UNIVERSITY OF CALIFORNIA, LOS ANGELES UCLA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

Graduate School of Education & Information Studies
P.O. Box 951521
Los Angeles, CA 90095-1521

CONGRESSIONAL TESTIMONY

Anne Gilliland-Swetland
Assistant Professor, Department of Information Studies
University of California, Los Angeles
Before the Subcommittee on Government Management and Technology
of the U.S. House of Representatives
Washington D.C.
April 4, 2000

Introduction

Mr. Chairman and Members of the Subcommittee, thank you for inviting me here this afternoon to present testimony with regard to the National Historical Publications and Record's Commission's efforts to preserve historical documents and electronic records and the bill to reauthorize the Commission for five years. I am a professor in the Department of Information Studies at the University of California, Los Angeles. My research, publications and teaching focus on the archival management of electronic records--those records that are electronically generated and or maintained by computers--and on digital archives development. I am particularly engaged with issues relating to the preservation of, and public access to trustworthy records in the digital environment. I am the co-Director of the NHPRC-funded US-InterPARES Project, the American component of an international project that is developing theoretical, technological, policy, and educational requirements for the preservation of permanent records created by electronic systems. I am also a member of the archival advisory board for the NHPRC-funded Methodologies for the Long-term Preservation of and Access to Software-dependent Electronic Records Project underway at the San Diego Supercomputer Center (SDSC). This latter project is building prototypes of information architectures that can be used to preserve software-dependent electronic records in a variety of institutional settings.

The Electronic Records Challenge

Many records generated by society in the course of its activities need to be preserved permanently as critical instruments of accountability, as means of protecting individual and corporate rights, and as sources of information for research and study. The ubiquitous implementation of rapidly evolving information and communications technologies means that almost all contemporary records that will require permanent preservation have an

electronic component, often with no paper counterpart. This implementation is also re-shaping traditional bureaucratic structures and affecting record creation and record-keeping processes in ways that we do not, as yet, well understand. While the implementation of information and communications technology has obvious advantages for activities such as electronic commerce, electronic government, clinical practice, and academic and industry research, it also presents society today with one of its greatest technological challenges—*how to guarantee the long-term preservation, trustworthiness, and accessibility of vast quantities of electronic records in the face of continual and rapid obsolescence of computer hardware and software, vulnerable and impermanent storage media, and manipulable electronic systems.*

As a society, we must now address the fact that, unlike with paper records, our present lives and activities as well as our future cannot tolerate even a small amount of benign neglect when it comes to electronic record-keeping and electronic records preservation. The moment when technological and fiscal support for electronic records is withdrawn within the organization that created them, electronic records instantly start to decay, sometimes disappearing almost immediately. Moreover, a single break in the chain of custody for electronic records or a mistake in backup or other media handling (to say nothing of security breaches and deliberate damage) can leave the integrity of electronic records as trustworthy sources open to challenge--a critical issue for anyone who must depend upon records.

A number of specific issues are particularly critical, and many of these have been at the center of the electronic records research projects funded by the NHPRC:

* *Identification of records:* Two of the most basic questions all organizations face in the electronic environment are *what is the record today?* and *how do we identify it for archival preservation?* As technological capabilities evolve and systems become increasingly networked, it is becoming increasingly difficult to establish what actually comprises electronic records. Organizations create, manipulate, and store operational data (e.g., geological data, market profiles, scheduling projections); associated transactional metadata (e.g., audit trails, use statistics); and strategic information (e.g., annual reports, committee documentation, executive correspondence, product designs, formulae and patents) within highly distributed networked record-keeping and information systems. The task of the records manager and the archivist is to determine which aspects of these system contents comprise "the record." Another complication is that the life-cycle of records in the digital environment is being affected by organizational knowledge management activities. Fewer records (especially operational data and transactional metadata) are likely to be systematically retired and sent to the archives because of their potential for being used in more than one way, or by more than one group of users. Electronic records are being re-used for projects other than those with which they were originally associated (e.g., data warehousing), analyzed and cross-compiled in new ways for management purposes (e.g., data mining), and re-tooled into information products (e.g., re-purposing, multi-versioning).

* *Evidential requirements:* meeting legal requirements for the admissibility of records as evidence sets the bar for record-keeping and preservation requirements at a higher level than that for any other kind of information. For example, it must be possible to demonstrate that records have been created and maintained appropriately in the course of daily activity, throughout both their active and their archival life. In order to understand how records were involved in decision-making and other business activities, it must also be possible for preserved records to be "rendered" in eye-readable form in the same way that they were presented to the records creator.

* *Juridical diversity*: organizations create electronic records for which inadequately-defined or even competing legal, regulatory, and professional requirements and standards exist. Moreover, not all organizations are subject to the same, or all records and record-keeping laws or regulations. Indeed, as electronic commerce and research activities becoming increasingly international and collaborative, electronic records will be created that must meet the requirements of multiple legal and regulatory jurisdictions.

* *Institutional diversity:* not all organizations creating electronic records have the same kinds of records, mandate, needs, users, or resources. We do not yet fully understand what is the same and what is different about electronic records creation and management in diverse organizational contexts, and, therefore, which electronic records management and preservation models can be applied and which not.

* *Technological diversity*: government, clinical, academic, industry and many other types of organizations employ multiple, constantly evolving, and frequently incompatible computer systems and operating environments to create their records. A major record-keeping system may migrate to a new software-hardware configuration as frequently as every three years, making it difficult or impossible to access and read records created on previous configurations.

* *Technological dependency:* due to the inherent characteristics of increasingly prevalent electronic media forms, it is no longer viable to consider preserving most records in non-electronic form, since to do so would result in the loss of critical information as well as an equally critical diminution of evidential value. Examples of complex record forms include:

  • *Multimedia:* systems that integrate multiple digital media simultaneously such as still and moving images and sound, often created as non-linear documents connected by a series of hyperlinks (such as World Wide Web pages).

  • *Smart documents:* contain embedded pointers to external sources of information that can be automatically accessed when a document is called up.

  • *Virtual documents and database views:* virtual documents that do not necessarily have any physical or long-term existence and are created through the juxtaposition

of several different pre-existing discrete pieces of data pulled together by of a set of relations or pointers.

• *Geographic Information Systems:* bring together maps and data regarding natural, demographic, and cultural resources in any number of layers to display data visually. For example, a database of addresses which can be converted into location points on a map; lines on a map indicating traffic volume on a highway; or demographic data such as ethnicity displayed as shaded polygons (shapes) on a map.

Not only is it difficult to convert such materials into paper of software-independent form, it is often strategically undesirable for the parent organization, whose knowledge management requirements may demand that as much organizational information (including records) as possible is maintained in searchable, interoperable, and collocatable electronic form.

* *Trustworthiness:* while trustworthiness is often a transparent quality of records, an overt effort has to be made both to ensure and to demonstrate the trustworthiness of electronic records, whether active or archival. Trustworthy records are critical in the present for accountable government, e-commerce, and for research; and in the future as accurate recorded memory. For electronic records to be trustworthy, they must be reliable and they must be authentic. Reliability is guaranteed by ensuring that appropriate record-keeping, security, and file maintenance practices and policies are in place and implemented for active records. Authenticity requires that it can be demonstrated that the physical and intellectual integrity of the records has not been compromised at any point during their life, including their archival life.

* *Awareness of records creators:* there is still very little awareness among records creators of electronic records issues, resulting in practices that are hazardous to the continued existence of the records, as well as to their evidential integrity. Many records creators are unaware that these documents and databases they are creating might be records that are subject to disposition requirements. Many activities that create electronic records, such as collaborative projects, have not identified which party should be responsible for the long-term management of the project's electronic records. Also hindering the development of an electronic records consciousness has been a lack of cost models and implementable systems requirements for preserving electronic records, and the necessary human expertise in electronic records management and preservation.

* *Public access:* counter-intuitively perhaps, very few preserved electronic records have yet to be made available online to the public, and this must surely be one of the next areas for electronic records research and development. Few archival organizations are yet sufficiently far advanced with their electronic records programs to consider making the records available online, and if they are, there is a dearth of models for them to follow in doing so.

* *Expertise:* one of the most pressing issues for the preservation of electronic records is the critically small pool of archival expertise currently available—perhaps no more than a couple of hundred individuals worldwide. It is critical that robust university and continuing education programs be developed that not only prepare current and future archivists to work with electronic records, but that also prepare a new generation of archival educators who are able to teach and conduct research in this area.

The Evolving Role of the Archivist

Responsibility for ensuring the preservation, authenticity, and accessibility of records that must be preserved permanently lies with the archival profession. New information and communication technologies, however, have not only transformed how business is conducted and the nature of the resulting record, they are also transforming the practices associated with archival management. The archivist is no longer a passive recipient of inactive or historical records, but a proactive advocate for the records who also:

- participates in the development of overall information policy and organization;
- provides archival input during the design and implementation stages of electronic record-keeping systems;
- devises strategies to communicate archival needs effectively to the resource allocators, systems designers, creators, managers, and end users of electronic records systems;
- conducts education and training programs in electronic records management and preservation within his or her organization; and
- identifies procedures for ensuring long-term preservation of electronic records.

The NHPRC Role in Addressing the Electronic Records Challenge

The National Historical Publications and Records Commission is the only national funding agency that is directly addressing the issues I have outlined above. With the articulation and implementation over the past nine years of its exciting and vitally important electronic records research agenda, the Commission has single-handedly been responsible for most of the knowledge gains and development activities that have occurred in this area in the past decade. This agenda, which addresses some of the most complex and resource-intensive technology-related issues that the archival and historical professions, records creators, and society as a whole have to face, has resulted in concrete outcomes such as the development of electronic records programs and pilot projects in many state and university settings, as well as sets of functional requirements, metadata schema, and industry standards for electronic record-keeping.

Another major outcome of NHPRC's electronic records research agenda is that it has attracted the kind of multi-disciplinary expertise and state-of-the-art research and computational resources that are necessary to tackle electronic records management and preservation in the most substantive ways. It should also be mentioned that the Commission also has an admirable track record in program development and advocacy in the area of electronic records. The Commission provides archivists and related professionals with advice and expertise on developing and carrying out projects, and

serves as a clearinghouse for information on electronic records activities. It also strives to build synergy between complementary electronic records activities. The NHPRC's new initiative to broaden the base of archival expertise in the area of electronic records addresses the important area of translating research outcomes into practice by building awareness, understanding, and expertise in electronic records management and preservation in the archival and record-creating communities.

I want briefly to discuss two of NHPRC-funded projects that are currently underway. These two projects exemplify not only the extensive and complex nature of research and development to date in the area of electronic records that has been funded by the NHPRC. They also exemplify how the NHPRC has worked to ensure interaction between complementary projects, its relevancy to a range of research communities, and its ability to facilitate projects that strategically leverage additional funding sources. Without the NHPRC's electronic records research agenda and its funding program, such research would simply not be possible.

The InterPARES and SDSC Projects

In June 1999, the NHPRC funded the International Project on Permanent Authentic Records in Electronic Systems (InterPARES). In January 2000, the Commission funded the San Diego Supercomputing Center's Methodologies for the Long-term Preservation of and Access to Software-dependent Electronic Records Project. The two projects are working closely together because of the inter-dependent nature of their research. InterPARES is generating theoretical, technical, policy, and educational requirements for the preservation of authentic electronic records based on an analysis of a wide range of organizational settings and legal and political jurisdictions. The SDSC Project is designing information architectures that will build upon these requirements and that will be scaleable to situations other than very large archival repositories such as the National Archives and Records Administration.

The InterPARES research project is a three-year project examining issues relating to the long-term preservation of electronic records in ways that maintain and demonstrate their authenticity. An inter-disciplinary team of researchers drawn from archival science, preservation management, library and information science, computer science, and electrical engineering, and an industry group representing global biocomputing and pharmaceutical industries, are working together with the national archives of several countries, including the United States National Archives and Records Administration, to identify and model the form, function, and structure of records contained in electronic systems in a variety of organizational and social contexts. Participating countries include the United States, Canada, Italy, The Netherlands, Sweden, Finland, France, Portugal, England, Scotland, Ireland, Australia, China and Hong Kong. In addition to funding from the NHPRC, major funding contributions have been made by Canada's Social Science and Humanities Research Council, the National Archives and Records Administration of the United States, and the Italian National Research Council. Universities and national archival institutions in participating countries have also committed financial and research resources to the project.

While such research models are now commonplace in some other disciplines, InterPARES is the first example of a collaborative, multi-funded, multidisciplinary project emanating out of the archival community. The model was adopted because the research would be tackling issues that are of critical concern to governments, industry, and archival institutions worldwide, but that have failed to be satisfactorily addressed unilaterally by any of those sectors. This approach has allowed InterPARES to:

- bring diverse disciplinary perspectives to bear on resolving seemingly intransigent questions;

- have enough scope, granularity, and depth of expertise to parse large research questions into smaller constituent questions, and thus divide the research into more manageable, yet synchronized components; and, equally importantly,

- generate the high profile for the research and its results that is required to get the attention of policy makers, standards developers, resource allocators, academicians, and others who play key roles in translating the outcomes of the research into practice and ultimately into the fabric of daily life.

The broad goal of the InterPARES Project is to develop the theoretical and methodological knowledge essential for the permanent preservation of records generated electronically, and, on the basis of this knowledge, to formulate model policies, strategies, and standards capable of ensuring their preservation. American InterPARES researchers are also focusing on the systems design implications of InterPARES by translating the resulting theoretical models, templates, and typologies into systems design requirements and metadata models for implementation in different organizational, social, and national domains, and across different types of electronic records.

The Project has been broken down into four domains: I. Conceptual Requirements for Preserving Authentic Electronic Records: II. Appraisal Criteria and Methodology for Authentic Electronic Records; III. Methodologies for Preserving Authentic Electronic Records; and IV. Framework for Developing Policies, Strategies and Standards. One additional group has been established that is developing a multi-lingual, multi-national glossary of terminology used in the project. One issue that any multi-disciplinary electronic records work faces is that the same terms are often used completely differently in different communities (e.g., archival and computer science usage of the terms “records” and “archives”).

As part of Domain I., the project is in the process of conducting and analyzing extremely detailed case studies of electronic records systems of diverse types in a range of organizations (for example, complex databases, geographic information systems, laboratory records, and interactive Websites in government agencies, universities, banking, biocomputing, and museum settings). From this analysis, we are deriving an understanding of the nature of the electronic record and the extent to which its intellectual, if not its physical form remains the same as that of traditional records. We are also building a typology of elements within different kinds of records that are crucial to the

establishment and maintenance of the authenticity of that record while it is still current and when it becomes historical. This typology will then become the basis of the technical and policy requirements for preservation management systems and strategies. Domains II and II are using modeling techniques to describe the components of the appraisal and preservation processes, and also to analyze different methods and strategies currently in place or being developed in archival institutions. Domain IV will take the results of the work conducted in the other domains, and address policy and standards implications.

The San Diego Supercomputer Center is the leading edge facility for the National Partnership for Advanced Computational Infrastructure (NPACI). SDSC's Methodologies for the Long-term Preservation of and Access to Software-dependent Electronic Records Project, which has just commenced, builds upon its experience working with the National Archives and Records Administration on the ongoing NARA/DOCT Electronic Records Management Project. The NHPRC-funded project will allow SDSC researchers to take what they have learned from working with NARA, as well as from the work of InterPARES and other recent and ongoing NHPRC-funded projects, and develop and test prototypes for preserving and making accessible software-dependent records in ways that are scaled to the needs and resources of different kinds of organizations such as state and local givernments and universities. The SDSC project will also include the creation of useful tools for archivists to use to preserve and provide access to electronic records over time.

Conclusion

Important and exciting as these projects and others currently underway are, they address only certain key issues, and there remains an immensely important role for the National Historical Publications and Records Commission to play with regard to furthering research and development in the area of electronic records management and preservation. Every organization in this country creates records, and very soon, some part of almost all those records will be electronic. Moreover, electronic commerce, as well as electronic government will need to rely heavily upon the trustworthiness of those records. There are many critical areas that still need to be addressed—translating research outcomes into practice through the development of basic and affordable software tools, the design and implementation of multi-faceted education programs for archivists and records creators, and the building of models for widespread access to archival electronic records, to name but a few.

Mr. Chairman, this concludes my prepared remarks. However, I would be pleased to answer any questions you or Members of the Subcommittee may have.