

On the Centennial of Peking University: Will Electronic Records Created Today Survive
to the Bicentennial?

Sarah E. Thomas

Carl A. Kroch University Librarian

Cornell University

Ithaca, New York 14850

USA

Abstract

Electronic records are becoming an increasingly important source of documentation of university administrative, legal, political, and historical activity. University libraries, as the home of archives, must take a leadership role in understanding the characteristics of electronic records, facilitating their transfer, description, and securing their long-term access.

In 1868, thirty years before the founding of Peking University, Ezra Cornell, a self-made man with only a few years of formal education, established Cornell University in Ithaca, New York. Cornell had prospered from his investment in a new technology, the telegraph, and he devoted a fortune to building an innovative, co-educational university --"an institution where any person can find instruction in any study." Ezra Cornell was a man ahead of his time, and his words still resonate today. In 1995 the University Library undertook the digital conversion of selected materials of the Ezra Cornell papers, including letters, diaries, photographs, and other documents of Cornell's

life. The purpose of the digitization project was both to gain insight into issues surrounding the digital conversion of manuscripts and older printed texts and illustrations as well as to increase access to the papers of our founder. To place these important manuscript materials in context for a wide audience, we produced a Web site "Invention and Enterprise: Ezra Cornell, A Nineteenth Century Life."

(<http://rmc.library.cornell.edu/ezra-exhibiy/entrance.html>). To our delight we found that 9 year-old students at a local elementary school had discovered the Ezra Cornell Web Site; they found the digitized papers invaluable in their unit on local and New York history. The Ezra Cornell Papers fit within a suite of materials which Cornell has converted in the Making of America Project, which includes 900,000 digital images converted from journal titles of importance in understanding the history of the nineteenth-century. The experience gained through these projects is helping us develop standards for digital conversion and imaging, new interfaces for improved access, and economic data on the cost and maintenance of digital data. One of the lessons we have learned from the Ezra Cornell papers is the difficulty of maintaining a stable environment for the media. The server on which the files were stored was simply a workstation on a person's desk in the Rare and Manuscript Collections Division of the Cornell University Library, and when the workstation was turned off, either intentionally or inadvertently, access to the Ezra Cornell papers ceased. In addition, the software we used to manage the scanned images became outdated, and the papers were unavailable for several weeks in the winter of 1997. These unfortunate incidents point to the need to have a solid plan for the permanent preservation of digital services and for securing their longevity through migration to the latest technology and software in advance of obsolescence. This is true

in the case of paper materials that have been digitized, but it is absolutely essential when we are working with digital documents that originate in electronic form, and for which no permanent paper analog exists.

Increasingly we are encountering material whose essence is digital. Of particular interest within the University context are electronic records. The long-term responsibility for these records falls under the purview of the University Archives, a component of the Division of Rare and Manuscript Collections of the Cornell Library that promotes knowledge and understanding of the university's origins, aims, programs, and goals; and facilitates effective records management. This includes permanent retention of official records and reports of the university, its officers and component parts; private papers of faculty, students, staff and alumni; maps and architectural records; audiovisual materials including still photographs and negatives, motion picture film, oral history interviews, audio and video tapes, and artifacts and ephemera documenting the university's history. For example, the University Archives holds student transcripts and grade reports; radioactive materials licenses and Safety Committee records; original patents and trademarks; and of course, the Ezra Cornell papers. The University Archivist routinely researches our files to determine underground utilities locations that may affect construction projects, the growth of the student body over the past two decades, or the height of McGraw Tower (173 1/2 feet) on top of which a mysterious climber speared a pumpkin in October 1997. It is the University Archives that provide background on

when the first Chinese student (Sao-ke Alfred Sze) enrolled at Cornell (1897) or that produces a picture of the Chinese students' club from the 1912 yearbook.

Like most bureaucracies, Cornell University has a complex administrative structure. We have a central administration and eleven colleges and professional schools, including seven privately funded units and four units supported in part by the State of New York. At Cornell, while records have been kept at both the college level and the university level, in diverse paper and electronic systems, documentary evidence is generated and organized largely at the college level. Cornell, as have many of its peer institutions in American higher education, has begun to move to a new environment that will use advanced technologies and integrated systems to simplify, streamline, and make more efficient the variety of administrative functions of this large, complex, and diverse campus community. "Project 2000," the central effort to automate major administrative systems, has begun redesigning administrative processes, policies, and structures at Cornell. Project 2000 will re-engineer five major administrative activities and their associated automated systems, including Finance, Student Records, Human Resources, Alumni Affairs and Development, and Sponsored Programs (Grants). Many documents will soon exist primarily in electronic form. The University archivist, knowing the demands placed on the paper records, has anticipated some of the problems of archiving electronic records. In 1997 Cornell University Library received funding of almost \$125,000 from the National Historical Publications and Records Commission of the National Archives and Records Administration of the United States to examine the relationship between existing paper records and future electronic documents and to develop recommendations for university policies and practices that address archival

concerns for the retention, preservation, and research use of electronic records. In order to do so, it will evaluate the functions and activities that should be present in an electronic records program, and identify the various incentives that can contribute to creator and user support for electronic records management concerns. In the current paper-dominated world, colleges and administrative units of the university review materials in their files and transfer them to the Archives when the documents are no longer used on an active basis. One of the incentives for transfer of files is to save space in the colleges. The Archives stores these copious files in a secure environment, organizes and preserves them, and provides access to them. When colleges and units no longer have the physical files choking their work area, will they be as likely to transfer valuable records to the University Archives, or will these electronic records suffer from neglect and disappear into the ether, creating a serious gap in our cultural and historical knowledge?

At the turn of the twentieth century, the University is making the transition from a primarily paper-based organization into an electronic organization. Project 2000 represents a strong movement on the part of the University to transform its way of doing business and to rely increasingly on digital data. Although the University's first concern is to build a system that supports the current use of data, the experience of the Archives leaves little doubt that there will be considerable, and possibly even increased demand for access to archival data in digital form. In anticipation of this need, the Cornell electronic records project has considered a number of key issues, to which it will seek answers during the next two years.

Among the questions the electronics records project will address are

1. How is record "ownership" determined in an automated, integrated system?

"Ownership" is currently one of the determinants in deciding who archives material. IN the paper environment, the unit which hold the physical object can be considered the owner, but in a distributed system in which units may contribute various pieces of digital data, the concept of ownership for archival purposes is much murkier. The Archives must capture all the relevant data to present a complete picture.

2. When the permanent "record" is actually being created "on the fly" from data stored in multiple databases, how should this virtual permanent record be compiled and stored?

3. What is the role of the University Archives in the administration of long-term storage of permanent records in electronic form?

Ultimately, the project staff anticipates making recommendations to the designers of the Project 2000 systems that will facilitate electronic archiving. Electronic records may include metadata that will trigger automatic archiving at certain time intervals and which will indicate levels of access to the data, much of which will contain privileged information and which will require authorization to use. The goal will be to build into the system a user-friendly component that gathers information from the creator of the data about the nature of the record being created. This information can serve both to launch the transfer of the electronic record from active use to archival status, as well as to provide the essential description elements for access to the data. As the number of electronic records grows, the volume of data will necessitate the automation of certain aspects of the record selection and retention process as well as the cataloging process. Automation of these activities provides an opportunity to create a more efficient and standard archival practice.

In addition to determining what data to archive, and when to archive it, Cornell and other universities must determine how to archive it in a way that ensures its long-term viability and accessibility. In a decade of experience with digital files, the Cornell University Library has encountered difficulty in providing continuously reliable access because of hardware instability and obsolescence, proprietary file formats, loss of institutional memory because of staff departures, inadequately documented software, and funding insufficiencies. As we move from the experimental and opportunistic phase of digital library development into a production environment, in which information originates digitally, it is imperative that we develop standards for digital archiving to ensure that we preserve today's information for use and analysis by future generations. While in the short-term, the safest technique may be to create paper duplicates of digital files, this will quickly prove to be both too cumbersome and inadequate to represent the true nature of the digital information. Presently, Cornell's strategy will be to develop a policy for refreshing and migrating data on a scheduled basis, and to investigate whether it is most cost-effective to store data locally, regionally, or nationally. The analysis will consider the need for immediate access or less urgent access and will take into account the readiness of external enterprises to provide dependable storage.

Electronic records are valuable sources of administrative, legal, political, or historical interest. As we enter the twenty-first century with an increasing reliance on digital data, it is vital that we establish policies for the retention and maintenance of digital files and that we create the technical infrastructure to undergird our policies and practices. Our

institutions must be able to document for the scholars of 2198 the world of 1998,
regardless of the format in which the records manifest themselves.