The Journal for the Society of North Carolina Archivists

Contents

SAA And SNCA: Communities of Archivists and Interests

Steven L. Hensen
Duke University

North Carolina State University KATHERINE M. WISSER

Mr. Hensen delivered the following essay in the keynote address to the Society of North Carolina Archivists on March 15, 2002 at The University of North Carolina at Greensboro, host for the Spring 2002 SNCA Meeting.

When I first came to Duke University in August of 1986, Robert Byrd, Director of the Rare Book, Manuscript, and Special Collections Library, wanted me to become more involved in local archival and manuscript matters. Towards this end, the first thing I did was to join the Society of North Carolina Archivists (SNCA) and attend the fall meeting that year in Greensboro. As I recall, it was a wonderful meeting and I had the opportunity to meet a number of North Carolina colleagues, many of whom I have worked with on various matters over the years. However, except for a meeting that was held at North Carolina State University a few years ago and another that was held at Duke, I confess that I have not attended many meetings nor maintained my membership. While I regret my lack of involvement, I suppose I could make excuses. Instead it struck me that maybe I could use this opportunity to think about and explore the relationship between regional archival associations and the Society of American Archivists (SAA).

At first blush, it might not seem that my own case presents an especially useful launching pad for such a discussion. When I arrived at Duke, I had already been involved in several national efforts. I was then in the midst of preparing the second edition of *Archives, Personal Papers, and Manuscripts*; I was also a member of SAA's Working Group on Standards for Archival Description (otherwise known as WGSAD), which was engaged in the process of exploring the outer boundaries of the universe of standards that might have some archival application or interest. I subsequently became involved in teaching workshops in both MARC-AMC and archival cataloging (the very first "official" MARC-AMC workshop given by SAA was done in Raleigh at the Department of Archives and History). Without detailing all my professional activities on behalf of both SAA and the profession generally, suffice it to say that it has continued unabated.

What I would like to do first of all is to discuss the important role that regional archival associations play, not only to their

necessity of electronic records management. Legal developments for electronic records can be difficult to traverse, but this chapter provides a clear summary of the on-going developments. Most importantly, this essay removes the legal framework from item-oriented consideration to address more fundamental issues of information content and knowledge.

Bruce Dearstyne's summary essay, "Riding the Lightning: Strategies for Electronic Records and Archives Programs," provides the perfect wrap-up for the complex topics dealt with throughout the text. Dearstyne includes checklists and outlines nine strategies that will be useful for archivists and records managers. These strategies are not meant to be implemented in isolation, but rather in consort with each other to strengthen the mission and the effectiveness of maintaining the human record stored on bytes.

One drawback of this work is the sole reliance on government institutions as its sample base. As electronic or machine-readable records began as governmental efforts, it is not surprising that the projects discussed represent government archives and records management. However, translating large governmental efforts effectively to smaller institutional parameters can often be challenging. While this text provides excellent insight into the governmental struggles and successes, it does not provide guidelines appropriate for smaller institutions. Despite this limitation, several themes are woven throughout, which apply to all types of repositories, regardless of size or governing institution: the need for direct involvement in the design process of electronic record generation, the need for practical technological savvy, the need for a clear understanding of the repository's inherent legal responsibilities, and the need for a better definition of what is to be saved. All of these themes cut across institutional and governmental lines, and should be considered at any repository grappling with the challenges of electronic record management and retention.

Each essay is well documented with print and online citations, enabling readers to continue a thread from a topic in the book. An index provides limited subject and name access, but given the book's size and segmentation (each essay representing a different aspect or case study), as well as descriptive subtitles, navigation throughout the text is easy. This volume is a useful reference tool for all repositories and an excellent launching point for archivists or records managers responsible for establishing an electronic records management program.

members and their regions, but also to the larger national scene. I would then like to examine several areas of national and international archival activity and see not only what "trickles down," but also what "bubbles up."

SNCA is part of a large network of regional archival associations. Last October I was invited to give the keynote address at the 30th anniversary meeting of the New England Archivists, a group I helped found in my first professional position at Yale University. When I left Yale for the Library of Congress, I became active in the Mid-Atlantic Region Archives Conference (MARAC). I have since had occasion to participate in various ways in meetings of the Council of Inter-Mountain Archivists, the Society of California Archivists, the Midwest Archives Conference, the Society of Southwest Archivists, the Society of Georgia Archivists, the Society of Florida Archivists, the Northwest Archivists, SNCA, the Society of Ohio Archivists, the Association of Canadian Archivists and others.

According to the SAA Web site, there are currently 55 regional or provincial archival associations in the U.S. and Canada (44 in the U.S. and 11 in Canada) and 10 national groups, including the two primary national associations, Academy of Certified Archivists (ACA) and SAA.

Given the large number of various flavors of regional associations and their rapid growth over the past 30 years and the inherent complexity of their relationship with respective national associations, it is perhaps surprising that there has been very little research or writing on the subject. However, three articles will provide some perspective on this phenomenon. These articles include a 1983 piece in the *American Archivist* by Patrick Quinn of Northwestern University, a 1988 article in the *American Archivist* by Tim Ericson, and a 1991 piece by Bill Maher in the *Midwestern Archivist*. It is perhaps instructive and certainly interesting that all three of these individuals come out of MAC—the Midwest Archives Conference, the largest regional association in the country as well as the region in which the largest concentration of SAA members can be found.

Pat Quinn's 1983 article identifies the environment in which many of the early regional organizations sprung up. In 1972 there were only five regional organizations, but by the end of that year there were eleven. This growth can be attributed in part to the report of SAA's Committee on the Seventies, which in turn was an outgrowth of the social ferment of the 1960s manifesting itself in the affairs of SAA. This report contained a number of recommendations on democratizing the Society,

opening it up more to women and minorities, and in general making the society more relevant and less conservative. It was, in effect, the fissure between the original, more "scholarly" association that had its roots in the historical profession and a new, more modern society, which recognized that there were "archival" issues distinct from those of historical research and that a Society of American Archivists ought to focus more on matters of the theory and practice of "archival economy," as it was so quaintly known in those days, as well as the social and political milieu in which archives functioned.

One of the more logical responses to these efforts to make things more democratic and responsive was to focus locally. After all, what better way to gain some purchase on the issues that concern *you* than to form your own organization. However, as Quinn notes,

> *the first regionals were organized very rapidly during a particularly fermentative period in the history of the American archival profession. The relationship between these newly founded organizations and the SAA had yet to be determined. It would be an understatement to say that there existed in the early 1970s a climate of mutual apprehension between many of the founding members of the new organizations and an important segment of archivists traditionally and exclusively committed to the SAA.*[1]

Quinn quotes an editorial from the Midwest Archives Conference *Newsletter* by way of illustrating this tension more bluntly:

> *Junior partner? Constituent Member? Competitor? These are a few of the possible forms that the future relationship between MAC and the SAA can take. At MAC's founding meeting last May in Chicago an attitude of suspicion and in some instances outright hostility toward the SAA was evidenced by a sizeable number of those in attendance. On the other hand it became clear at the SAA Annual Meeting in Columbus [my first, I might add] that a number of long-time SAA members viewed the emergence of regional archival organizations such as MAC as a distinct threat to the SAA. At this point the question of how the regional groups will relate to SAA is open. Those who are less than enthusiastic about the SAA maintain that it is an elitist organization composed in the main of administrators who have little concern for the problems of the secondary level personnel, archivists from small institutions, and para-professionals. Other criticisms of the SAA range from the contention that it is a "do-nothing organization"*

nent record of that workplace. It should not be surprising that electronic generation and maintenance of records have become an important aspect of the twenty-first century archivist. Within this framework of innovation, "What Is an Electronic Record," by Roy C. Turnbaugh, details the cultural and legal definitions of "record," and how that is problematic for electronic records. Turnbaugh's solution is a continuum on which records, information and data interact and a tiered management system works in consort with this continuum.

"Implementing Requirement for Recordkeeping: Moving from Theory to Practice," by Timothy A. Slavin, as its subtitle suggests, shifts from a theoretical to a practical approach in the maintenance of electronic records. Through the discovery and implementation of functional requirements, Slavin points to three lessons learned by the Delaware State Archives, including an assessment of the role of record keeping in an organization, a better understanding of the functional requirements and a knowledge of the development of systems, and a need for technological skills on the part of the archivist or records manager. This essay serves as the opening of the practical experiences section of the book and allows the reader to obtain a growing understanding of a complex issue, while at the same time become exposed to real-life experiences in the repositories.

The next three essays each outline a specific case study, but the necessity of all three studies is demonstrated by the scope and depth at which they are covered. In "Obstacles and Opportunities: A Strategic Approach to Electronic Records," Robert Horton highlights the need for collaboration between systems developers and designers and records management professionals by exploring Minnesota's Information Systems Concept. Next, John McDonald discusses the challenges facing the Canadian National Archives as they try to act as facilitators for subsidiary repositories in "Government On-line and Electronic Records: The Role of the National Archives of Canada." Finally, Alan S. Kowlowitz discusses New York State's e-government initiatives and their impact on governmental archives and records management institutions in "Playing the Electronic Angles and Working the Digital Seams: The Challenge and Opportunities State Electronic Government Initiatives Present to State Archival and Records Management Programs."

Following these case studies, Lee S. Strickland outlines the legal ramifications of records management through which archivists and records managers must navigate in "The Law of Electronic Information: Burgeoning Mandates and Issues." The various laws provide an essential backdrop and illustrate the

importantly, Moss implies), to "reassert the primacy of the original source" (12). To begin traveling this path, archivists must learn how people (not just academics) study, interpret, and make history: how they do research, how they view history, and how they see our role in their endeavors. We must consider how we can encourage the production of archival research and the sure-footed development of our profession. We must learn how others do these things in other places, even if this is another country and a foreign one for us. This little volume is an admirable step in that direction.

Duke University   PAULA JEANNET MANGIAFICO

Dearstyne, Bruce, ed. *Effective Approaches for Managing Electronic Records and Archives.* Lanham, MD: The Scarecrow Press, Inc., 2002. x, 167 pp. ISBN 0810842009.

Electronic records management and retention is one of the most critical issues impacting archivists and records management professionals today. Increasing numbers of records are born digital with a minefield of legal and theoretical ramifications for the profession. This volume, edited by Bruce Dearstyne, offers an introduction to the challenges repositories face in trying to establish an electronic records management program. By introducing a theoretical framework and proposing several different approaches through concrete examples, it clearly presents the issues that archivists and records managers are facing in a world of dynamic, technologically savvy record creation. Inherent in this process is a re-evaluation of standard archival perceptions in a world that is creating more and more "records" with less and less tangibility.

The first two essays establish the theoretical framework nicely. We learn in Richard E. Barry's "Technology and the Transformation of the Workplace: Lessons Learned Traveling Down the Garden Path" that while technology has become a buzzword that impresses grant-awarding agencies, it is really nothing new. Technology has consistently impacted the workplace, and thus also impacted those that create and maintain the perma-

to the allegation that it is impossible for less well-heeled archivists to attend functions because of the prohibitive costs.

Some SAA members are equally apprehensive about the regional organizations. They view MAC and an attempt to circumvent, undermine, and even destroy the SAA. They see the new organization as a challenge to the authority and prestige of the SAA.[2]

The editorial concludes on a more positive note:

In sum, MAC and the SAA are not and should not be competitors. MAC and the SAA should complement and reinforce each other's functions. A spirit of cooperation and reciprocal trust should be fostered between MAC and the SAA. MAC should develop a working relationship with the SAA wherever possible but should not affiliate either as an institutional or constituent member at this point. A harmonious relationship ultimately will evolve as each organization defines its own role and recognizes the other organization's merits. The benefits of mutual respect and cooperation will eventually prevail.[3]

Before I discuss how this dynamic has evolved over the last twenty years, I want to look at Bill Maher's exhaustive examination of some of these issues in his article "Cooperative Competitors: Local, State, and National Archival Associations."[4] In this article Maher stakes out the position that "Archivists' professional development is critically dependent on a diverse system of archival organizations. Through meetings, publications, and committee work, these organizations provide members with education, experience, and legitimization. Regional archival organizations are enviably positioned to fulfill archivists' educational and socialization needs."[5]

He starts by defining how they fulfill those needs. "Archival organizations play a key role by serving as advocates for archivists and providing a forum for discussion of relations with collateral professionals, such as librarians, records managers, historians, and museum specialists." Further, he points out with respect to the variety of organizations that have emerged to meet the more specialized needs of archivists, whether those special needs be geographically focused, by type of repository, or by material type or subject focused, that such diversity, "if kept in balance…can move the archival profession forward by improving both its techniques and its responsiveness to specialist constituent groups."[7]

He concludes his general comments by noting "…the growth of regional, state, and local archival associations is the result of a commitment to democracy, as much as it is a manifestation of the overall growth in size and complexity of archival work."[8] He then offers a fairly complex analysis of the generic goals, activities, and outcomes of organizations, particularly as they relate to regional archival associations as embodied especially in MAC.

While these statements represent fairly standard organizational analysis, it is useful to review his four major points:

- *Once organizations are firmly established, they are driven by a desire to survive.*

- *Organizations…provide legitimacy and credibility for the individual professional.*

- *Organizations fulfill a role as an exercise and proving ground through volunteer opportunities which develop and refine skills in critical thinking, writing and presentation, interpersonal relations, program planning and budgeting.*

- *Organizations fulfill a complex socializing and humanizing role through meetings and committee work which can create an environment in which the technical details of archival work can be mixed with the personal interest of members.*[9]

It is fairly easy to see how SNCA and SAA (or MAC, MARAC, etc.) fit into these organizational models. Moreover, they all continue to refine the ways in which they represent and meet the needs of their members. While SNCA did not emerge out of the ferment of the 70s or in direct reaction to SAA, it did arise out of a specific need in the North Carolina region (given the extraordinary geographical breadth of our state, let us think of ourselves as a "regional" organization, rather than a strictly local or state one). These reasons are clearly articulated in Bill Maher's generalizations regarding MAC and remain as valid today as they did nearly twenty years ago at SNCA's founding.

I would like to suggest, however, that the dynamic between SAA and SNCA—and indeed *all* of the regional and local archival associations—has changed. Furthermore, these changes do not in any way diminish the importance and relevance of the regional associations, but instead offer opportunities to become even stronger and more relevant. This new dynamic has emerged because, unlike the old "do nothing" days to which Pat

tion, often colored by a fetishization of the physical artifact, in turn encourages users to fetishize archival materials — in some cases, leading to a life of criminal obsession. In another strong essay, Judy Dicken examines the collecting practices that have either helped or hurt the corpus of twentieth-century literary archives in Britain, discussing the problems raised by appraisal methods, modern formats, competition among collectors, and privacy issues; in the end she asks whether Britain can afford not to have a national collecting strategy in place. In all of these writings, even in the study of local Anglican church history, American archivists will recognize the many familiar problems and concerns that color their world of work and archival thought.

The student essays are worth getting to know, but the real piece of interest in this collection is Michael Moss' "The Scent of the Slow Hound and the Snap of a Bull-dog: The Place of Research in the Archival Profession." Granted, it is a somewhat disjointed and dense essay, but if one wishes to learn something about historians, post-modernists, British archivists, diplomatics, epistemology, and which areas are crying out for archival research, in only twelve thought-provoking pages, read this essay. The reader will learn something about Dickens as well. Do not be daunted by the occasional German word (the reviewer was totally frozen to her chair by the sudden appearance of *Geschichteswissenschaft*). Such things can be looked up or given educated guesses. Where else can one get even a mention of a fascinating debate between post-modernists and archivists, a debate over authenticity and truth which directly pertains to our profession, but which has not uttered a peep here in the States (or at least if it has, it really was a peep). In these pages one reads a compelling defense of our profession and its loyalty to the importance of primary sources, and a warning to the profession that the debate is going on without us.

Is this important?, one asks in desperation, glancing from a volume of essays to piles of desk paper or finding aids awaiting encoding. Is it even within our scope of activity to dwell on theoretical issues and even philosophical teasers about truth, permanence, and uniqueness? Yes, it is, if we are in fact committed to becoming a profession. As Moss writes, we cannot rely on historians or anyone else to defend us. No matter what kind of archives we work in, all of us at the very least must present our professional goals and foundational practices clearly and concisely to others who come in contact with our archives and rare materials, and be able to explain their significance, not only so that we will be recognized as professionals, but also (and more

B O O K   R E V I E W S

New Directions in Archival Research. Margaret Procter and C. P. Lewis, editors. Liverpool : Liverpool University Centre for Archival Studies (LUCAS), 2000. 145 pp. Introduction, b/w illus. ISBN 0953796302.

This trim paperback may look positively tiny when viewed on the shelf next to the recently published 657-page collection of essays entitled *American Archival Studies: Readings in Theory and Practice*, but anyone interested in the future of archival theory and research should take it down from the shelf and give it an evening.

The editors, one a records manager at the University of Liverpool, the other a Lecturer in History at the same institution, have put together a set of five essays originally presented as dissertations by graduate students matriculating from the British system of archival studies (in this case, the program at LUCAS), prefaced by an introductory essay on the place of archival research in the world by Michael Moss, University Archivist at Glasgow University and holder of the only British Chair in Archival Studies.

Although they form an unevenly balanced collection, the six essays all offer something of interest, though it is very unlikely all six will capture the reader's interest equally. The least likely to succeed is the essay on the history of the Church of St. Mary in Disley (Cheshire); although this inquiry into local Anglican church history is very well documented and would be fascinating to someone with a passion for such history, it is too out-of-sync with the rest of the volume, which is oriented towards theoretical issues. It does, however, offer a detailed case study underscoring the importance of diplomatics, which makes it resonate in a satisfactory way with the much more theoretical essay by Michael Moss. The other essays offer various levels of academic work: some are decidedly polished pieces of critical thinking, while others are less successful, in particular the essay that asks the question, "Should archivists be considered professionals?" (we should not be so burdened, comes the unfortunate answer). But these writings range over a wide-open territory of archival research and ask provocative questions. In a fascinating essay, "The Fetish of the Document: An Exploration of Attitudes Towards Archives," Helen Wood studies how archivists conceive of the nature of archival materials and how archivists present them to users, and whether this presenta-

Quinn referred, SAA is now stronger and more active than ever. It is more active not only because of the strength and commitment of its members, but because of the overall strength and commitment of the profession, as encompassed by *all* archivists, whether members of SAA, or SNCA or any of the other many regional and local organizations. There is a synergy at work in which the totality is much greater than the sum of the individual parts.

I would like to explore the changes in SAA over the past twenty years and what those changes mean for the regional archival organizations, and conversely, what the regionals now mean to SAA—in short, to examine the "percolator" effect I invoked at the outset.

I laughed when I read the MAC editorial that referred to SAA as a "do-nothing" elite organization. I remember very well those days, and I am not so sure they were much off the mark in that assessment. SAA is hardly that today. Membership is growing and is much more inclusive, not only with respect to professionals working at various levels within their respective institutions, but also in terms of a much more inclusive definition of what constitutes an archivist. Many of the so-called "allied" professionals referred to in the past—the librarians and museum staff, for example—are now very much a part of the archival community and have found a very comfortable, congenial, and relevant home within SAA. The organization addresses some of the high level concerns of the profession, including national information policy issues, preservation, and digital access, and it also has roundtable and committees that deal with some very specific concerns.

Quite beyond the "bigger tent" that is SAA today (and I mean that in both its size and inclusiveness connotations), I want to focus on three specific areas of activity in which SAA has been particularly active, and which all have immediate impact on American archivists, archivists in North Carolina, and SNCA members. These are public advocacy, standards—especially descriptive standards and education.

Advocacy: Pat Quinn, in the article I quoted earlier, mentioned the frustration that many archivists felt in the 1970s with SAA's unwillingness to get involved in public issues, and how this frustration in some ways contributed to the growth of some of the regional associations. While it is true that SAA was famously hesitant and even downright unwilling to make public pronouncements on the public issues of the day, it must be said in SAA's defense that they were often being pres-

sured to take positions on issues that had nothing directly to do with archives, such as the war in Vietnam, the bombing of Cambodia, and the entire range of radical, political issues that had so galvanized public opinion. At the same time, it was recognized that the "social fabric" issues that grew out of this same atmosphere of ferment, especially those relating to gender and racial equality, were very much of concern to an organization like SAA. Like most institutions of that period, it carried much of the social and cultural baggage that was part of the times. To the Society's credit, it did form the Committee of the Seventies to examine these issues. The report the committee issued resulted in many important changes in the structure of the Society and in general raised our collective consciousness on the social and cultural dimensions of an organization which had heretofore been seen largely in one-dimensional terms. The extent to which SAA, and indeed many of the regional associations, now reflect the full racial, gender, and sexual identity diversity of its members is, I believe, directly related to how we as a profession confronted and resolved those issues.

At the same time SAA as an organization was also generally disinclined to take public positions on archival matters. This was in part because of the large influence that the National Archives still held in SAA governance. In a government agency the workplace culture is such that opinions and pronouncements of any sort are actively discouraged. To the extent that they are occasionally allowed, they must always be framed in resounding and emphatic disclaimers: along the lines of "the opinions expressed herein are mine and mine alone and do not in any way reflect the official or unofficial position of anyone else in my office, department, or agency, nor do they reflect the views of any members, living or dead, of this, any preceding or future government or officer of such government..." Such an atmosphere does not exactly encourage "untrammeled inquiry" or "fearless sifting and winnowing," to quote the foundation documents of the University of Wisconsin.

Furthermore, the Society was a much different organization in those days with different expectations from its membership. The major feature of membership in those days was the timely receipt of the quarterly *American Archivist,* convening an annual meeting in a place that was not too expensive and had a sufficiency of local archivists to serve on the host committee, and desultory involvement in some committee work. You must remember that being an archivist was not always as "cool" or "out there" as it is today. The profession was characterized if not by a certain timidity, then certainly a low public profile. It was also still dominated to some degree by historians, who frankly were more interested in

(http://www.library.cornell.edu/imls/index.htm).

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).

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).

access to documentation than in proper procedures for managing it.

It really was not until the 1980s when the archival community started to come together and feel some sense of identity and pride in what they were doing and how they were doing it. I would argue that much of this was the direct result of the emergence of distinctly archival standards. Archivists also started releasing a number of white papers on such topics as preservation, access, restrictions, appraisal, and other broader issues. These white papers had the effect of laying a fundamental foundation to develop policy positions, while also empowering archivists to speak on issues without feeling like they were going out on a limb or doing so in a vacuum.

In the early 1990s this came to a head when the SAA Council formally opposed the nomination of John Carlin as Archivist of the United States and testified before Congress on this matter. Fortunately, Governor Carlin is a forgiving person for, as you know, he was confirmed; in fact, he attended his first SAA Council meeting two months after that confirmation. Since then, we have enjoyed a warm working relationship with him and he goes out of his way to meet with or speak to SAA Council whenever it's in session. He understands that our opposition was not personal, but principled. The profession is now at a point where it is comfortable with its standards and principles and is no longer hesitant about taking an emphatic public position when those principles are threatened or compromised. Just over the past few months SAA has been at the forefront of public championing of the archival ethic—in the case of the President's illegal Executive Order on access to presidential papers, the anomalous transfer of Texas gubernatorial records out of state custody, and the equally questionable transfer of Mayor Giuliani's mayoral records from public custody. We have also joined in an *amicus* brief in Eldred v. Ashcroft before the Supreme Court challenging the validity of a copyright law that takes no account of the rights of the public or of traditional notions of "fair use." These activities are ones, which SAA ought rightfully be involved on behalf of the entire profession, and they validate and strengthen both the individual archivist and the profession. I have had no prouder moments in the thirty plus years of my professional life than I had in speaking and writing publicly and representing the Society over the past few months in an effort to protect our principles and to raise public awareness of the importance of archivists and archives.

Standards: I believe that the development of descriptive and

other standards has done the most to give the archival profession legitimacy, confidence, and a sense of cohesion that allows us to speak amongst ourselves with one voice, and to speak clearly and cogently to the world at large about the content and value of both our holdings and our work.

For many of you, the work of the National Information System Task Force (or NISTF) is part of archival pre-history—somewhere back there with the invention of acid-free folders and stainless steel paper clips.

The work of the National Information Systems Task Force[10] began in 1977 under the auspices of an National Endowment for the Humanities (NEH) grant to SAA. It was initially an attempt to reconcile a territorial dispute between the *National Union Catalog of Manuscript Collections*, published by the Library of Congress, and the *Directory of Archives and Manuscript Repositories in the United States*, published by the National Historical Publications and Records Commission. Not surprisingly, it quickly became clear that there were larger descriptive issues at stake. The work of the task force was a lengthy and often contentious process. Among the difficult issues with which NISTF grappled was the apparent hostility felt by many in the archival community towards anything that smacked of librarianship, and the firm belief that since archives were unique, they required unique approaches, and standards (especially library standards) could thus never be applied. Add to this mix the sentiment that the methodologies and principles of archivists were somehow fundamentally different than those employed by their more library-oriented "manuscript curator" colleagues — perhaps a vestige of the "archives/historical manuscripts" dichotomy dating back to Sir Hilary Jenkinson in the early 20th century. As I just noted, our collections were unique; we knew them better than anyone and we knew what our users required. In many cases, we had developed long-standing internal systems for the administration and use that seemed to work just fine and the idea of other archivists -- let alone librarians -- telling us how to work was plainly impertinent and unacceptable.

Thus, NISTF had to address this resistance by first determining whether there was any substance in the long-standing dispute between "archivists" and "manuscript curators" over various matters of theory and practice. Towards this end, Elaine Engst of Cornell University conducted a thorough study of descriptive practices in a wide variety of repositories. Her unpublished report, "Standard Elements for the Description of Archives and Manuscript Collections,"[11] clearly demonstrated

Source List

Berger, Barbara, 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.

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)

Engst, Elaine D. and H. Thomas Hickerson, 1998. Developing collaborative structures for 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.

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. 1993. "Digital-to-Microfilm Conversion: An Interim Preservation Solution. *Library Resources & Technical Services*, 13 (Oct. 1993), 381-401.

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, et. al., 2001. "Preserving Cornell's Digital Image Collections: Implementing an Archival Strategy," A report submitted to the Institute for Museum and Library Services.

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.

[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" (1998). This point might be extended to librarians in general.

[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 (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.

[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.

[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.

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

[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's article in the issue of *Library Hi Tech* devoted to the Mann Library (1984).

[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.

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

that there was no significant difference between the descriptive approaches of these two groups and that, in the words of Tom Hickerson, "there are common methods of archival description which could be integrated into a broadly applicable set of standards."[12] More importantly, however, Engst's report helped lay an essential foundation for the subsequent development of a unified data elements dictionary, which was the first step on the road to adapting the MARC format for the purpose of describing (or, more specifically, "cataloging") archives and manuscripts. At the time this work was going on, it was not altogether clear to the members of the task force that it was possible or desirable to describe these materials in the same systems used for describing other library materials, but it was already obvious that the superstructure used by the library (the MARC formats) could easily be adapted to archival purposes. The result was the US MARC Format for Archival and Manuscripts Control (MARC-AMC).

Thus, this work accomplished three major things in bringing the larger archival community together and giving them both internal and external legitimacy.. First, it established a common descriptive vocabulary among what had heretofore been seen as distinctly separate branches of archivy.

Second, it provided archivists with the opportunity to integrate their holdings into what had formerly been systems entirely devoted to bibliographic processing, giving archival materials new visibility and legitimacy. What had formerly been buried in such printed resources as Hamer's *Guide to Archives and Manuscripts in the United States*[13] and the *National Union Catalog of Manuscript Collections* was now thoroughly commingled with other resources. This bibliographic propinquity helped librarians, archivists and scholars realize that these materials were simply different physical formats for what was essentially a vast seamless web of cultural resources, and that archival materials were very much a part of this web. Accordingly, the process of converting the bibliographic networks into cultural resources databases began.

Third, and perhaps most important, the realization that we were all custodians and gatekeepers of cultural resources brought us together in new ways and built new communities out of what had heretofore been seen as disparate—even competing—institutions. On a national/local axis, it created a standard that was developed, maintained, and promulgated by SAA, but was embraced at the local and regional level. Moreover, many of the MARC-AMC workshops that were offered by SAA were given in conjunction with regional or state archival

association meetings.

The development of archival cataloging rules and the more recent emergence of Encoded Archival Description (EAD) have been natural and related extensions of the phenomena started by NISTF. Nationally focused and funded efforts designed to bring greater rigor and standards to bear on that most basic of archival activity—description. On a strictly North Carolina local level, these efforts have had a profound effect, spawning the North Carolina-EAD project in which Duke University, North Carolina State University, The University of North Carolina at Chapel Hill, and the North Carolina Office of Archives and History are all working collaboratively to develop a commonly agreed-upon set of encoding standards and encoding tool-kits that will facilitate what we hope will be virtually automatic encoding of finding aids for archives, manuscripts, and a wide variety of other cultural resources. Running virtually parallel to this project and built upon the same foundation of archival standards is the breathtakingly ambitious "NC-ECHO" or "Exploring Cultural Heritage Online." This project, which is working towards developing an online database of institutional and holdings information from over 700 cultural repositories in our state. These repositories are roughly defined as any place that has "stuff" and is open for the public to use it. Such projects represent the absolute pinnacle of possibilities when national and regional archival efforts are functioning with perfect synergy.

Education: Finally, let me turn briefly to the subject of education and training. As noted in the Maher and Quinn articles, education is one area in which there is more overlap between regional and national activity. It is also the area of potentially greatest collaboration and cooperation. To the best of my understanding, there are two major areas in which SAA is engaged in educational and training issues. The first and most obvious is through the workshops and training offered by the SAA Education Office. SAA has been offering a rich variety of workshops to members and non-members for several decades now. While many of these are held in conjunction with SAA annual meetings, the greater percentage are given around the country (and world, in some cases) at various places—often in conjunction with meetings of regional or local associations. As a veteran of giving nearly forty workshops myself, I can tell you that a substantial number of these were held as part of meetings of such organizations as MAC, the Society of Florida Archivists, the Council of Intermountain Archivists, the Society of Georgia Archivists, the Society of Southwestern Archivists, the Association of Canadian Archivists, and the Society of California Archivists.

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 Kodak Photo CD 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 standards 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 surrounding the creation and maintenance of digital collections.

Notes

[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

in which the visionary (Kenney) and the pragmatic (Hickerson, Rieger) dimensions have been integrated into a common perspective.

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 institution 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 of 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. The collection of data on projects meant for general use will be 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.

This kind of collaboration between SAA and the regionals is especially fruitful because the workshops often offer training in specialized areas or emerging technologies and standards where there may often be very little local expertise. In addition, the content of these workshops can be, and often is, tailored somewhat to the special needs of archivists in a certain locale or region in order to relate the subject matter to their particular institutional or regional setting. Furthermore, the experience of taking such workshops with local colleagues and friends gives the archivists a better opportunity to bond over the topic and contributes towards creating a critical mass of expertise in a region which can then be leveraged for use in future collaborative projects.

Education is also an area in which SAA is especially susceptible to suggestions and new ideas from regional associations and groups. Not only does education fulfill one of the primary mandates of SAA, but it is also a good way to develop and encourage new membership.

The other means by which SAA deals with the subject of education is through its Committee on Education and Professional Development (CEPD) and its Archival Educators Roundtable. While this group directly advises the Education Office, they have also directly addressed the larger questions surrounding appropriate curriculum and training for professional archivists—both within the setting of formal academic programs and in pre- and post-professional training. The guidelines that govern the activity of this group are worth examining:

- Education and professional development are essential to the continued advancement of the profession;

- Graduate education and continuing education need to be addressed in a coordinated manner, ensuring that developments in both areas are based on a common understanding of the needs of the profession at all levels;

- Education and professional development offerings must be responsive to the forces and circumstances that could or should shape the profession.;

- Education and professional development should be a cooperative enterprise involving various participants, including SAA, other national, regional, and local archival organizations, graduate-level academic programs, employers, and related professional associations.

The last point about involving "other national, regional, and local archival organizations" in this work has been a critical piece of their efforts to develop the so-called "MAS guidelines," and has been a prominent element in preceding efforts, such as the Task Force on Continuing Education. The work of this committee in conjunction with various regional and local associations will have increasing impact on the way in which we train archivists in both our formal academic programs and less formal workshops and symposia and on the skills, which our new employees will bring to the job. It will directly affect the projects we undertake, the basic sort of work we do as archivists, and the impact that our work will have on our own profession and on the world in which we live.

We are living in a very different archival world than the one in which I began my career over thirty years ago. One of the biggest differences is the enormous professional, personal, and regional diversity that I believe is one of the profession's greatest strengths. SAA continues to grow in membership, in what it provides for its members, and in its growing role as a legitimate and respected advocate for our particular corner of the information universe, of which we are now a recognized vital and legitimate part. We no longer have to bow our heads and stare at our shoes while we mumble in embarrassment that we are archivists. One of the reasons and indeed perhaps the wellspring for that professional strength is the strong network of regional, state, and local archival associations. One of the things I have tried to emphasize here today is that "national vs. regional" is not an "either-or" question. The connections existing among all us are both synergistic and symbiotic: in the first case we are indeed stronger as a whole than the individual sum of our parts; and in the second case, we thrive upon each other and indeed need each other to survive. It is as hard to imagine SAA without the SNCAs as it is to imagine SNCA without SAA.

Notes

[1] Patrick Quinn. . "Regional Archival Associations and the Society of American Archivists," *American Archivist,* 46:4 (Fall, 1983), 436.

[2] Midwest Archives Conference *Newsletter*, Vol. 1 (January 1973): 3, quoted in Quinn (op cit), 436-7.

[3] Ibid.

[4] Maher, William J. "Cooperative Competitors: Local, State, and National Archival Associations," *The Midwestern Archivist*, 16:2 (1991), pp. 105-13.

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 exemplary 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 proliferation 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 cite 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

terns 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. 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 program 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.

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,

[5] Ibid., 105.

[6] Ibid., 105.

[7] Ibid., 106.

[8] Ibid., 106.

[9] Paraphrased from Ibid., 108-109.

[10] NISTF was formed in 1977 by the Society of American Archivists with funds from the National Endowment for the Humanities. Its members consisted of Richard Lytle, chair, and David Bearman, project director, both of the Smithsonian Institution; Maynard Brichford, University of Illinois; John Daly, Illinois State Archives; Charles Dollar, National Archives and Records Administration; Larry Dowler, Yale University; Max Evans, State Historical Society of Wisconsin; Steven Hensen, Manuscript Division, Library of Congress; Tom Hickerson, Cornell University; Charles Palm, Stanford University; and Nancy Sahli, National Historical Publications and Records Commission. For a detailed summary of the work of NISTF, see Richard Lytle, "An Analysis of the Work of the National Information Systems Task Force," *American Archivist*, 47:4 (Fall, 1984), pp. 357-365.

[11] Elaine Engst, "Standard Elements for the Description of Archives and Manuscript Collections," unpublished report delivered to the National Information Systems Task Force, 1979.

[12] H. Thomas Hickerson, "Archival Information Exchange: Developing Compatibility," Paper presented at "Academic Libraries: Myths and Realities," Proceedings of the Third National Conference of the Association of College and Research Libraries, Seattle Washington, 4-7 April, 1984.

[13] Philip Hamer. *Guide to Archives and Manuscripts in the United States* (New Haven: Yale University Press, 1961).

tential 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.

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 (1996) has termed the "network society" 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 (1991) 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." (6). With the substantive view, "technology constitutes a new type of cultural system that restructures the entire social world as an object of control." (7). We should avoid this situation because 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 pat-

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 (Kodak Photo CD technology and the EPOCH optical disk "jukebox"). In the last two years, Cornell has undertaken a study with funding from Institute of Museum and Library Services (IMLS) to plan strategies for addressing issues relating to long-term data management. Working from 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 cents per image per year does not sound like much, the resources necessary to 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 po-

Electronic Text Projects Revisited

Russell S. Koonts
North Carolina State University

When "First Temple of the Atom: Electronic Text Projects on a Shoestring Budget" was published in *The North Carolina Archivist* 62 (Winter 2000), the North Carolina State University Libraries' Special Collections Department had just finished its first foray into the digital realm. With the optimism and naiveté of a child, I extolled the virtues and simplicity of creating digital projects in that article. During the intervening eighteen months, much has changed since then, and I would like to revisit that article. While I still feel that such projects are worthwhile, I now have a much greater respect for the amount of time and knowledge required for the successful completion. However, there are two basic rules to heed when venturing down the digital path:

1) KISS (Keep it Simple, Stupid). In other words, start small with a defined project on a processed collection.

2) Stealing someone else's work in the digital realm is not plagiarism. It is hacking, and hacking is the sincerest form of flattery.

Digital Projects at the NCSU Libraries' Special Collections Department

Within The NCSU Libraries' Special Collections Department, digital projects have taken two distinct paths: 1) Science and Technology Electronic Text Projects, and 2) History of Forestry Web Site. The Science and Technology Electronic Text Project was established in July 2000 in order to provide full-text, on-line access to unique resources from our holdings. This project includes selected pre-1930 engineering books and entomology articles, and selected manuscript materials (see "First Temple of the Atom: Electronic Text Projects on a Shoestring Budget", *The North Carolina Archivist* 62 (Winter 2000)) The History of Forestry Web Site project was established in December 2001 when, in partnership with the Biltmore Estate Company and the Forest History Society, we received a 2001-2002 Library Services and Technology Act (LSTA) grant from North Carolina Exploring Cultural Heritage Online (NC ECHO). The initial phase of this project will consist of EAD encoded finding aids and digitally accessible images and documents from various Carl Alwin

Schenck-related collections at the three institutions. In addition, we will digitally reproduce the 14-volume set of Romey B. Hough's *American Woods.* Each volume of *American Woods* consists of approximately 25 plates containing 3 sections (transverse, radial, and tangential) and copious notes detailing roughly 350 varieties of American trees. The following table gives a quick comparison of the two projects.

Science E-Text	History of Forestry
Pilot project to develop skills (~150 documents)	Grant-funded project promising 19, 505 images
Selected documents based on text content	Items selected based on text and visual content
Full text versions available on-line (very labor intensive)	Limited full-text versions available on-line.
Access through project description page	Access through the Schenck EAD finding aid, project description page, or Libraries' OPAC
All items digitized in 24-bit color at 400 dpi saved as TIFF	Items digitized based on original format (8-bit grayscale or 24-bit color at 400 or 600 dpi
File size of 8.5x10 page = ~45MB file	File sizes range from 8MB – 110MB
Files encoded in Teixlite	Employ both EAD and Teixlite (XML flavors)
Access provided to both XML and HTML	Access provided to both XML and HTML (finding aid also converted to PDF)
Files converted from XML to HTML with a Perl script	Files converted from XML to HTML (and PDF) with an XSLT/XSLFO style sheet
Images presented as 72dpi thumbnail with a link to a 200 dpi version	Images presented as 72 dpi thumbnail with a link to a 200 dpi version (and a link to a 600 dpi image in the case of *American Woods*)

ards 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 outlined. 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, which has enabled them to acquire high-end storage capacity with

terials.

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 "culturally 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 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 stand-

Digitization Process

Regardless of the project, the digitization process at NC State consists of seven distinct steps: 1) Selection; 2) Digitization; 3) Transcribing/Converting the Text (Optional); 4) Encoding; 5) Validating/Parsing; 6) Creating Surrogate Documents; and 7) Providing Access.

1. Selection

Digital Projects at the NCSU Libraries' Special Collections Department have allowed us to re-examine our normal course of business. Through the implementation of an EAD program, we have fundamentally changed our entire approach to processing. For example, instead of selecting a particular record group and processing the collection from beginning to end, we are in the process of creating folder level inventories for all University Archives record groups. The EAD inventory acts not only as the temporary finding aid, but also serves as our collection control mechanism and will be the basis for future full-scale processing. I mention EAD here for two reasons. First, in selecting projects for digitization projects, it is imperative that the materials are processed and that access to the collection exists. Secondly, access to our digital projects is often provided through the use of the <dao> (Digital Archival Object) tag in the finding aid. For example, our first digital project provides access to all of the documents through the project introduction page. Additionally, all of the items included in the project that were selected from the Department of Nuclear Engineering Subgroup are linked from the finding aid. The actual selection of documents is a joint effort of the specific curator and the collection processor.

2. Digitization: Scan Once Approach

For our first project, we decided to scan all items at 24-bit color and 400 dpi regardless of the original format. The result was a fairly consistent file size of about 45MB per 8.5x11 sheet of paper. For our current project, we make decisions on the resolution and color based on the original document. As a rule of thumb, color photographs are scanned at 600 dpi and 24-bit color, black and white photographs at 600 dpi and 8-bit grayscale, and manuscript documents at 400 dpi and 8-bit grayscale. File sizes range from about 8MB to 112MB. These master images are saved as TIFF files and burned onto Maxell Gold CD-Rs for off-line storage. For the next phase of our project, we will use Mitsui Gold CD-Rs, an NC ECHO recommendation.

To produce the web images, we open the files in Photoshop and

run an "action" (basically a Photoshop macro) that produces two surrogate images. First, the image size is reduced by fifty percent, the resolution is changed to 200 dpi, and the image is converted to a JPEG and saved. This large JPEG is then reduced by fifty percent, resolution is reduced to 72 dpi and saved as a thumbnail JPEG.

3. Transcribing/Converting the Text

Through the Science and Technology E-Text projects we attempt to make full-text, searchable versions of the rare materials available on-line. Due to the nature and time frame of our NC ECHO grant, this is not possible for the Forestry History web-site. During this project selected documents will be transcribed to varying degrees. Although the projects differ in their full-text approach, the method for achieving the full-text digital version does not. In most cases (where there is clear, typed text) the TIFF image is opened in Text-Bridge Pro and converted to text through Optimal Character Recognition (OCR). We tried OmniPage Pro, but our limited experiments exhibited a higher accuracy rate in Text Bridge. Once the text has been processed, it still needs to be proof-read, as even the best OCR software has only a 98% accuracy rate. In cases where the text is handwritten, or the typeface is of poor quality, we simply re-key the entire document.

4. Encoding

Up to this point, there have been limited amounts of investment in terms of training. Once you reach this point, however, the learning curve has the potential to become steep. Depending on the nature and long term use of your project, you will need to decide on an encoding language. If you have no plans for future development, then simple HTML encoding is sufficient. If, on the other hand, you plan to continuously develop your project and make use of emerging technology, you should probably use an XML-based language. For all of the text-based documents in both projects, we are using the XML version of the TEILite subset of the Text Encoding Initiative (TEI) encoding language. I learned the TEILite encoding language by attending David Seaman's Rare Book School session on creating digital texts (currently around $700 per week excluding travel and food); by visiting Natasha Smith at Documenting the American South at The University of North Carolina at Chapel Hill; and through trial and error.

Once you have a grasp of TEILite (or any other encoding language: we use EAD, TEILite, and MathML in our projects), you will need the software to encode the document. Here, again, you have some decisions to make. There are a number of authoring software packages available today. The two I have used, and will mention,

sources 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 and may, in fact, be their most important innovation.

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 to glean some of the larger conclusions to be drawn from Cornell's experience.

Issues in Electronic Preservation and Access

Cornell has achieved a position of leadership in no small part because of the volume of 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 ma-

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 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 re-

are XMetaL and NoteTab. XMetaL is created by SoftQuad, comes ready to use right out of the box, has numerous user interfaces, comes preprogrammed with a validator/parser, and costs around $500 ($250 for educational users). NoteTab is created and maintained by Fowkes Software (Erik Fowkes), requires a great deal of programming, requires helper applications acquired elsewhere, and has no native viewing interface. However, it is either free for the Light version or $20 for the Pro version.

We have made the choice to use NoteTab in our projects. I like to refer to NoteTab as Notepad on steroids. Like Notepad, it is a simple text editor. Unlike Notepad, you have the ability to create templates, and clips (similar to Word macros) that make repetitive tasks much easier. With these templates and clips in place, our students answer a series of simple, straightforward questions and produce a valid TEILite document without ever having to learn the language. Such advantages were the results of an enormous investment and commitment on our part. In addition to the work we have done to tailor NoteTab to our uses, there is a growing community of NoteTab users who create and share libraries and there is also a NoteTab listserv where the answer to many questions can be found.

XMetaL	**NoteTab**
Ready to use right out of the box	Requires a great deal of programming
Has numerous user interfaces	Requires helper applications acquired elsewhere
Comes preprogrammed with a validator/parser	Has no native viewing interface
Costs around $500 ($250 educational use)	Is either free for the Light version or $20 for the Pro version

5. Validating/Parsing

When you decided to create an XML document, you either knowingly or unknowingly agreed to abide by the XML rules set forth by the World Wide Web Consortium (W3C for short). The W3C states in its XML recommendation (February 1998, *http://www.w3.org/TR/2000/REC-xml-20001006*), that all XML documents must be valid. As we have chosen TEILite, our documents must be valid according to the rules set forth in the TEI-Lite document type definition (DTD). This DTD states which tags can be used where, in what order, and within which other tags. There are numerous tools that will parse your document to ensure its validity.

If you use XMetaL, the parser comes pre-installed. If you use NoteTab, you will need to choose and install the parser of your choice. We use XMLINT (Microsoft) and a native parser in the Internet Explorer 6.0 browser. XMLINT is free and can be downloaded at *http://tabor.millikin.edu/sts/xml/xml/tools/xmlint.zip.*

Another functional parser is Xerces-J (*http://xml.apache.org/xerces-j/*). The parser reads the DTD into memory and "walks through" your document and checks the document for encoding errors. If it locates an error, it creates an error message. Once you fix that error, you continue to parse the document until you receive the message that the document is valid.

6. Surrogate Documents

Since TEILite/XML is not currently pre-coded into any browsers, you need to create a style sheet that tells the browsers how to handle the TEILite/XML. Once again, you have choices to make. There are two style sheet W3C recommendations, Cascading Style Sheets (CSS) and eXtensible Stylesheet Language (XSL). If you choose CSS, you can display your documents in Internet Explorer (IE) 5.0 and higher, Opera, or Netscape 6.0. However, XSL was created specifically for transforming XML-based documents. Presently only Internet Explorer 5.0 and higher support XML/XSL documents.

The W3C working draft is supported in IE 5.0 and higher, and the W3C recommendation is supported in IE 6.0. To complicate matters, between the working draft release and the version 1.0 recommendation, the W3C changed XSL drastically. The XSL descendants from which you can choose are the eXtensible Stylesheet Language Transformation (XSLT), XML Path Language (XPath), and eXtensible Stylesheet Language Formatting Object (XSLFO). The latter is used in converting XML docu-

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 (IMG) 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, resulting 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) The consistent pattern of good human resources decisions is an example of another exemplary practice in the CUL. It would be difficult to attribute this practice 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.

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 Photo CD 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 Photo CDs, 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 cooperation 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,

> *was 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*

ments to PDF format.

In our first digital project, we wrote a Perl script that converted our TEILite to HTML. In all of our current digital projects, we use a combination of CSS, XSLT, XPath, and XSLFO.

Now that we have the TEILite/XML document and the XSL style sheet, how do we deliver it to the public? If we just offer the XML, we exclude users who do not have Internet Explorer 6.0. If we deliver HTML, do we create static HTML pages or convert "on-the-fly"? If we create static HTML pages, which application do we use? If we create HTML "on-the-fly" do we do it on our server or in the user's browser?

We provide access to up to three documents from our web page: XML, static HTML, and occasionally, PDF. To create the static HTML, we use freely available helper applications. I have experimented with MSXSL (Microsoft *http://msdn.microsoft.com/downloads/default.asp?url=/downloads/sample.asp?url=/msdn-files/027/001/485/msdncompositedoc.xml*); Saxon (Michael Kay *http://users.iclway.co.uk/mhkay/saxon/*); XT (James Clark *http://www.jclark.com/xml/xt.html*); and Xalan-J (Apache *http://xml.apache.org/xalan-j/index.html*).

I prefer either XT or Saxon, with XT being the most forgiving in terms of reporting errors. To create our PDF files, we use FOP (Apache *http://xml.apache.org/fop/*). Again, all of these require downloading, installing, and then programming NoteTab with the correct commands.

7. Providing Access

As mentioned above, one method of access in the NCSU Special Collections is through the EAD document. Another method is through the NCSU Libraries' OPAC. Yet another is through commercial search engines (Yahoo, Google, etc.). Finally, the NCSU Libraries acquired Blue Angle MetaStar, a search engine that uses the TEILite/XML encoding that we employed.

For example, if a document contains the following three tags:

<name type="person">Sir Walter Raleigh</name>

<name type="corporate">Raleigh Ale House</name>

<name type="geographic">Raleigh, North Carolina</name>

With an XML-enabled search engine, one can search for all instances of the City of Raleigh without producing false hits of personal or corporate names.

Do I still think that digital projects are worthwhile? Yes. Do I

still think they are necessary? Yes. Do I still think they are simple? Yes, but not quite as simple as when I wrote the first article. But then again, I am a bit spoiled. I work in a supportive institution, with great individuals, and a great deal of collaboration. The amount of idea sharing and project development that has taken place between departments and over cups of coffee clouds my judgment on the ease of digital projects. I do feel, however, that every institution, no matter what size, has the ability to create a successful digital project. If you remember to keep it simple — start small with a defined project on a processed collection — and to hack, your project can be a success.

Links and code available at:

Science and Technology Electronic Text Projects:

http://www.lib.ncsu.edu/archives/etext/

History of Forestry Web-Site (Coming July 2002):

http://www.lib.ncsu.edu/archives/forestry

Links to software and helper application:

http://www4.ncsu.edu/~rkoonts

Links to NoteTab EAD clip libraries and templates:

http://www4.ncsu.edu/~rkoonts/notetab

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

The development of *Utopia* was conducted with extensive front-end input from potential users. 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 orientation of the project avoided some of the difficulties with understanding substantive patterns of use. It is a luxury not available to projects with a more "extramural" focus.

gy 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 goals 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.

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

Reaching Out to Researchers: A Model for Web-based User Education for Archives and Manuscript Collections

Jill Katte
Duke University

Abstract

This paper discusses the need for and presents a model of web-accessible instruction and orientation resources to support users of archives and manuscript collections. Traditional archives orientation and user education practices, as well as web-based library instruction techniques, are reviewed to create a framework of types of information that should be provided and types of resources that best provide this information. A sample of thirty special collections department web site at large research institutions are examined in depth using this framework to discover what user education resources are being implemented in the field. Applying the evaluation framework and findings, a model for web-based archives user education resources is offered, featuring content on archival orientation, intellectual access, physical access, and utilization.

Introduction

The widespread adoption and use of the Internet has given archives and manuscript collections not only opportunities to offer research tools to remote users, but also challenges to determine the best ways to support these services. Increasing numbers of repositories are encoding their finding aids and mounting these digital resources on the World Wide Web. According to recent research, making finding aids accessible through the web and implementing Encoded Archival Description expand the roles these finding aids perform (Tibbo and Meho, 2001; Gilliland-Swetland, 2001). The digital finding aid is not only a guide to a collection, but also a searching and retrieval tool. Because the finding aids are more widely accessible on the web, more people will be likely to discover them while performing standard web searches. In addition to students and scholars, members of the general public may retrieve archival finding aids while conducting personal research, but many individuals may not be familiar with these tools or know how to use them.

Before institutions began digitizing their finding aids, researchers discovered collections by consulting published research guides and union catalogs, by following citations to collections in other published works, and by word-of-mouth from col-

leagues. After determining where the relevant collection was held, researchers either traveled in person to the repository to use the collection finding aids or requested to have a copy of the finding aid mailed or faxed to them. In either of these situations, reference archivists had the opportunity to communicate with the user before, during, and after the user worked with the finding aid. The archivist could educate the user on what a finding aid was, how to use the finding aid to navigate the collection, and how to find other archival finding aids related to his or her area of study.

Web delivery of finding aids and their use as retrieval tools, however, does not automatically offer this archival reference component. An individual whose information need may best be met with archival materials may retrieve a finding aid from a web search without understanding what it is or how to use it. This unmediated interaction with the repository through an unfamiliar descriptive tool might well discourage many prospective archives patrons. A straightforward solution to this problem is to provide web-accessible user instruction information in conjunction with web-accessible archival finding aids.

This paper will examine the ways in which archival repositories currently are presenting user education resources on their web sites, including which elements of traditional archives user instruction are being transferred into web-based resources, and which, if any, additional resources are being offered. First, a brief analysis of conventional archives reference services will be provided to determine what types of information could be included in a digital user education resource. Then, web-based user education resources that have been implemented in libraries will be considered to discover if any library user instruction techniques could be applied to archival tools.

Following the literature review, the research methodology of the study will be described in detail, including the sampling frame, the sample size, definitions of concepts used in the analysis, and the system of evaluation of the web resources. Then, the findings of the web site survey will be presented, summarizing the current practices of the archival institutions examined in the study, as well as highlighting some of the methods that are not widely used in the field.

Using the information collected in the study and the sources consulted in the literature review, a model of practical recommendations for useful web-based archival user orientation resources will be proposed. Adding electronic user education resources to complement web-based finding aids is a necessary

film 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.

A further initiative undertaken at the Mann Library 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 has partnered 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 projects exemplify some of the approaches adopted in the Mann Library as part of the larger history of the growth of Cornell's digital strategy. 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.

DAC, CIDC, and Institutional Infrastructure

Cornell's commitment to integrating communications technolo-

tion 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 concert 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 National Endowment for the Humanities (NEH), the Rockefeller Foundation, the U.S. Department of Agriculture, 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 micro-

step in ensuring the widespread use of archival collections in the digital age.

Literature Review

Traditional Archives User Education

Most archives and manuscript collections offer traditional user orientation and reference services. During typical reference interviews, the archivist usually explains the general research guidelines for the institution or collection and helps the user understand how to find information using finding aids or catalogues specific to the institution. The interview is also a way for the archivist to ascertain the researcher's information needs, which helps the archivist recommend particular collections within the repository that may contain useful materials for the researcher. This section of the paper will summarize the most important components of the orientation and reference interview and will examine whether these components can be transferred into a web-based medium.

Until recently, not much emphasis was placed upon reference services in archival settings. In fact, "the term reference was not found in the titles of any articles in *American Archivist* between 1938 and 1950" (Whalen, 1985). Because archivists were trying to manage increasing amounts of paper and organizational records, reference services may not have been the highest priority. Archivists were more concerned with description, preservation, organization, and classification of records.

Although it was not emphasized initially, archives reference service is extremely important to the successful use of archival materials. Access to these materials is frequently neither easy nor self-evident. Archives users are much more dependent upon archives staff than users of libraries (Tissing, 1984). It is usually more difficult for an archives user to orient him- or herself to the repository. This is because there exists no universal classification system for archival collections and because most repositories maintain closed stacks for material storage and security. As a result, user self-service in archives is virtually non-existent. Archives users must interact with public services staff in order to find and retrieve the records they need. Therefore, "enabling the researcher to use archival records effectively and efficiently is the central tenet of the modern archivist's mission" (Cox, 1992).

High-quality reference service is made up of three essential elements: the researcher, the reference archivist, and the records (Chalou, 1984). In an ideal situation, these three elements

come together in a sort of "convergence," meaning that the researcher and reference archivist work together to determine which records will be of interest to the researcher. The pertinent records are retrieved for the user, and the researcher's information needs are met. In order for such a convergence to happen, each of the three elements must be optimized. The records must be preserved properly, arranged, and housed, and finding aids and inventories must be created so that the user can find records quickly and effectively. Reference staff must be receptive to user needs and able to locate records to meet the user's research goals. Researchers must understand how to use the archives and its finding aids to support their research.

Archives administrators do have control of the organization and level of access they provide to their records and the level of service reference archivists offer to users. They do not, however, have control of how much knowledge or experience in the use of archival resources their researchers have before they arrive. Therefore, an integral part of providing high-quality reference service is the orientation interview. In fact, the main problems researchers have in finding and utilizing archival records is due to a lack of orientation to archival research tools and methods (Whalen, 1985). In his research on user studies at the National Archives and Records Administration, Conway (1994) identifies user education and training as an important aspect of archives reference service that requires further research:

> The user study demonstrated that a significant portion of researchers, on a continuing basis, lack experience with primary research methodologies in general and with specific agency procedures. To assist users in making efficient use of the holdings without significant direct assistance, practical research should be undertaken to determine the most effective mechanisms and procedures for compensating for inexperience.

The archives orientation interview is made up of two critical components. First, the archivist should explain the basic information a researcher would need to know when using records in that particular repository. This information would include reading room procedures, reproduction policies, records organization, and a brief orientation to understanding and using finding aids. Second, the archivist should help the researcher define his or her research objectives, determine the scope and depth of the research, and recommend archival resources for the researcher to utilize.

> 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 1988, when discussions for what would become the Chemistry Online Retrieval Experiment (CORE) began.[7] The project, which actually got underway in 1990, was a joint venture of the Mann Library, the Chemical Abstracts Service of the American Chemical Society, Bellcore, and Online Computer Library Center, Inc. (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 not only gave the Mann Library staff explicit information about how their users interacted with their digital products and allowed for improved user orientation, but it also gave them an actual example of how subsequent projects might work. This was certainly a factor in the willingness of the staff to adopt this mode of access provision.

Similar to the Olin Library, the Mann Library experienced a serious problem with the degradation of books from the mid-nineteenth 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 that some 350,000 volumes in the Mann Library Collec-

(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. 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 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*

According to an archives reference study conducted by Robert Tissing (1984), most reference archivists surveyed did not use a checklist or guide for conducting the reference interview. Using the information collected in a survey, Tissing compiled a checklist of topics to cover during the archival orientation interview. The topics included in this checklist will serve as a framework for the archival web site content analysis provided in this paper.

The other goal of the reference interview is to understand the user's needs and identify records that will meet those needs. Research in archives follows a needs-oriented rather than question-oriented model (Cox, 1992). The face-to-face interview allows the reference archivist to ask questions that will help the user communicate the research needs, the scope and depth of inquiry, and the length of time he or she is able to devote to the research (Chalou, 1984). Using this information, the archivist may suggest collections containing material pertinent to the researcher's goals.

In the traditional model, resource identification relied heavily on the memory of the archivist (Whalen, 1985). The belief was that the "most valuable finding aid is the reference person, not a prepared guide, index or inventory" (Chalou, 1984). The reference archivist had to be "omniscient," knowing everything about every collection within the repository, and, therefore, able to recommend all of the relevant collections to any particular researcher (Pugh, 1984).

Realistically, the archivist will not know everything about every collection, especially if he or she personally did not process the collections. Many archival finding aids describe materials according to their creation and reflect the organization of the institution that created them (Whalen, 1985). Subject retrieval can be difficult when records are organized in accordance with the principles of provenance and original order. Archival materials are unique and activity-centered, and they are not classified like library materials (Pugh, 1984).

In order to remove some of the reliance on the archivist's memory, subject indexing of collections can improve the resource recommendations archivists give to their patrons. Assigning multiple subject access points to each collection can make archival resource recommendation more helpful and more meaningful. Good subject indexing allows reference archivists to rely less on their ability to memorize information about collections and more on their searching and retrieval skills when identifying relevant resources (Whalen, 1985).

The traditional orientation and reference interview provides information to both the researcher and the archivist. The researcher learns about the policies and procedures for the successful use of the resources in the archival repository. The archivist understands the researcher's needs and suggests resources for him or her to incorporate into research. In a web-based environment, similar types of information must be exchanged, but largely without any face-to-face interaction. In order to determine how to translate the user orientation and education information communicated in the reference interview into an Internet-compatible format, user orientation resources employed by libraries will be considered. Some of the strategies for user education developed for library patrons may also be suitable in an archives environment.

Web-based Library Instruction

In recent years, libraries have employed web-based user education tools to complement or replace traditional classroom-based or face-to-face instruction. Although libraries provide different types of resources and services to their users compared to archives, some of the electronic user education techniques used in libraries could be modified to work in an archival setting. Various strategies developed by libraries for electronic user education will be examined here, and each will be evaluated for adaptability into an archival user education web presence.

As electronic, web-accessible library resources become available, the need for web-accessible library education increases (Vander Meer, 2000). Distance education programs rely upon web-accessible library resources to serve remote learners. Internet-based library instruction assists these off-site researchers in successfully utilizing library resources. Web-based library instruction also gives library educators added flexibility. Librarians can use the web to instruct users in an asynchronous environment, "without the constraints of the fifty minute, one-shot class" (Dewald, 1999). Users can choose what library skills they are interested in obtaining and can progress at their own pace. The interactive component of selecting and applying library information resources on the web actively engages learners, which promotes long-lasting skill retention.

Libraries employ several types of web-based tools for instruction: Frequently Asked Questions (FAQ) and navigational aids, customizable interfaces, tutorials, email and chat reference, and research guides and handouts (Vander Meer, 2000). Navigational aids are guides to finding specific resources both on the library's web site and within the physical library facility.

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 Foundation). 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, projects that illustrate a process of building on the knowledge acquired in other parts of the institution.

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

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 was 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, and 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 than 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

For example, the University of Nebraska libraries feature a "How do I…" link that leads users to an easy-to-understand list of FAQs, such as "How do I find a journal article?" Each question leads the user to help screens and on-line resources related to the question. Similarly, the University of Arizona provides a "How to Find" menu that directs the user to information on using the library catalog, finding journals, and searching databases. Though these navigational aids could be extremely helpful to new library users, they are often difficult to find on the library home page because they are buried under several layers of departmental web pages. If prominently featured on the home page or linked strategically to particular resources, navigational aids can be extremely useful in an archival setting. A possibility could be to include a "How to Use the Archives" guide including links to information on different aspects of archives use.

Library web page customization allows users to modify the library web interface according to particular profiles. Librarians can use this customization feature to target instruction to users who are in particular classes or who have set up particular research interests in their profile. Examples of library web page customization are "My Gateway" at the University of Washington and "My Library" at North Carolina State University. Researchers could save finding aids or research guides created by the archival repository on their personalized library desktop for ease of use. Although extremely promising, customization may be difficult for some smaller institutions to implement because these projects require extreme amounts of computer programming and support. Interface customization may also require a high level of computer literacy on the part of the researcher, which might discourage beginning computer users. Also, many archives users who are not students attending or researchers employed by the institution may not want or be able to set up a profile. A workable option would be for a manuscript repository to offer its users digitized archives resources that could be accessed through a customizable user profile, but that are not dependent upon it.

Library instruction tutorials provide hands-on research exercises that teach users library skills by example. Many tutorials are built to accompany student research projects, including the course-integrated instruction modules created at the University of Illinois at Chicago (Koening, 2001). These course-specific tutorials are particularly successful because they pair library instruction with a specific context of research needs, the course project (Donaldson, 1999). A drawback to the tutorials is that they require a considerable amount of time to complete. Many

students said they probably would not use a library tutorial unless it was specifically assigned in class (Vander Meer, 2000). Tutorials may not be the best method for delivering archives user orientation information electronically because of the time commitment involved in taking the tutorial from start to finish. Some potential archives users may be discouraged because the tutorial does not provide quick access to the information they need. The tutorial might, however, be a viable option for archival repositories that directly support assigned coursework.

E-mail and chat can provide library users with information tailored to their needs. The library user could email a particular research concern or difficulty to the library reference or instruction contact and could receive a reply within minutes. Chat allows a patron to communicate with a librarian while searching for information, which can help the user identify ways to improve research strategies. Although email and chat features provide ways for library users to get very specific help, they are very labor intensive on the part of the librarian. Other user instruction resources should be provided so that the email and chat provisions become a last resort if these resources do not address the specific question. Archives can implement email and chat capabilities, but user education information can be standardized and disseminated in other ways more efficiently.

Handouts and research guides provide step-by-step instructions on how to use library resources. Many times these are pathfinders and other library orientation documents that have been edited and digitized for web-accessibility. These guides can be focused around particular topics or particular library resources, providing context-specific information in an easy-to-read document. Users, however, need to be able to find the guides, so strategic placement of the information at the point of use is extremely important. For example, information on how to search databases should be found at the database gateway. Archives also could implement web guides to research, links to finding aids, online exhibits, and the repository home page.

Each of the numerous types of web-based library instruction has its own primary applications. Some of the strategies would work well for archival collections, while others might not. The techniques that could work most effectively for archives also seem to be the most straightforward. Navigational aids like "How to" pages can lead the archives user to the instruction he or she needs. Research guides can provide basic information in a standardized, widely available form, and can give step-by-step instruction for conducting archival research. These resources could also direct the user to other collections within the reposi-

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 technologically 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 nineteenth 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 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 had 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 work-flow solution. It was decided that the volume of images to be produced would be too large to be handled adequately in-

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. Trialability is particularly significant 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: "relatively 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. Observability is 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.

Making of America 1

The *Making of America* (http://library5.library.cornell.edu/moa/)

tory or in other repositories that relate to a particular research area. Email and chat could serve as a backup to field any questions not addressed by other resources.

Methodology

I selected the institutions included in this study from the Carnegie Classification of Institutions of Higher Education, 2000 edition, "Doctoral/Research Universities – Extensive." All of the 151 institutions in this Carnegie Classification served as the sampling frame. The Carnegie web site listed the public institutions first, alphabetically by state, then the private, Not-for-Profit institutions alphabetically by state. I entered these, in the order presented by the Carnegie Foundation, into a spreadsheet, and then enumerated them. Using a random number table (Babbie, 2001), I selected a random sample of thirty institutions for the web site evaluation.

For each of the institutions in the sample, I found the institutional web site, then navigated to the library web site, then attempted to find the web page for the manuscript department, archives department, or special collections department. Because the organization of each library is slightly different, it was difficult at times to find the bulk of the manuscript collections. Some institutions have a manuscripts department, others group manuscript collections with rare books in special collections departments, and others pair manuscript collections and personal papers with the archival records of the institution. As a way to handle this disparity, I tried to identify the department where the greatest concentration of manuscript collections and personal papers were located at each institution and evaluated this department's web resources.

Once I identified the appropriate archives web site at each institution, I examined the types of user education resources offered by the repository. The literature outlined in the previous section defines a number of different types of user education resources (Vander Meer, 2000). These resources can be categorized into several groups:

- **Frequently Asked Questions** are usually a page of links designed to help users locate specific resources both on the archive's web site and within the physical archives facility.

- **Interface Customization** allows users to modify the library or archives web interface according to particular profiles. Librarians and archivists can use this customization feature to target instruction to users who

are in particular classes or who have set up particular research interests in their profiles.

- **Tutorials** provide hands-on research exercises that teach users skills by example. Many tutorials are course-integrated, or built to accompany student research projects.

- **Research Guides** provide step-by-step instructions on how to use specific resources or how to conduct research in particular subject areas.

I did not examine e-mail and chat reference in this study. Although email and chat provide a means for archives users to get individualized assistance, it is time-consuming on the part of the reference archivist. Other user instruction resources can be provided so that email and chat provisions become a back-up if the resources provided do not address a specific question. Archives can offer email and chat provisions, but standardized user education information can be offered quickly and easily in other ways.

After I determined what types of resources an institution provided, I conducted latent content analysis to interpret and record the types of information being communicated by these resources. The typology I utilized is based upon the "Basic Guide for the Orientation Interview" developed by Robert W. Tissing (1984). The types of information for which I searched are:

- **Definition of primary resources**: Some remote users may have little or no experience with archives or manuscript collections. This category includes information about what an archive is, what types of documents are housed in an archive, and why these are significant to research.

- **How to interpret finding aids**: Researchers unfamiliar with archival repositories may not know how to use a finding aid to navigate a research collection, and many finding aids are not equipped with explanatory texts. Resources in this category would describe the parts of the finding aid to help the researcher interpret it more effectively.

- **Reader requirements**: For security reasons, many archival repositories require users of their collections to submit an application, show one or more forms of photo identification, obtain a researcher card, or acquire a recommendation to use collections. Resources in this category should list these requirements for research.

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 proactive 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

1,000 volumes selected for scholarly value on the basis of recommendations by a faculty panel and by citation analysis. Selection criteria included format (no larger than 8.5x11), content (no large or color illustrations), and circulation (heavy). According to 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 application of new technologies to old problems is a matter of 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 projects undertaken at Cornell, it will be useful to discuss the process of innovation adoption.

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.

1. Relative advantage. Rogers defines relative advantage as "the

- **Photoduplication policies**: The archives web resources in this category should mention whether photocopying is allowed or if there are photocopiers available for use in the reading room. In addition, information about photographic services, like scanning, optical character recognition, or film photography will be included in this category.

- **Copyright information**: Researchers need to know whether the resources they are using are copyrighted or in the public domain. Archival user education resources in this category explain copyright restrictions and how to seek permission to use copyrighted materials.

- **Publication guidelines**: Some archival repositories require the researcher to submit a copy of any publication referencing the materials in the collection. If the repository communicates such a requirement, it would be included in this category of web-based user education resources. Also guidance on how to cite materials in the collection will be categorized here.

- **Reading room procedures**: The method for requesting and viewing documents varies from repository to repository. Information that fits into this category includes whether documents need to be requested by the researcher in advance, what kinds of materials can be taken into the reading room with the researcher, and whether there are outlets and data ports for laptop computers.

- **Resource identification**: This information instructs the user on how to find resources pertinent to their research objectives, by describing subject indexing and how to use on-line catalogs to locate archival collections. This may also include lists of resources related to a particular subject area, which can help researchers identify relevant collections.

As I analyzed these web sites, I used a data collection worksheet to quickly record and code the information. For each resource type and information type, I simply checked to see if the information was present within the institution's web page. I did not evaluate whether or not I felt the information provided by the repository was helpful, just whether or not any information in each category or of each type was presented.

Findings and Implications

User Education Resource Types

First, I examined the ways in which user education information is offered by each institution via their Web site. The resource types that I recorded were Frequently Asked Questions pages, interface customization, tutorials, and research guides. Of all of the resource types, the most widely implemented is the research guide, with twenty-six of thirty, or 87%, of the sample institutions providing these guides. The prevalence of this type of user education resource might be related to the goals of the manuscript library. If a repository's objective in creating a web site is to increase access to finding aids and other collection metadata, the library will focus a significant portion of its web site to helping users find relevant collections. Some of these guides help the user understand the relationship between the manuscript collection, the finding aid, and the on-line public access catalog at the institution. A majority of these resources assist users in accessing materials by grouping collections based on particular subject areas, which can help researchers identify similar or related collections.

The other resource types I considered in the study were implemented much less frequently. Only five institutions included a Frequently Asked Questions page in the archive or manuscript web site. A good example of FAQ in a special collections context can be found at Louisville University (http://library.louisville.edu/uarc/faq.html). This resource anticipates the questions a researcher may have when visiting the web site and attempts to provide appropriate answers to these questions.

One institution in the sample, North Carolina State University (http://www.lib.ncsu.edu), offered library web interface customization. The Special Collections and University Archives departments could be accessed through this portal, but it was unclear whether any targeted user instruction resources were created to take advantage of the customized user-profiles created by researchers. One feature that could be leveraged by the Special Collections department is the "New Titles" element. As the archival repository acquires new materials, collections, and records on a particular subject, a notification could be linked to user profiles created with this subject interest. This could be an implicit way to educate researchers about the subject areas being developed within the collection, as well as what types and formats of materials are available in the manuscript and archives departments.

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 and the incremental nature of this process.

The CLASS Project and the Diffusion of Innovations

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 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

nology 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 by 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.

No institutions in the sample provided a tutorial to instruct remote users. This may be because tutorials are complex and time-consuming to plan, create, and maintain. The instructional value of tutorials, however, may make the time and effort required to create them worthwhile. One example of an archives tutorial can be found at an institution not included in the sample, Auburn University Special Collections and Archives. The tutorial is entitled "What is an Archive and How Do I Use It?" (http://www.lib.auburn.edu/archive/user/). Although some sections are still under construction, the main elements of the tutorial provide an overview of the processes, terminology, resources, and research methods associated with using archival collections. The tutorial also features a Frequently Asked Questions section and the taxonomy of Auburn University finding aids.

User Education Information Types

The framework of information types used in the analysis includes the definition of primary resources, how to use and interpret a finding aid, reader requirements, photoduplication, copyright, and publication policies, reading room procedures, and resource identification information. Repositories included in the sample each featured at least one of the information types, but none of them included all of these informational elements.

The information type that is most widely implemented is resource identification information. Twenty-five out of the thirty institutions in the sample provided resource identification information for remote users. Resource identification information is most often provided by the research guide format. The prevalence of this information across many institutional web sites also probably is related to the goals of the repository in increasing access to the collection. In addition to instructing researchers in the use of tools like the on-line catalog, some archival repositories synthesize their own tools for researchers, such as subject-focused research guides to the collections held within the repository. Examples of both kinds of resource identification information can be found at the Southern Historical Collection at the University of North Carolina at Chapel Hill. The web site provides information on finding resources in the on-line catalog (http://www.lib.unc.edu/mss/msshelp.html) and offers subject research guides to the collections (http://www.lib.unc.edu/mss/finding.html).

The next information type in terms of frequency of implementation is photoduplication information or policies. Of the thirty

institutions examined, twenty-one provided some information about whether materials could be photocopied, photographed, microfilmed, or duplicated digitally. While many institutions in the sample included photoduplication information on their web site, the level of information provided varied widely. Some repositories included a very brief sentence or two indicating that photocopying provisions were available, or that photoduplication of archival materials in the repository was at the discretion of the archivist. Others included several paragraphs about the various photoduplication methods available to researchers, and some even included fee schedules for these services. An example of detailed photoduplication information can be found at the Massachusetts Institute of Technology Institute Archives and Special Collections, (http://libraries.mit.edu/archives/research/reproduction.html).

Reader requirements, such as presenting identification, confirming institutional affiliation, or completing a registration form, are sometimes included on the repository web sites. Of the thirty institutions examined, nineteen, or 64%, included some information of this type on their web sites. Columbia University's Rare Book and Manuscript Library offers a clear example of this kind of information under the heading Use of Collections (http://www.columbia.edu/cu/lweb/indiv/rare/about.html#Use):

> *The collections of the Rare Book and Manuscript Library are available for use of Columbia University faculty and students, for those in affiliated institutions, and for readers not affiliated with Columbia who are engaged in scholarly or publication projects. Readers are asked to register at the Rare Book reference center and to provide a current photographic identification card. We recommend that new readers make appointments before coming to the library and to familiarize themselves with the hours that we are opened, the times that we retrieve materials for readers, and the location that the materials are stored. (Some materials cannot be retrieved without 24 hours advance notice.) Non-Columbia undergraduates are required to apply in advance and to provide letters from their faculty advisers explaining their need to use the manuscript and rare book collections.*

While resource identification information helps researchers access collections intellectually, reading room procedures help researchers access collections physically. Fourteen of the thirty repositories in the sample, or almost half, include information of this type on their archives web sites. The University

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. 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. This section will 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 tech-

Institutionalizing Success: The Growth of a Digital Strategy in the Cornell University Library System

John Foster
University of North Carolina at Chapel Hill

Introduction

Viewed purely in terms of volume, the library system at Cornell University has established itself as 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 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. In addition, CUL has participated in organizations relating digital preservation and access, principally the Council on Library and Information Resources and the Research Libraries Group, and has expanded 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

of Chicago Special Collections Research Center (http://www.lib.uchicago.edu/e/spcl/using/access/guidelines.html) provides a good example of reading room procedures. These guidelines explain how materials are requested, what can or cannot be taken into the collection's reading room, whether provisions are available for the use of laptops, and methods for the proper handling of the library's materials.

Only ten institutions in the sample, or roughly 30%, feature explanatory information about the primary resources in their repositories, what a manuscript collection or institutional archive contains, why the materials are there, and why the materials could be significant to researchers. Some institutions provide this information in narrative form, such as the University of Idaho Special Collections and Archives (http://www.lib.uidaho.edu/special-collections/Primary.Sources.html). Their guide integrates an overview of primary resources within the repository with resource identification information. The L. Tom Perry Special Collections Library at Brigham Young University (http://www.lib.byu.edu/~scm/sampler/index.html) also conveys information about the kinds of resources in its collections through the use of a "Sampler," or a group of digital images which provide a graphical representation of the variety of materials in the collection, such as manuscript pages, letters, early printed books, photographs, and ephemera. This method informs the user visually and educates them by example.

A vital information type for users of collections is the copyright policy, which informs the user on the legal rights status of the materials, and it also can help to protect the institution from copyright litigation and protects the rights of the copyright holders. Of the thirty institutions examined in the study only nine include a general copyright policy. Several of these institutions use the required copyright statement provided by Section 108 of the U. S. Copyright Act to address not only copyright but also photoduplication concerns on their web sites:

> *The Copyright Law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials.*
>
> *Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specific conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship or research." If a user makes a request for, or later uses, a photocopy or*

reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Other institutions also include more specific copyright and usage information, tailored to the contents of the repository. An example of this can be found at the Manuscripts Department at the University of North Carolina at Chapel Hill (http://www.lib.unc.edu/mss/copy.htm).

Publication guidelines inform the researcher about how to cite materials found within the collections and if there are any policies or fees associated with scholarly or commercial publication of the materials within the collection. Five institutions in the sample provide publication guidelines on their web sites. The Special Collections Department of the Honnold/Mudd Library at Claremont Graduate University (http://voxlibris.claremont.edu/sc/access/services.html#citing) gives an example of publication guidelines on their web site. Also Auburn University Special Collections and Archives uses its tutorial to present guidelines for citing many different formats within the collection, including letters, diaries, photographs, maps and plans, and online resources (http://www.lib.auburn.edu/archive/user/rescitation.html).

The least implemented user education information type in the study is finding aid interpretation information. Only one institution of the thirty in the sample offers any kind of help for users who may not understand what a finding aid is or how to use it to navigate a collection. An example of finding aid interpretation information can be found at the Auburn University Special Collections and Archives web site (http://www.lib.auburn.edu/archive/user/rescitation.html) within the tutorial. They feature a guide to understanding the different parts of an archival finding aid, including instruction on how to use a container list to find materials within a collection

While each institution offered at least one type of user education resource and at least one type of user education information, none of the archives or manuscript repositories offered all of them. The next section of the paper examines some possible reasons for the findings of the analysis. After the results of the study are considered, a model for workable, inclusive archives user education resources for the web will be proposed.

Web sites Cited

AU Special Collections & Archives, Archival Tutorial. 31 July 2000. Auburn University. March 2002 <http://www.lib.auburn.edu/archive/user/>.

Carnegie Classification of Institutions of Higher Education. 2000 edition. The Carnegie Foundation for the Advancement of Learning. March 2002 <http://www.carnegiefoundation.org/Classification/index.htm>.

Manuscripts Department Copyright Information, UNC-Chapel Hill. 17 April 2000. University of North Carolina at Chapel Hill. March 2002 <http://www.lib.unc.edu/mss/copy.htm>.

Manuscripts Department Helpful Hints, UNC-Chapel Hill. 17 April 2000. University of North Carolina at Chapel Hill. March 2002 <http://www.lib.unc.edu/mss/msshelp.html>.

Manuscripts Department Online Finding Aids and Guides, UNC-Chapel Hill. 15 June 2001. University of North Carolina at Chapel Hill. March 2002 <http://www.lib.unc.edu/mss/finding.html>.

NCSU Libraries. North Carolina State University. March 2002 <http://www.lib.ncsu.edu>.

Primary Sources: Personal Papers and University Archives. University of Idaho. March 2002 <http://www.lib.uidaho.edu/special-collections/Primary.Sources.html>.

Rare Book and Manuscript Library, Background. 12 March 2002. Columbia University. March 2002 <http://www.columbia.edu/cu/lweb/indiv/rare/about.html#Use>.

Reproduction Services: Institute Archives and Special Collections: MIT. Massachusetts Institute of Technology. March 2002 <http://libraries.mit.edu/archives/research/reproduction.html>.

Sampler. Brigham Young University. March 2002 <http://www.lib.byu.edu/~scm/sampler/index.html>.

Special Collections Guidelines. University of Chicago. March 2002 <http://www.lib.uchicago.edu/e/spcl/using/access/guidelines.html>.

Special Collections Services. 5 July 2001. Claremont Graduate University. March 2002 <http://voxlibris.claremont.edu/sc/access/services.html#citing>.

University of Louisville, University Archives, Frequently Asked Questions. January 2002. University of Louisville. March 2002 <http://library.louisville.edu/uarc/faq.html>.

Yale Style Manual, Table of Contents. 27 September 2001. Yale University. March 2002 <http://info.med.yale.edu/caim/manual/contents.html>.

Chalou, George. 1984. "Reference." In *A Modern Archives Reader: Basic Readings on Archival Theory and Practice*. Washington, D.C.: NARS: 257-263.

Conway, Paul. 1994. *Partners in Research: Improving Access to the Nation's Archive, User Studies at the National Archives and Records Administration*. Pittsburgh: Archives and Museum Informatics.

Cox, Richard J. 1992. "Researching Archival Reference as an Information Function: Observations on Needs and Opportunities." *Reference Quarterly*. 31 (Spring): 387-397.

Dewald, Nancy H. 1999. "Web-Based Library Instruction: What is Good Pedagogy?" *Information Technology and Libraries*. 18 (March): 26-31.

Donaldson, Kelly A. 1999. "Library Research Success." *Internet and Higher Education*. 2, no. 4: 237-51.

Gilliland-Swetland, Anne J. 2001. "Popularizing the Finding Aid: Exploiting EAD to Enhance Online Discovery and Retrieval in Archival Information Systems by Diverse User Groups." *Journal of Internet Cataloging* 4, no. 3-4: 199-225.

Kocour, Bruce G. 2000. "Using Web-Based Tutorials to Enhance Library Instruction." *College and Undergraduate Libraries*. 7, no. 1: 45-54.

Koening, Melissa and Eric Novotny. 2001. "Online Course-Integrated Library Instruction Modules as an Alternative Delivery Method." In *Library User Education*. Lanham: Maryland: Scarecrow Press.

Pugh, Mary Jo. 1984. "The Illusion of Omniscience: Subject Access and the Reference Archivist." In *A Modern Archives Reader: Basic Readings on Archival Theory and Practice*. Washington, D.C.: NARS: 264-277.

Tibbo, Helen R. and Lokman I. Meho. 2002. "Finding Finding Aids on the World Wide Web." *American Archivist*. 64 (Spring/Summer): 61-77.

Tissing, Robert W., Jr. 1984. "The Orientation Interview in Archival Research." *American Archivist*. 47 (Spring): 173-178.

Vander Meer, Patricia. 2000. "Pushing the Limits: Creative Web Use in Libraries Related to Instruction." *Research Strategies*. 17, no. 4: 237-56.

Whalen, Lucile. 1986. "The Reference Process in Archives: An Introduction." *The Reference Librarian*. 13 (Fall/Winter): 1-9.

Yakel, Elizabeth. 2000. "Thinking Inside and Outside the Boxes: Archival Reference Services at the Turn of the Century." *Archivaria*. 49 (Spring): 140-60.

A Model for Web-based Archival User Education

As increasing numbers of finding aids and archival materials are digitized and mounted on the web, archives reference service will also change dramatically. Researchers might retrieve the archival materials they need on their own computers while sitting at home or in their offices (Whalen, 1985). As a result, the face-to-face interaction between reference archivists and researchers could be lost. To provide the user with the information he or she will need, "reference services must move beyond the reference desk, reference room, and stacks" (Yakel, 2000).

The results of the study show that archives and manuscript repositories are realizing this changing nature of reference and user education provisions at their institutions. Each institution provides a subset of the user education resource types and information types considered in the study in order to serve and educate users taking advantage of electronic access to collections. Some resource types and information types were widely implemented, while others appeared in few archives' web sites.

Figure 1. Resource Types Featured on Archives Websites
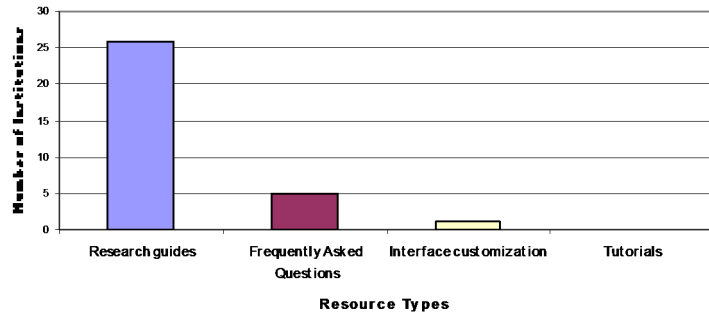
Based on the results of the study, most institutions provide research guides and Frequently Asked Questions, while many fewer offer tutorials or interface customization (see Figure 1). This may be because most FAQs and research guides are static documents that can be created quickly and easily. Tutorials and interface customization, on the other hand, are usually dynamic and require a higher level of expertise and a greater time commitment to create and maintain interactive sets of documents. In the future, however, tools may be designed to make more complex user education resource types easier for archivists and librarians to design and maintain. Most of the institutions in the study featured only one resource type on their web site, usually research guides. Archives may need to consider offering more than one type of resource to serve users who may have diverse information needs and different levels of experience in the use of archival materials.

repository and how to conduct research there, the reference interview also helps the user identify resources pertinent to his or her research objectives. Just as subject indexing helps the archivist recommend more relevant collections to the user, web-based subject access to collections can help the user find research collections independently. While basic subject access cannot be a substitute for the service provided during the reference interview, subject-indexed search tools and research guides can help the researcher frame his or her inquiry within the descriptive vocabulary of the institution to arrive at a starting point for research. The goal of the user education resources implemented on an archives web site should be to answer as many questions and remove as many doubts in the minds of the remote user as possible. In situations where the repository does not offer a particular service, such as photocopying, it is preferable to mention that these services are not available rather than leaving them unaddressed and potentially ambiguous.

Finally, to answer any questions the researcher might have that could not be addressed by the basic repository information guides or subject access provisions, email and chat reference can be utilized. Unlike the more general information resources, these services can resolve specific challenges for the user by providing information tailored to the information need. Email and chat most closely resembles the face-to-face reference orientation interview, but take place in the virtual realm.

Providing user education information electronically can help remote users understand the methods of archival research more completely. Information can be presented that educates the researcher about institutional policies and procedures, as well as subject information about collections contained within the repository. Frequently Asked Questions pages, research guides, tutorials, interface customization, and email and chat reference are practical techniques for disseminating this information electronically. As increasing numbers of resources are offered electronically, web-based archives education resources can provide essential support and high-quality reference services that users will require for research endeavors of the future.

Works Cited

Babbie, Earl. 2001. *The Practice of Social Research*. Belmont, CA: Wadsworth: A22.

Figure 3.

|	Archival Orientation	Intellectual Access	Physical Access	Utilization
Definition	Nature, significance of materials; definition of primary sources	Locating relevant materials within collection using metadata	Requesting, viewing collection materials	Policies concerning end use of archival materials
Information Types	Primary research	Resource identification; finding aid interpretation	Reader requirements; reading room procedures	Photoduplication; publication; copyright policies
Implementation	FAQ; research guide; tutorial	FAQ; research guide; tutorial; interface customization	FAQ; research guide; tutorial	FAQ; research guide

Conclusion

In summary, the basic elements of the orientation interview can be translated into easy-to-use web-based resources. Information from the Orientation Checklist (Tissing, 1984) that could be included are the types of materials held in the repository, what an archival collection is, what a finding aid is and how to interpret it, hours of operation, and rules and regulations of the repository such as duplication, copyright, and publication policies. Much of this standard information could be presented to the user in a "Frequently Asked Questions" page or an "Archives User Information" research guide, or presented using tutorials or interface customization.

In order for these web documents to be effective, they need to be easy to find, easy to read, and easy to understand. The location of these help screens should be intuitive and could be context-specific. In addition to linking them to the repository's home page, help screens could be placed at the point of use. For example, the information on how to interpret finding aids should be linked to the finding aid listing, or even included in the style sheet applied to each of the repository's finding aids. The web pages should be clear and user friendly (Dewald, 1999). The goal of the organization of these documents should be to allow the user to scan the pages quickly to find the information he or she needs.

In addition to giving the user general information about the

Figure 2. Information Types Featured on Archives Websites
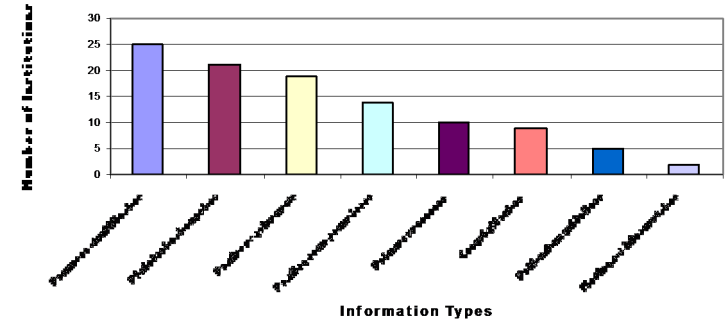
The information types that were most widely implemented were those that focus on access to collections, such as resource identification information, reader requirements, and reading room procedures (see Figure 2). Information about the end use of collections, such as copyright policies and publication guidelines, and information to help orient users, such as information about primary sources and finding aid interpretation, were implemented less frequently. This may be because the foremost goal of the archivists in creating an archives web site was to support and increase access to the collections held at the repository, resulting in an emphasis on access information but not many other types of user education information. In other words, since the archivists primarily may be concerned with helping users access materials, they may not be as concerned with helping researchers to put the resources to use.

Orientation information may not be presented on many archives' web sites because archivists might assume that people visiting their web site will already know what primary resources are, what a finding aid is, and how to use these materials in their research. With the availability of finding aids on the Web, however, more individuals who have little or no experience in archival research will retrieve finding aids in their standard web searches.

In order to fully understand the reasons why certain information types have been included or omitted, the archivists who designed the web sites should be surveyed. This is an area where further investigation could clear up any ambiguities in

the data, and could provide explanation for the implementation choices that have been made.

User education resources can make research in archival collections less daunting for beginners and more efficient for experienced researchers. The resource and information types scrutinized in the content analysis can be organized into a model with four main categories: Archival Orientation, Intellectual Access, Physical Access, and Utilization. Each category will be defined and strategies to implement the category within an archival repository's web site will be suggested.

Archival Orientation

Archival orientation resources provide context for the user's research by explaining the nature and significance of the materials within the repository. These resources may also include basic information about what primary sources are, what formats appear within the collection, and any interesting historical notes or subject strengths of the collection. Of the information types examined in this paper, the most appropriate type for this category is the definition of primary resources.

This information can take the form of a Frequently Asked Questions page, including questions such as "What is an archive?" "What is a manuscript?" "What is a special collection?" or "Why are these materials important?" A research guide could also work to present this information under a title such as "About Manuscript Collections," "About Primary Sources," or "Help." The first one or two introductory screens in an archives tutorial also could provide this orientation and contextual information. Another resource type that could provide some of this archives orientation information is the web exhibit. Users can get an idea of the kinds of resources available in an archival repository by looking at digital images or textual examples of them.

Intellectual Access

Intellectual access tools help the researcher find relevant materials within the repository. These are distinguished from Physical Access Resources defined below because at this stage users are locating the metadata associated with physical materials, not the physical materials themselves. Information types belonging in this category are resource identification guides, including online catalog or archival information system instructions and subject guides to the collection, as well as finding aid interpretation information. These resources help the user negotiate the various systems and metadata schemas to locate relevant information sources. This could be presented in a FAQ

featuring questions like "How do I find materials in the archives?" or "What is a finding aid and how do I use it?" Research guides can offer step-by-step instructions on how to search and retrieve using available bibliographic tools and databases. Tutorials are especially well-suited to this kind of results-oriented learning, and could lead the user through a series of practice search exercises. Interface customization and research profiles could support subject-related research guides.

Physical Access

Physical access tools inform the user of the rules and procedures associated with requesting and viewing the collection's materials. Information types to feature here are reader requirements and reading room or circulation procedures. Other kinds of information that could be grouped with physical access are hours of operation and location of the repository, whether there are researchers for hire, or if there are fees associated with remote research requests. This information could be presented as FAQs, such as "Once I find a relevant collection, what do I do next?" "How do I request materials?" "What happens in the reading room?" or "What if I can't come to the library in person?" Research guides could be created under titles like "Viewing Materials" or "Research Procedures." Tutorials giving instruction on intellectual access could end usefully by including guidance for requesting and viewing materials within the repository.

Utilization

Utilization information deals with the end use or application of the information discovered from the materials within the repository. Utilization resources help the researcher use materials in the collection appropriately and legally. Information types that could be included in this category are photoduplication policies, copyright policies, and publication guidelines. This information can be disseminated effectively through FAQs such as "Can I make photocopies/take photographs / scan or optically recognize these materials?" "Can I publish this?" or "How do I cite this?" Research guides could be created as "Usage Policies" or "Copyright and Publication Guidelines."

The following chart (Figure 3) represents a matrix of the four archives user education resource categories, their definitions, associated information types, and implementation resource types.