit information about how their users interacted with their digital products and allowed for improved user orientation, but also gave them an actual example of how subsuquent projects might work. This was certainly a factor in the willingness of the staff to adopt this mode of access provision.

---

[7] Employment of electronic resources at the Mann Library actually extends back into the 1970s, but for purposes of the developmental trajectory being described here, the CORE project may be taken as a useful (if somewhat artificial) starting point. For a more long term approach to this topic see Susan J. Barnes article in the issue of *Library Hi Tech* devoted to the Mann Library (Barnes, 1984).

Similar to the Olin Library, the Mann Library experienced a serious problem with the degradation of books from the mid-19th century. Although one might assume that scientific collections primarily contain very recent materials, the Mann Library maintains a large historical collection as well as contemporary sources. The decay of these materials was no less severe than in other branches of the library system. It was estimated the some 350,000 volumes in the Mann Library Collection alone were threatened by embrittlement. The Mann Library's service mission and the need to address the brittle books problem were the motivating issues behind a series of projects designed to preserve and provide access to materials relating to agriculture and home economics.

In 1993, the Mann Library participated in a collaborative project with the U.S. Agricultural Information Network to construct a framework for preservation of the historical literature of agriculture at the national level. The program grew out of work that had been in progress since 1989 toward establishing a bibliographic basis for preservation of literature relating to agriculture. A national panel of scholars, led by Mann Library Special Projects Librarian Wallace Olsen, identified 4,500 books that constituted the body of fundamental work in this field. This study led to the publication of a seven-volume bibliography on the topic that formed the groundwork for a number of subsequent projects. This program sought to coordinate the activities of the National Agricultural Library and the libraries of the various land grant institutions across the country to work toward the preservation of endangered materials in the field of agricultural theory and practice.

Beginning in 1994, the Mann Library undertook a program to identify and preserve materials crucial to the study of agricultural practice. This work was done in con-

cert with Anne R. Kenney and the DPC, and involved application of the same microfilm based approach that had been used in the Olin Library. Jan Olsen's commitment to exploring the potential for augmenting the preservation and access provision missions of the library led to the application of the methods used in the brittle books preservation project to give the work done on the core literature of agriculture greater practical value. The resulting program, called the Core Historical Literature of Agriculture (CHLA), was conducted with funding from the NEH, the Rockefeller Foundation, the USDA, the U.S. Department of Education, and the U.S. Agricultural Information Network, as well as some funding provided by the College of Agricultural and Life Sciences. It was intended as a conscious attempt to use digital technology to improve access and thereby to integrate this literature more fully into the intellectual life of the school.

The technical dimension of this project was organized along the lines established in MOA 1. Here we see the effect of the CLASS project and MOA 1 as institutionally visible trials that yielded significant prestige goods Items were scanned at 600 dpi bitonal and the scans used for the creation of archival quality microfilm copies of the items. The page images were then scanned with Textbridge Optical Character Recognition software. The resulting documents were left uncorrected and used to reference the scanned images. Access to the collection is provided via the same software package used in MOA 1. One significant difference between MOA 1 and CHLA was the copyright status of the materials selection. While exclusion from copyright protection was one of the selection criteria for MOA 1, this was not the case in CHLA. The selection process for CHLA was primarily guided by the academic significance of the works in question. On the one hand, the primacy of this consideration is a positive aspect since it constitutes a commitment to

academic standards as opposed to purely legal considerations. On the other, it does mean that the scope and progress of the project has been somewhat limited. Efforts to secure copyright permission have slowed the process, and as of this time only 815 of the proposed 4,500 works have been digitized. Nonetheless, the project is exemplary in its combination of digital technologies with the imperatives of scholarly work.

The Mann Library has extended this work to a second subject area resulting in the creation of a collection focusing on literature for the education and advising of homemakers: the Core Historical Literature of Home Economics (CHLHE). This project has recently received a sizeable grant from IMLS to digitize 1,500 items from before 1925 (which should simplify most of the copyright hassles).

A third major initiative that the Mann Library has undertaken relates to provision of access to serials in the field of agriculture. The Essential Electronic Agricultural Library (TEEAL), while not involving actual digitization on Cornell's part, represents an innovative use of digitized materials. Cornell haspartnered with the owners of digitized serials to provide a package of materials for creation of knowledge resources in developing countries. The journals in the base package cover the years 1993 to 1996, with 1997 and 1998 available in the upgrades. These resources, which Cornell estimates would cost $600,000 in the developed world, are made available to a list of developing countries for prices ranging from $5,000 to $15,000 (with subsequent yearly charges for updates).

These are just three examples from the history of digital projects in the Mann Library. There are many others, but the purpose here is to illustrate some of the approaches adopted in the Mann Library as part of the larger history of the growth of Cornell's digital strategy rather than to catalog such efforts. The third and final subsection will discuss

the trajectory of development from the Digital Access Coalition to the Cornell Institute for Digital Collections and some of the projects associated with this process. The goal will be to further illustrate the general approaches to digitization employed at Cornell as well as to chart the growth of institutional structures that exemplify Cornell's "projects to programs" approach.

2.3. DAC, CIDC, and Institutional Infrastructure

Cornell's commitment to integrating communications technology into the mission of the library system and the university at large has resulted in the establishment of institutions to facilitate this process. The Cornell Institute for Digital Collections (CIDC) was founded in 1997 and grew out of an earlier institutional grouping called the Digital Access Coalition (DAC). The DAC was founded in 1992 with the explicit goal of promoting the use of communication technology to better meet the educational and scholarly goals of the institution, to coordinate work among various segments of the library system, and to encourage the formation of partnerships between the university and external entities. It was a collaborative effort between Thomas Hickerson of the Division of Rare and Manuscript Collections (RMC) of the CUL, and Dr. Geri Gay, an associate professor of communications. Their goal was to establish an organization that would cross institutional and disciplinary boundaries in order to facilitate the use of digital imagery to present collections in an integrated manner.

The DAC was of particular importance given the discontinuous nature of Cornell's larger institutional structure. Hickerson recognized that the segmented nature of the libraries made duplication of effort significantly more likely, as well as reducing the

chance for effective and integrated presentation of library resources. Further, his partnering with Gay constituted an attempt to bring teaching faculty into active participation in the process of developing digital resources. Dr. Gay was a particularly apt partner in this endeavor. An associate profession in the Department of Communication, Gay specialized in interactive media. In 1988 she founded the Interactive Media Group, a working group to study the potential applications of technology to education. In addition, the inclusion of a faculty member from the Department of Communication in the DAC was significant in and of itself. The Department of Communication is a subunit of the School of Agricultural and Life Sciences. Gay's inclusion functioned to forge a link between the efforts in the statutory and endowed segments of the university.

Hickerson and Gay were aware that there were a number of projects involving digitization and networking technology going on independently of each other. Their first move was to organize a meeting for all of the various players, most of whom were unaware of the existence of the others. The meeting demonstrated that, for the most part, they shared common perspectives and a common technical terminology. The DAC was formed to promote and coordinate these projects.

The fundamental principle on which the Digital Access Coalition was founded was that digital technology could be used to produce materials that would act as surrogates for the items that they represented. An important element of the program was the attempt to move beyond the limitations to use resulting from the storage of digital information on localized media. In a paper delivered in 1997, Thomas Hickerson characterized the approach by noting that, "The use of World Wide Web technology provided a critical transition from our initial CD-ROM applications. In the case of some users and uses, the

change was not initially viewed as an improvement, but the Web has quickly become the standard for networked access to digital collections." (Hickerson 1997, para. 3). Between 1992 and 1997, the DAC conducted a range of projects seeking to integrate information technology with special collections and to bring the latter into the pedagogical mainstream. These included *Utopia*, the Project to Democratize Access to Scholarly Sources, and the *Louis Agassiz Fuertes Collection*.

Utopia was an internally funded collaborative project created in 1994 to provide digital surrogate images of fifteenth and sixteenth century European art and architecture, conducted by History of Art Department, College of Arts and Sciences; the Knight Visual Resources Facility, College of Architecture, Art, and Planning; the Herbert F. Johnson Museum of Art; and the Rare and Manuscript Collections. The original plan was to scan images drawn from various collections and to store them on Kodak Photo CDs for use in classroom settings. Scanning of the materials was outsourced to a company called Boston Video, and resulted in the creation of over 4,500 images on 47 CDs. The images were then cataloged using Kodak's Shoebox image management software package. The life of the project makes clear the marriage of innovation and necessity that often characterized work in these early days of digitization. The stated goals of the project had been to facilitate classroom use of these materials, and especially to bring undergraduates in contact with them. The CD Rom-based nature of the project facilitated this, but also proved to be not entirely suited to the user orientation of the project. Looking back in the project in 1997, Noni Korf Vidal, who had worked as the interface design manager on the project, noted:

> The act of creating digital resources is a transformative act that demands a reorganization of how we define our goals and guiding principles as well as how we

> implement and improve our resources. As resource creators we may try to keep the goals of our project separate from its specific implementation, yet the user experience of our resources is greatly influenced by their implementation. In our move from a CD-ROM-based resource to one available on the Web, we found that professors who had adopted *Utopia* on CD-ROM experienced disappointment and dismay when some of their activities were not supported within a Web-browser. Toolsets and features to which users had become accustomed were taken out of their hands. Had we been able to follow the tenets of user-centered design we might have chosen to remain offline, but being digital means accepting change, including changing beliefs in how we can best serve our users. (Vidal 1997, para. 7)

More serious than user difficulties was the fact that Kodak eventually decided not to support their Photo CD technology in the long term. This necessitated the migration of the data to a Filemaker Pro database management package, also facilitating presentation on the web.

The development of *Utopia* was conducted with extensive front-end input from potential users. While this improved the prospects for effective use, this process also had pitfalls. In the same paper cited above, Noni Korf Vidal noted:

> In the online environment, there is an inherent dilemma in the practice of User-Centered Design. User-centered studies are predicated on a fairly strict notion of who your users are. Once you make a Web resource, however, determining who your users are is impossible. Having tailored a resource to specific users, it is possible that it "won't fit" the needs of users outside your defined community. (Vidal 1997, para. 9).

These concerns led the design team to carefully reassess (in their terms “reframe”) their approach to taking potential use into account. Since *Utopia* was intended, in the first instance, for use by professors and students at Cornell (rather than by some amorphous grouping of web-based users or the public at large), obtaining dynamic user feedback allowed the designers to have a clearer understanding of the ways in which the collection was actually being used. This reflects an awareness of the considerations raised at the end of the preceding section vis-à-vis user studies, although the predominantly local orienta-

tion of the project avoided some of the difficulties with understanding substantive patterns of use. This is a luxury not available to projects with a more "extramural" focus.

A second important project undertaken under the auspices of the DAC was the *Louis Agassiz Fuertes Collection*, a project to create a prototype digital collection using materials relating to the prominent naturalist. Conducted at about the same time as *Utopia*, *Fuertes* involved collaboration both inside the CUL (the collection was created on behalf of the Kroch Library that houses Cornell's special collections), as well as with institutions external to Cornell University. *Fuertes* came about in concert with a collaborative venture involving the DAC, Eastman Kodak, the Commission on Preservation and Access, and the University of Southern California. Entitled KLIC (Kodak Library Image Consortium), the project sought to explore (and promote) the uses of Kodak's PhotoCD technology in academic libraries and archives. Using this technology, *Fuertes* was intended to produce a unified collection of digital surrogates whose projected community of users would not be limited to professors and students at Cornell. The project brought together images of *Fuertes*' artwork with materials contained in Cornell University's special collections, as well as those in the Johnson Museum of Art, and the Laboratory of Ornithology.[8] The bulk of the digital products were made by scanning images of original materials from 35mm slides. The scanning was outsourced to Boston Photo Inc. The products were originally stored on Kodak PhotoCDs, but were migrated to the Sun server after the demise of that technology. *Fuertes* illustrates a number of important aspects of the DAC: cooperation between separate elements of the CUL structure, as well as coop-

[8] . One side effect of this project was the eventual transfer of items from the Laboratory of Ornithology to the CUL, thus allowing it to receive proper storage and care.

eration between private corporations and government entities in the interest of building projects that might be beyond the resources of the individual partners.

The Project to Democratize Access to Scholarly Sources was rather more ambitious than *Utopia* or *Fuertes*. The project was initiated under the auspices of the DAC in 1994. It built on techniques and approaches developed in the course of the work on *Utopia* and *Fuertes*. The subject matter for this project was the papers of Cornell founder and namesake Ezra Cornell, which were housed in the Rare and Manuscript Collection of the Kroch Library. The digitization of this collection provided an opportunity to build on the knowledge gained in other projects. As Engst and Hickerson noted, their choice of subject matter,

> [W]as based both on the richness and diversity of Cornell's collections and on the recent initiation by the University Library and Cornell Information Technologies of a national, multi-institutional project called "Making of America: Creating Electronic Pathways to our heritage." While "The Making of America" project emphasized published sources, this proposal would use manuscript and graphic materials to document political, social, and technological developments in the United States. (Engst and Hickerson 1998, 12)

Their choice of Ezra Cornell also had the benefit of attracting funds from a private donor (a Cornell alumnus). Construction of this project would yield valuable experience in providing access to collections containing a large variety of items and involving a wide range of activities. It also provided staff from the Instruction Media Group with an opportunity to investigate information seeking and use among students. Finally, the project was also an opportunity to develop collaborative practices among various segments within the university, practices that could then form a basis for more extensive institutional development.

The evolution of the project was marked by the sorts of conflicts that often occur in collaborative work between groups with differing goals and institutional cultures. Difficulties relating to division of funds highlight the problems caused by Cornell's institutional structure. Since the CUL and IMG were in fiscally separate segments of the university (CUL in the endowed portion, IMG in the College of Agricultural and Life Sciences with statutory funding) the budget for the program had to be structured in order to account for this division. Engst and Hickerson also note scheduling difficulties arising from this same division (Engst and Hickerson 1998, 14). A subtler problem was the difference in focus between the two groups involved in the program. While the staff at the Rare and Manuscript Collection was interested in the substantive content of the project, those from IMG were more interested in the technological dimension, and this resulted in a certain degree of friction. Engst and Hickerson point to a number of factors that smoothed some of the problems. Interestingly, great emphasis is placed on the importance of hiring Noni Korf Vidal in 1996. Vidal, who also worked on the later stages of *Fuertes*, had studied under Professor Gay. She had also done a certain amount of library work, which gave her "some understanding of library concerns." (Engst and Hickerson 1998, 15) This is an example of another exemplary practice in the CUL: the consistent pattern of good human resources decisions. It would be difficult to attribute this to any particular factor in the process. Nonetheless, the growth of the CUL digital programs has been assisted by their ability to put the right people in the right places and to exploit the talents and vision of people already working in the system.

The programs initiated under the auspices of the DAC show a pattern of growth in practical knowledge and technique, but also in vision. Taking advantage of emergent

technology and while integrating practices developed in the CLASS project (and later in MOA 1), these projects have built on that technical knowledge and used it to create more effective, focused, and user oriented products. In addition, the staff of the CUL system has employed a collaborative model, both with external organizations and with other segments of the library and the university to augment the system's own resources. This model has been shown to be effective in terms of mobilizing the financial resources necessary to create digital collection. Perhaps more importantly, it has become an effective model of knowledge management in which skill and perspectives from various sources are combined for more effective planning and production. The final portion of this section will briefly discuss the institutional structure that has been constructed to provide support for this model.

In the final chapter of *Moving Theory into Practice*, Anne R. Kenney proposes a strategy for the long-term maintenance of digital assets. She writes:

> The move from projects to programs is based on the premise that *digital collections are institutional assets*. Institutions must safeguard these investments to maintain their long-term value and utility. Cultural thinking must shift away from viewing digital imaging efforts as short-term or experimental. (Kenney and Rieger 2000, 153)

This strategy has been exemplified in the transition from the DAC, which was essentially project-oriented, to the CIDC, which has a much more program-oriented, and therefore sustainable, approach. Funding in academic institutions is invariably subject to competition, much as we might wish it were otherwise. The Digital Libraries and Information Technology (DLIT) section (of which CIDC is a part) has become the second largest unit in the CUL Central Services budget and the fifth largest expenditure over all. The role of the CIDC and DLIT as institutional advocate and defender of digital assets cannot be

overestimated. DLIT has become the institutional advocate for the CUL system's strategy for creating and employing digital resources. CIDC has become an effective means of coordination and knowledge management, both in terms of the technical know-how required for digital projects, as well as in promoting a collaborative model for structuring the projects themselves. There is a sense in which this may be the most important development of all. Digital products are fragile and knowledge resources even more so. The transition from a project-oriented to a program-oriented approach in the CUL has laid the groundwork for the prescriptive work done in the CUL. This may, in fact, be their most important innovation.

3. Digitization and Leadership

In this section I will discuss some of the larger issues surrounding Cornell's strategy for developing sustainable programs in digitization and access provision. Cornell has managed to address some of the most difficult problems surrounding the creation of digital assets in the course of completing a large number and variety of projects. They have also gone a step further and entered the field of prescriptive benchmarks and best practices. Further, through their presentation of NEH sponsored workshops on digitization, their cooperative work with the Research Libraries Group, and their work on *RLG Diginews*, the staff in the CUL have explicitly and self-consciously taken on a leadership role in this area. The first part of this section will be concerned with issues of a practical nature, relating to technology and data management and the methods the CUL adopted to approach them. In the second part of this section, I will discuss the prescriptive work

done by the CUL staff and try glean some of the larger conclusions to be drawn from Cornell's experience.

3.1. Issues in Electronic Preservation and Access

Cornell has achieved a position of leadership in no small part because of the volume of the their digital products. Their collection currently comprises roughly 2.5 million images, nearly half a terabyte of data. More importantly, they have been very active in proposing standards and best practices. This commitment to formalizing and systematizing knowledge and technique is unparalleled. In this section I will examine some of the major issues relating to digital preservation and archiving and attempt to highlight Cornell's approach to addressing them as a means of providing an anatomy of innovation.

One of the most significant problems facing digitization programs is that of maintenance of the data over the long term. While the magnitude of this issue for Cornell is rather larger than for most other libraries, it is a problem of general import for all research libraries and archives creating digital assets. Much of the thinking on this topic has been devoted to preservation at the national level, but several themes of interest to this paper recur. In particular, we see a continued call to create institutional infrastructure in order to facilitate a distributed approach to preserving cultural heritage and scholarly materials.

A study conducted between 1994 and 1996 by the Taskforce on Archiving of Digital Information, put together by the Committee on Preservation and Access of the RLG (Committee on Preservation and Access 1996) recommended the establishment of a national network of repositories held to broad standards for archival preservation of "cul-

turally significant" electronic data. Starting with an analysis of the issues, such as technical obsolescence, data migration, and the legal issues surrounding the materials themselves, the taskforce then sought to build these into a conceptual framework from which appropriate standards could be developed. From a practical standpoint, the solution to preservation in the long term was seen as migration.

> Copying from medium to medium, however, also suffers limitations as a means of digital preservation. Refreshing digital information by copying will work as an effective preservation technique only as long as the information is encoded in a format that is independent of the particular hardware and software needed to use it and as long as there exists software to manipulate the format in current use. Otherwise, copying depends either on the compatibility of present and past versions of software and generations of hardware or the ability of competing hardware and software product lines to interoperate. In respect of these factors -- backward compatibility and interoperability -- the rate of technological change exacts a serious toll on efforts to ensure the longevity of digital information. (Andre, et al. 1994, 7).

Much of the report is devoted to attempting to assess the changes that new technologies will effect on the information environment. There is also an extensive discussion of the fiscal dimension of the problem, although this is necessarily in rather general terms. One interesting feature of the report is its focus on the need for "deep infrastructure." Although this concept is never precisely defined, it seems to call for the development of institutional resources as a necessary foundation for addressing the technological issues.

A similar approach is taken in a report published in Great Britain by the National Office of Digital Archiving (National Office of Digital Archiving 1997). This report looks at these issues in the context of England and Ireland. Although the methodology differs somewhat from that used in the study discussed above, the conclusions that are reached are substantially similar: there needs to be some sort of institutional body to coordinate standards and practices for the maintenance of electronic records on the national

level. Further, it is argued that the funding for this institution should come from public sources, so as to ensure that appraisal decisions are made on the basis of considerations of substantive historical and cultural value. One assumes that the premise here is that funding in the context of a national framework is less susceptible to the financial constraints of the moment, which seems rather a questionable proposition. In any event, the report also addresses legal and commercial considerations, as well as those of timescale (i.e. should the goal be to preserve these records in perpetuity). An interesting facet of this particular report is that, rather than presenting one unitary position for all of the contributors, a range of positions are outlines. For instance, on the issue of timescale alluded to above, the opinions expressed run the gamut from a complete renunciation of any sort of temporal consideration to rigorous controls on what will be selected based on the institution's capacity to maintain all of the item's functional dimensions.

Both of these studies attempt to address the problem of long-term digital asset management in terms of institutions that might be established in order to oversee and coordinate. As such, both are relevant to the project at hand since they attempt to integrate technical and institutional approaches to the problem. The analytical undercurrent of both reports is an attempt to address the fundamentally unstable nature of technology and technological development by thorough, well-considered, and flexible institutional structures. Some portions of this institutional approach have been exemplified at Cornell. The foundation of institutional entities such at the CIDC and its parent unit, the Digital Library and Information Technologies section of the CUL, were important steps in the formation of institutional infrastructure for sustaining digital projects. The extensive production of digital materials in the various projects undertaken at Cornell, coupled with the

segmental structure of the library system, has created a complex data management problem. CIDC is a forum in which representatives from various elements of the library system can meet to coordinate projects and to create an enhanced environment for knowledge management. It is also an institutional player of increasing prominence, thus putting it in position to defend digital assets in the course of the overall competition for funding within the university.

It is also worth noting at this point the continuing relationship that Cornell has established with Sun Microsystems. This has enabled them to acquire high-end storage capacity with minimal expenditure. At the time of this paper's writing, Cornell is preparing to transfer their digital assets to a new Enterprise 4000 server with the capacity to hold a terabyte of data. Obviously this will allow a dramatic expansion of their digital assets. It is also evidence of the continuing strategy employed by the CUL of working in partnership with industry.

The myriad products created in the course of digitization projects have saddled Cornell with a memory overhead problem running to around half a terabyte. The issue of archiving this data is of particular relevance since the potential for digitized representations to function in an archival manner was a stated goal of the studies conducted by the DPC at Cornell in the early 1990s. At this time, the technological environment is such that digitized materials cannot be viewed as archival-that is, likely to survive for the indefinite future. The onus then falls on the storing institution to effectively plan to deal with issues such as data migration, corruption, and loss. The history of digital projects at Cornell contains several instances of forced migration out of obsolete media (Xerox PhotoCD technology and the EPOCH optical disk "jukebox"). In the last two years, Cor-

nell has undertaken a study with funding from IMLS to plan strategies for addressing issues relating to long-term data management. Working a framework similar to that found in the CPA and NODA projects, the report recommends "the establishment of a centralized depository within the Library's Digital Library and Information Technology (D-LIT) infrastructure for ensuring a cost-effective preservation strategy over time" (Kenney, et. al. 2001).

The report also details the institutional resource commitment necessary for data archiving and the expenses incurred in managing the data from MOA 1 in the last five years (Kenney et. al. 2000, 25). According to the report, these costs have run to $1,075,017.37, or $1.18 per image. While $.24 per image per year does not sound like much, the resources necessary for maintain a large digital presence are staggering. These are costs that must be met by the institution itself without assistance from the agencies that originally funded the production of the collection. This is a considerable commitment, even given the fact that Cornell's endowment ranks in the top fifteen in the United States. The data from MOA 1 comprise over 900,000 images and a total file size of 175 gigabytes, but this is only a portion of the overall digital collection numbering roughly 2.5 million images and thus, the one million dollars in maintenance to date is only a portion of Cornell's data archiving expense.

Information technologies have dramatically increased the potential for providing remote (and otherwise broader) access to rare or unique materials. However, this potential comes with unavoidable costs. While it is possible to obtain outside funding for the creation of digital products, their long term retention involves the maintaining institutions and the users in cycles of technology consumption that are complex and costly, and that

promise to become more so. Some of these problems can be addressed via the well-considered application of technology. Some problems, such as the necessity of obtaining ever-larger storage and transmission capacity in order to maintain current levels of functionality, are less susceptible to such measures.

3.2. Technology for Technology's Sake

Marx argued that commodities have a fetish character under capitalism, i.e. that they appear to have qualities such as value in and of themselves. The advent of what Manuel Castells has termed the "network society" (Castells 1996) has given rise to a fetishism of technology, in which the application of technology to various spheres of life is promoted merely (or primarily) on the basis of the value of technology itself. Andrew Feenberg (Feenberg 1991, 2-23) has argued that there are essentially two modes of viewing technology: the instrumental and the substantive. According to the instrumental view, technology is merely a tool, without significant content of its own: a mere medium for achieving whatever ends to which it might be set. "Given this understanding of technology," says Feenberg, illustrating the most optimistic assessment, "the only rational stance is unreserved commitment to its employment." (Feenberg 1991, 6). With the substantive view, "technology constitutes a new type of cultural system that restructures the entire social world as an object of control." (Feenberg 1991, 7). This is a situation that we should wish to avoid, since it transforms the substantial aspects of human life and work into media through which technological capacities are expressed. The instrumentalist thinks of technology as a means to the end of controlling his environment, in one respect or another. The substantivist worries that this control motive becomes an end in itself.

Feenberg himself proposes a third way that he terms the "critical theory of technology." He seeks to overcome the excessively optimistic approach characteristic of the instrumentalists, while stripping away the more hysterical pessimism of the substantivists. Feenberg's position is that technology has the potential to augment the scholarly and humane values of society, but only if its application is tempered by the recognition that technology is not an end in itself. What is needed is an approach to technology that takes into account both the possible advantages of its employment and the potential effects on patterns of work and community life.

Clearly, it would be absurd to suggest that librarians and archivists are blithely and unthinkingly diving into technology as a solution to every problem. The capital-intensive nature of most technological enterprises, coupled with the generally penurious nature of most libraries and archives, acts as a natural brake on such behavior. Nonetheless, it is clear that a certain fixation on technology has developed within this intellectual and professional community. This is visible in one sense in the intensifying tendency to promote the development of resources at the expense of personal engagement (for instance the growth of distance learning programs via the essentially passive medium of the Internet).[9]

A further matter worth noting is that use of resources from remote locations is much more difficult to track in any meaningful way. The development of the larger digital strategy at Cornell has shown a sensitivity to user needs with a greater strong dimension. The process of construction of their various projects shows a commitment to include users (both scholars and students) to ensure that the products are intellectually substantial and fit into actual patterns of scholarship. The study of substantive use of the CORE pro-

[9] On the implications of this trend see Standish, 2000.

gram conducted by the staff of the Mann Library is a clear manifestation of concern with these issues.

There has been a certain amount of speculation involved with the move to provision of access to scholarly materials via network technology. The conviction that information technology is the coming thing, and the parallel revision of institutional missions to reflect this belief serves to focus attention on the technicalities of the process rather than on its place in larger patterns of scholarly work. Given the significant capital requirements associated with digitization, both in terms of up-front costs (see Puglia 1999, and Besser and Yamashita 1999), as well as in terms of the overhead costs stemming from storage and migration, it is surprising that use tracking in this area has been so limited. In recent years there have been more extensive attempts to develop understandings of the substantive use that is made of digital collections.

3.3. Theory and Practice

Another element of Cornell's strategy has been to take the lead in establishing standards and diffusing the knowledge gained in the course of their work. Their program of publications and workshops has gone much of the way to establishing such standards without the intervention of a standards body at the national level. *Moving Theory into Practice* is, in many respects, the centerpiece in this model. The structure of the work highlights many features of Cornell's style. The presentation revolves around describing practical experiences, rather than providing a cookbook for digital projects. Also included are commentaries from a wide range of practitioners in the field. This lends the book a breadth of perspective that more purely technical accounts lack. This method is exem-

plary of actual practices within the CUL: distributed rather than unified, both internally and externally. Rather than attempting to create a rigid hierarchy through which to control the creation of digital projects, the approach in the CUL has been to stress coordination and consensus. Thus, the last five years have seen the proliferations of offices and projects involved in various aspects of applying network technology to academic environments.

Cornell has built on their institutional commitment to a program-oriented attitude through spreading the fruits of their experience, most prominently in *Digital Imaging for Libraries and Archives* and *Moving Theory into Practice*. These works are outgrowths of the knowledge management strategy that has characterized Cornell's internal methodology. The latter work in particular is evidence of this. Cornell has consistently benefited from the vision of its top players. The early work done in the DPC was facilitated by the commitment of Stuart Lynn, the Vice President for Information Technology. Although his position was outside the library hierarchy, he was alive to the potential applications of network technology in the library setting. Sarah Thomas, who has occupied the position of University Librarian since 1996, has been similarly active in promoting the digital strategy. Her work in her previous position with the Research Libraries group has allowed the staff of the CUL to forge a connection with this organization that has become integral to the prescriptive dimension of their work. Anne R. Kenney's vision for the development of digital assets has been evident in a variety of areas: her work on projects in the DPC and in the Mann Library, her work with Oya Rieger on *Digitization for Libraries and Archives* and *Moving Theory into Practice*, and the NEH funded workshops presented annually by CUL. Finally, the work of Thomas Hickerson in facilitating intra-

institutional cooperation and in building the institutional infrastructure to promote and defend digital assets has been key to the growth of the program-oriented approach. One could certainly site other examples of the contribution of various individuals to the growth of the strategy as a whole. Indeed, one of the CUL's strengths has been the ability to recruit and retain personnel who are highly qualified and who fit ideologically with the program. Another has been a division of function in which the visionary (Kenney) and the pragmatic (Hickerson, Rieger) dimensions have been integrated into a common perspective.

4. Conclusion

The position of leadership that Cornell University has assumed in the field of creation and preservation of digital materials is based in three main factors:

1. *The considered application of new technologies to the traditional mission of the university*. The work done by the DPC, both in development of the CLASS System and in their approach to preservation of, and access to, brittle books indicates a willingness to exploit new technologies without being exploited by them. Projects such as *Utopia*, Invention and Enterprise, and the Core Literature of Agriculture have made effective use of digital and communications technologies to provide access to materials in the context of clearly defined missions and with clearly defined user groups in focus. The seducements of technology for its own sake are considerable, particularly in the context of an institutional with ample resources. Cornell has been exemplary in combining these resources with the vision to create projects that substantively contribute to the educational and scholarly mission of the university.

There are gaps in information that might be derived from user studies in order to give a clear impression of actual patterns of use of some of the collection, although the study of the CORE program is a step toward filling that gap. The proponents of digital libraries have been somewhat reticent about collecting substantive data about the ways in which these collections are used, often preferring to rely on passive measurement tools such as tracking transaction logs. Clearly, actually locating individual users and getting them to agree to participate in a study is a difficult proposition. Whereas the study of the CORE project benefited from the fact that the community of use was strictly localized, the mounting if projects on the Internet means that the community of use is practically unlimited. Anecdotal evidence seems to suggest that people searching for information via the Web do so in part because they value the anonymity that it allows. This will make the collection of data on projects meant for general use quite challenging. Lacking actual data gleaned from use, it is even more important that teachers and scholars who might potentially make use of these products are included in the process of creation. The examples of this mode of operation are legion in the case of Cornell's digital projects.

2. *The recognition of the importance of inter-institutional collaboration in order to achieve productive synergies and economies of scale*. Cornell's partnerships with various corporations (Xerox, Kodak, Sun) have allowed them to effectively position themselves to make use of emerging network technologies. While this collaboration has sometimes had less than optimal long-term results (the situation with the Xerox PhotoCD technology), the benefits have been considerable, both in terms of acquiring technology and in acquiring the technical expertise as a foundation for future developments. In particular, the continuing partnership with Sun Microsystems has provided Cornell with storage and

data management capacity to pursue digital projects on a scale unavailable to most other institutions.

Cornell's collaborative work with other universities, particular their partnership with the University of Michigan in the creation of the MOA 1 project has provided a model of effective cooperation in this area. Each institution pursued the strengths of their particular collections and managed to share technological expertise in an effective way. The cooperative work done to provide access via the Michigan Middleware software package and the use of "dirty" OCR to facilitate text searching are examples of this approach.

3. Finally, the incremental process of growth in the CUL has resolved itself into a mission with broad scope. There is a distinct pattern of employing technology, not as an end in itself or merely as a means of adding prestige to the library system, but to augment the traditional functions of the institution. There has been an extensive commitment by the library staff to plan effectively and to establish standard and benchmarks to facilitate the larger process of integrating technological developments and the mission of the university. Staff in the CUL have contributed extensively to the literature relating to many of the issues discussed above, including the technical aspects of digital to microfilm conversion (and vice versa), the creation and maintenance of metadata, and the significance and potential for further application of digital technologies to the larger mission of libraries and universities (as well as other issues). Perhaps most importantly, there has been an extensive commitment of time and resources in coming to terms with the long-term issues surround the creating and maintenance of digital collections.

Appendix A. IRB Consent Form

Invitation to Participate in a Research Project

Introduction
You have been invited to participate in a case study of the history and structure of digitization projects in the library system at Cornell University. This study is being conducted by John Foster under the supervision of Dr. Helen R. Tibbo in the School of Information and Library Science at the University of North Carolina at Chapel Hill.

Volunteer Status
Your participation in this study is voluntary. You may refuse to participate, or to answer specific questions questions, or withdraw from this study at any time without penalty or any effect on your present or future professional status.
.

Purpose and Procedure
The purpose of this study is to analyze the history and current structure of programs for digitizing historical and cultural heritage materials in the library system at Cornell University. The goal of the study is to provide models and guidelines for the development of similar programs at other institutions. If you agree to participate, you will be interviewed about factual and technical matters relating to the project mentioned above.

Time Commitment
The entire procedure will take approximately one hour to complete.

Risks
No more than minimal risks are expected during any phase of this study. Participants will, in the main, be asked to clarify historical information and technical and practical matter that fall within the normal conduct of the participant's work.

Benefits
Benefits of this study will help to develop other programs of this sort and to reduce the potential for lost resources due to insufficient planning or modeling, or inefficient practices in future programs.

Confidentiality
All information obtained in this study is strictly confidential. Your name will NOT be used for any portion of this study unless specific and explicit permission is obtained beforehand.

For Further Information
Any questions that you may have about this study can be answered by John Foster (919 928 0178, fostj@ils.unc.edu), or his faculty sponsor Dr. Helen R. Tibbo (919 962 8063, Tibbo@ils.unc.edu). Academic Affairs Institutional Review Board oversees research conducted at the University of North Carolina at Chapel Hill that involves human participants. Questions or problems regarding your rights

as a participant should be directed to Barbara Goldman, Chair, AA-IRB, 201 Bynum Hall, CB #4100, Chapel Hill, NC 27599, (919 962-7761, aa-irb@unc.edu).

Before You Sign This Document

By signing below, you are agreeing to participate in a research study. Be sure that any questions have been answered to your satisfaction and that you have a thorough understanding of the study. If you agree to participate in this study, a copy of this document will be given to you.

Participant's Signature:

Date:

Print name:

Experimenter's Signature:

Date:

Appendix B. Sample Questions

1. Describe your work in digitization programs at Cornell University.

2. What educational and professional experiences prepared you for work in this area?

3. What approaches have been used to solicit funding from internal sources for digitizing particular collections?

4. What approaches have been used to solicit funding from external sources (both governmental and private) for particular digitization projects?

5. How do the digitization projects fit into the institutional structure of the library system?

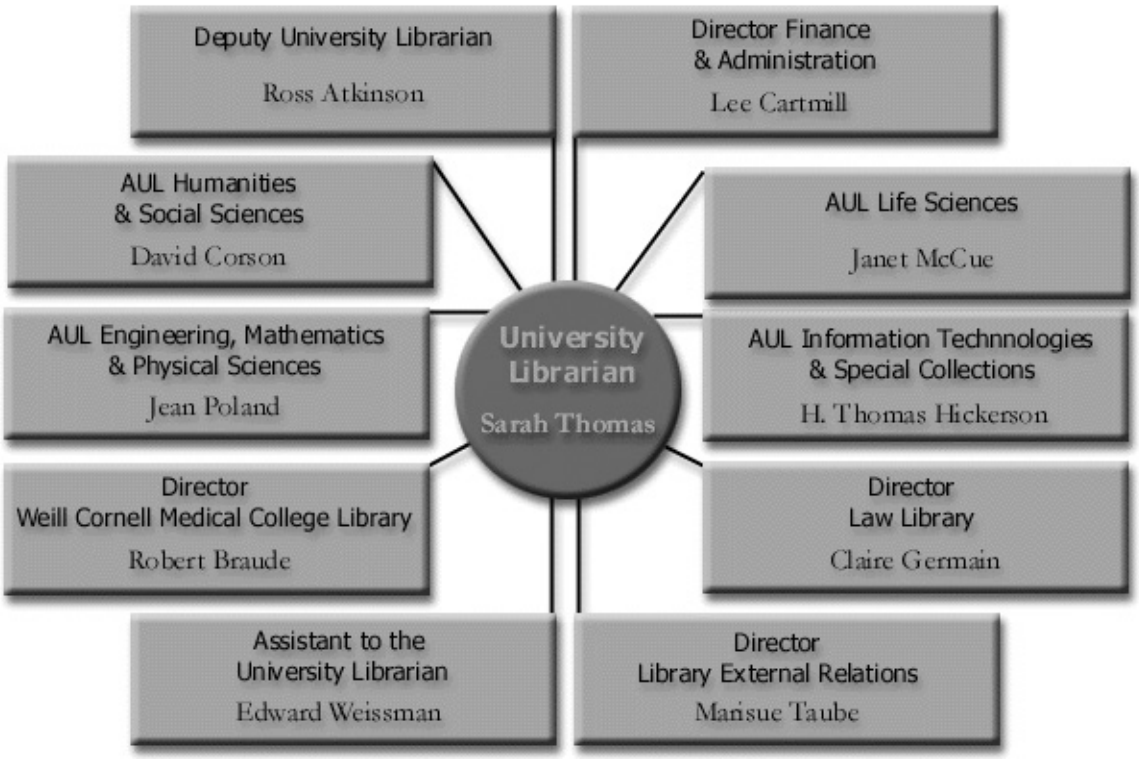
6. What sorts of factors structure decisions about which and what portions of collections will be selected for digitization?

7. What provisions have been made for long-term preservation of, and access to, digitized collections?

8. What sorts of technological factors affect planning for future projects and maintenance of current collections?

9. How have the rapidity of cycles technological obsolescence affected the development of digitization programs?

Appendix C. Organizational Chart



(The original version of this chart is available at the Cornell University Library Information site: http://campusgw.library.cornell.edu/cgi-bin/dj.cgi?section=about&URL=about/about.html

Source List

Committee on Preservation and Access. 1996.Preserving digital information: Report of the Taskforce on Archiving of Digital Information. Mountain View, Ca.: Research Libraries Group. (http://www.rlg.org/ArchTF/tfadi.index.htm)

Berger, 1996.Preservation and digitization: Principles, practice, and policies. Paper presented at conference. National Preservation Office (UK) 1996 Annual Conference. 3-5 September, at the University of York.

Besser, Howard, and Robert Yamashita. 1999. "The cost of digital image distribution: The social and economic implications of the production, distribution and usage of image data", Report produced by the School of Information Management & Systems, University of California, Berkeley.

Castells, Manuel. 1996. *The rise of the network society*. Cambridge, Mass.: Blackwell Publishers.

Elaine D. Engst and H. Thomas Hickerson, 1998. Developing collaborative structures fir expanding the use of university collections in teaching and research: A case study. Chicago, Ill.: Society of American Archivist

Entlich, et. al., 1996. Testing a digital library: User response to the CORE project. *Library Hi Tech*, 14(4), 99-118.

Feenberg, Andrew. 1991. *Critical theory of technology*. New York: Oxford University Press.

National Office of Digital Archiving (UK). 1997. Responsibility for digital archiving and long term access to digital data. *JISC/NPO Studies on the Preservation of Electronic Materials*. (http://www.ukoln.ac.uk/services/papers/bl/jisc-npo67/digital-preservation.html).

Hickerson, H. Thomas. 1997. Realizing new means: Networked access to research collections. *Museums and the Web: An International Conference*. 16-19 March, 1997. (http://www.archimuse.com/mw97/speak/hickerso.htm).

Kenney Anne R., and Oya Rieger. 2000. *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, Ca: Research Libraries Group.

Kenney, Anne R., and Stephen Chapman. 1996. *Digital Imaging for Libraries and Archives*. Ithaca, New York: DPC, Cornell University Library.

Kenney, Anne R. 1993. Digital-to-microfilm conversion: An interim preservation solution. *Library Resources & Technical Services*, 13 (Oct. 1993), 381-401.

Kenney, et. al., 2001. Preserving Cornell's digital image collections: Implementing an archival strategy. A submitted to the Institute for Museum and Library Services. (http://www.library.cornell.edu/imls/index.htm).

Olsen, Jan. 1994. Cornell University's Albert R. Mann Library: A prototype for today's electronic library. *Library Hi- Tech*, 12(3), 32-7.

Puglia, Steven. 1999. `The costs of digital imaging projects', *RLG DigiNews* (3 (5) 15 October 1999, (http://www.rlg.org/preserv/diginews/diginews3-5.html).

Rogers, Everett M. 1962/1995. *The diffusion of innovations*. 4th ed. New York: Free Press.

Vidal, Noni Korf. 1997. *Utopia's* approach to user-centered design. *Museums and the Web: An International Conference*. 16-19 March, 1997. (http://www.archimuse.com/mw97/speak/vidal.htm).