II. ELECTRONIC RECORDS MANAGEMENT, PRESERVATION, AND ACCESS

A. National Archives' Electronic Records Archives

The increasing use of electronic records in Congress during the last five years has dominated records management issues. Information technology professionals within Congress, legislative agencies, and the National Archives and Records Administration recognize that long-term preservation and access to electronic records will be critical to the twenty-first century's history of the U.S. Senate and House of Representatives. The National Archives and Records Administration is presently undertaking research and development programs to achieve this long-term preservation and access in partnership with the Department of Defense, the San Diego Supercomputer Center, and other electronic research institutions. The preliminary results indicate that while the volume of today's electronic records is massive, the challenge is not insurmountable. A technological breakthrough promises the development of an Electronic Records Archives at NARA that can preserve any kind of electronic record, free it from the format in which it was created, retain it indefinitely, and enable requesters to read it on computer systems now in use and coming in the future.

Through agreements with the Defense Advanced Research Projects Agency, the National Science Foundation, and the Army Research Laboratory, NARA is co-sponsoring research at institutions such as the Georgia Tech Research Institute, the University of Maryland, and the San Diego Supercomputer Center, the leading-edge site of the National Partnership for Advanced Computational Infrastructure. This research is developing a comprehensive information management architecture for long-term preservation and access to digital information, as well as applying advanced technologies that are seen as core to the next generation national information infrastructure and as key enablers of electronic commerce and electronic government. These technologies include XML, which is being tested as a powerful and flexible means for preserving, managing, and delivering very large and highly diverse collections of electronic records.

The first round of research in these initiatives produced very promising results, articulating both the key components and the core processes of an Electronic Records Archives capable of han-

dling the hundreds of millions of electronic records that NARA will need to preserve. When implemented, the Electronic Records Archives will be capable of handling millions of records, accommodating a variety of electronic record formats, and providing continuing access to authenticated and preserved electronic records into the indefinite future. The concepts advanced in the research projects were demonstrated through empirical tests, including tests on electronic records of the Congress. There are substantial research and development tasks that remain to be addressed in order to translate these promising beginnings into operational capabilities. This initiative, however, is critical to the future of the archives of Congress, Executive branch agencies, and other repositories of electronic records. Section F describes the status of XML within Congress.

In the meantime, both the Senate and the House have revised their records management instructions and manuals to encompass National Archives guidelines and directives on electronic records management. Until electronic records management systems with the capability of long-term preservation and access are developed, congressional offices are making hard copies of their computer-generated records. The Center has accessioned electronic records from the Ervin Committee, the Congressional Budget Office, the Senate Home pages of the 104th and 105th congresses, the Joint Inaugural Committee Home Page, and the home page of the House Committee on Small Business. Other computer records, indexes to records systems, and word processing files have come to the Center with accessions of paper records. The Center, with other Archives offices, is studying how to preserve these disks, CDs, and tapes containing proprietary software and other encumbrances to their preservation and access.

The Center for Legislative Archives and the Electronic and Special Media Records Services Division at the National Archives have followed the work of the Legislative Branch SGML Technical Committee. The Technical Committee began in 1997 to coordinate the development of major systems in the Senate and House for the creation, transfer, publication, and preservation of all bills, resolutions and amendments. The membership includes staff members from the information technology offices in the Senate, House, Government Printing Office, Library of Congress, other legislative agencies, and in 1998, the National Archives as observers. During the monthly meetings of the Technical Committee, questions of long-term preservation are directed to the Archives, and in September, 1999, a presentation on the National Archives/San Diego Supercomputer study was given by the National Archives project director.

14

There have also been periodic one-on-one meetings between the information technology staffs of the House and Senate and National Archives staff. These meetings have included updates on research projects, exchanges on the Legislative Information Systems, and discussions about requirements and capabilities at the National Archives for long-term preservation and access to electronic records. As the current period of planning and prototyping evolves into one of significant transfers of electronic records systems, the communication between congressional offices and the National Archives will increase in frequency and intensity.

B. Technology Task Force

Established at the December 9, 1996 meeting of the Advisory Committee, the task force was charged with six goals: (1) Locate and identify major electronic information systems planned or currently in use in the House and Senate, including all systems that maintain official records; (2) Identify the permanently valuable information in these systems; (3) Develop a list of National Archives and Records Administration transfer options for electronic records; (4) Discuss and make recommendations for verification, security, and authenticity; (5) Offer recommendations on electronic record-keeping systems and data migration and preservation; and (6) Present a report in the fall of 1997.

At the September 1997 meeting, the Task Force reported that the House and Senate had made progress on inventorying existing systems. As a result, the House discovered that data from its old mainframes needed to be converted to transfer media specified by the National Archives. This was completed with assistance from the Library of Congress. The Senate archivist reported that the inventory of Senate systems was completed and that permanently valuable records in these systems were identified. The National Archives provided guidance on transfer procedures and options.

Recommendations on electronic record-keeping and security were compiled and distributed through three Senate publications: the *Records Disposition Procedures for the Office of the Secretary of the Senate; Records Management Handbook for United States Senators and Their Archival Repositories;* and *Records Management Handbook for United States Senate Committees.* Recommendations on electronic record-keeping for the House are presented in the section on electronic records in the *Committee Resource Guide.* Through these publications and office briefings, staff were alerted to the need to manage and preserve substantive e-mail.

With assistance from the National Archives, Senate committee offices transferred a variety of electronic records to the Archives.

These included indexes, home pages, and word processing tapes. Retiring senators sent mail management systems data to their designated repositories. It is apparent with the passing of each year that greater and greater amounts of information reside in electronic form. The Senate Sergeant at Arms estimates that at this time about forty percent of mail received by the Senate is electronic.

The Clerk of the House pointed out difficulties in working with changing technology and emphasized the need to continue to work with the National Archives as technology evolves. Consequently, it was agreed not to present a task force "final" report, but to continue to monitor progress to ensure that Congress' needs are met. Regular updates have been featured as the new legislative systems are developed.

C. Legislative Branch Standards

In 1996, the chairman of the Committee on House Oversight and the chairman of the Senate Committee on Rules and Administration directed the Clerk of the House and the Secretary of the Senate, respectively, to work together to establish common data standards for the exchange of legislative information. This directive was also included in the Conference Committee report for the FY97 legislative branch appropriations bill.

The Secretary and the Clerk created a working group on data standards and contracted with outside experts[3] to assist in this process. The working group also surveyed the offices and support agencies in the legislative branch who were engaged in the creation, exchange, and publication of legislative data.[4] Based upon the work of the consultants and the results of the survey, the Clerk and the Secretary submitted a joint report to the committees recommending the establishment of a data standards program for legislative information.[5] In April 1997, the report was approved in a joint letter from the chairman of the Committee on House Oversight and the chairman of the Senate Committee on Rules and Administration.

---

[3] Mulberry Technologies, Inc. Their report to the Clerk and Secretary is entitled: SGML for the Legislative Branch agencies; An Overview of Existing Applications and Recommendations for Development Approaches, January 1997.

[4] The agencies surveyed were: House Information Resources, Senate Sergeant at Arms, Government Printing Office, Congressional Budget Office, General Accounting Office, Congressional Research Service, and Library of Congress.

[5] *Recommendations for a Data Standards Program for Legislative Information*, A Report Prepared by the Secretary of the Senate and the Clerk of the House for the Senate Committee on Rules and Administration and the House Committee on House Oversight, April 1997.

The report confirmed the need for data standards to: (1) ensure continued effective management and control of legislative information; (2) more easily exchange data on a timely basis; (3) enable the officers of the House and Senate to certify the accuracy of legislative data; (4) help to reduce costs in preparing, managing, retrieving, and printing information; and (5) improve the functionality and flexibility of the systems used for preparing and managing legislative data. It was especially important to agree upon standards at that time because the primary systems for creating legislative data are becoming increasingly difficult to maintain and are being made obsolete by modern technology. As offices and agencies began to make plans to replace their current systems, there was, and continues to be, a clear need to agree upon standards for the creation and exchange of legislative data.

The management of data standards is a dynamic process and requires an ongoing management and policy structure. The report of the Secretary and the Clerk therefore included recommendations for the establishment of a data standards program, the appointment of data standards managers for each house to manage the program, and the establishment of a coordinating policy committee and a technical committee to provide ongoing oversight and support.

The report also recommended that several standards be approved and authorized for use as warranted by the requirements of specific documents and the capabilities of the various offices that create them. The report recommended Standard Generalized Markup Language (SGML) as the primary data standard because it is an international standard that is not controlled by any single vendor; it can be implemented independently of any specific hardware or software; it allows data to be tagged for content rather than format; and it can support a variety of output formats for printing, for creating CD ROMs, and for publishing on the Web. The report also recognized the need for other standards, and therefore recommended the use of Hypertext Markup Language (HTML), a subset of SGML especially designed for rapid display and linking of data on the Web; and Portable Document Format (PDF), widely used as a format for displaying data on-line as an image of the printed document.

The technologies that affect data standards will continue to develop. The report noted that ". . . standards will evolve over time as technology and the capacity of offices and agencies to adopt these technologies evolves."[6] The report anticipated that the data standards managers and the coordinating and technical committees

---

[6]Ibid, p. 7.

would have to assess new standards as they emerge and make recommendations for their use by the legislative branch as appropriate.

D. Senate Legislative Information System (LIS)

The LIS is a mandated system (Section 8 of the 1997 Legislative Appropriations Act, 2 U.S.C. 123e) to provide a "comprehensive Senate Legislative Information System" to capture, store, manage, and distribute Senate documents. A Year 2000 compliant LIS Document Management System (LIS/DMS) was successfully deployed in December 1999. The Advisory Committee closely monitors this project to ensure that the information is in formats that can be transferred to the National Archives for preservation and that the record copies (i.e. authoritative version) of legislative documents are identified and maintained permanently.

In 1998, the Secretary of the Senate created a task force to address two issues in relation to the initial development of LIS: archival preservation of the electronic documents in the LIS system and designation of the official "record copy" of legislative documents. The task force included staff of the Historical Office, the LIS project team, and staff from the National Archives. The task force concluded that:

- textual documents identified as "record copy" in the *Records Disposition Procedures for Offices of the Secretary of the Senate* constitute the official copy for purposes of documentation;
- LIS system data should be preserved in the National Archives because of its research usefulness in the electronic format;
- LIS System Requirements did meet current National Archives' standards as specified in 36 CFR 1228.188 for preservation of electronic files. It was noted that bills and other legislative information in LIS will be created in SGML, as specified in 36 CFR 1228.188.

The task force recommended that:
- the National Archives participate in a data standards project led by the Secretary of the Senate and the Clerk of the House;
- if regulations defining acceptable formats for archiving digital audio and video are established, additional coordination with the National Archives should be sought;
- as LIS is developed, any documents that are unique to LIS should be identified and designated as "record copy";
- because the Senate, House, and other legislative agencies are the de facto active repositories for research on historical information until electronic storage requirements exceed the capac-

18

ity of LIS to maintain this data, the timing and content of transfers of LIS data to the National Archives needs to be determined.

In 1998, the LIS project staff analyzed and reviewed systems requirements, related projects (e.g., LOC LIS Retrieval System), and initiatives at the Senate and other agencies, and gathered information integral to the implementation of the LIS. The Committee Scheduling application, developed and deployed during the year, replaced the older system. This system enables the Daily Digest Office to schedule committee and subcommittee meetings and allows all Senate users to retrieve information about committee meetings and hearings via a convenient web-browser. The Amendment Tracking System (ATS), also deployed in 1998, enables staff to scan floor amendments as they are received at the Bill Clerk's desk. Within 20 minutes, Senators and staff can view the text of the amendment from their computers.

During 1999, staff focused on the development of LIS/DMS and its interfaces to other legislative systems. The system was successfully deployed in December 1999.

For 2001, the LIS staff will develop enhancements to the LIS/DMS. The staff also will study the LIS Senate Recording Studio transcription and closed captioning project, LIS retrieval enhancements, and the retention, distribution, and archival policies and procedures project. The Recording Studio transcription and closed captioning project involves establishing the mechanism that must be put into place to make Senate Recording Studio data available within the LIS system. The retention, distribution, and archive policy project implements the capture and archiving of historical information collected and made available through the LIS.

E. House Document Management System

In 1996 the Clerk of the House presented to the Committee on House Oversight a plan to create a House Document Management System. This information management initiative proposed an enterprise-wide approach to the creation, distribution and maintenance of legislative information that endeavored to make enormous improvements in the cost, accuracy, timeliness and efficiency of the process. Paramount to this effort is the establishment of common data standards for the exchange of legislative information. To that end, the Clerk invested heavily in and led the development of the document type definitions (DTDs) which are necessary to provide the framework for a document using common data standards. This project is proceeding on a cooperative basis between the House,

19

Senate, Government Printing Office (GPO), Library of Congress and the Congressional Research Service. DTDs have been developed for a number of legislative documents including Bills, Resolutions, Amendments and Conference Reports. Additionally, Document Analysis workshops for the U.S. Code, Committee Reports, and Compilations have been completed.

With the progress made in DTD development and following the House Systems Development Life-Cycle Policy, the Clerk determined that a feasibility study was needed to provide an analysis of specific objectives, requirements, and system concepts as an evaluation of alternative approaches. In 1999, the Clerk was directed by the Committee on House Administration to initiate an SGML/XML feasibility study including staff from the House, Senate, Government Printing Office, Library of Congress, and Congressional Research Service.

This feasibility study was designed to provide specific information on:

- Customization and evaluation of several SGML/XML editors using a subset of bills and/or resolutions (without tables or graphics).
- Creation of XML style sheets for delivery to customers using XML-aware browsers in addition to Portable Document Format (PDF) delivery.
- Evaluation of SGML/XML capabilities of the Government Printing Office supported *Microcomp* formatting system to produce hard copy documents.
- Use of digital signature or other appropriate technology to allow users to determine if the subject documents have been inappropriately altered or tampered with.

F. SGML/XML Feasibility Study

The Secretary of the Senate, with concurrence from the Senate Rules and Administration Committee, joined the Clerk of the House in establishing the Bills, Resolutions, and Amendments Feasibility Study.

The primary purpose of the XML Feasibility Study was to determine whether or not XML and available XML tools provide a viable option for drafting legislation. In addition, the Study was designed to provide a collaborative development environment for the House and Senate to explore XML encoding as a data standard for exchange of data within the legislative branch. The Feasibility Study also enabled a more thorough validation of the Bills DTD. In both regards, the Feasibility Study provided some clear successes,

although it also demonstrated some potential challenges and opportunities.

One main challenge is that legislative drafters have been working in a DOS-based application for many years with keyboard shortcuts that could be viewed as akin to a combination of stenography and typesetting functions. Both the migration to a Windows environment and application, and the migration to a rules-based editing environment (DTD) offer challenges that also provide an opportunity to enable legislative drafters to concentrate more on legislative drafting than on format and display of the end product. The House Office of the Legislative Counsel and the Office of the Clerk are encouraged by the progress to date and are enthusiastic about the possibilities these new tools will provide. Additionally, the GPO and the LOC have expressed their support for this approach and their study participants are additionally encouraged by the accomplishments of the Feasibility Study.

Although the Feasibility Study did not produce a product that currently enables legislative drafters to produce legislation, it provided insights about the state of the technology and technology's applicability to the legislative drafting process. Additionally, it confirmed that while there are fundamental business process differences between the House and Senate, XML could be relied upon as a common data standard although possibly implemented differently by the two chambers. While the Senate Office of the Legislative Counsel recognizes the potential for XML to enhance the drafting process in certain ways, the office was reluctant to fully endorse the use of the new technology until more progress is achieved for certain drafting and editing functions. On the other hand, although currently available tools are still somewhat immature and still emerging, the Clerk, the Secretary, and HOLC concluded that we should continue to expand efforts to use XML along with associated tools for the creation, editing, and exchange of legislative documents.

In November 2000, the Committee on House Administration, acting on the Clerk's recommendations to the committee following completion of the feasibility study, authorized the adoption of XML as a data standard for the exchange of House legislative documents and authorized the Clerk to develop a program plan for the House document management system initiative using XML as the data standard. The Committee further stipulated that the House deploy resources to proceed with appropriate XML conversions and software customization, together with other projects to foster development and customization of editing environments and expand development of document type definitions for additional legislative products.

21

Concurrently, the Senate Committee on Rules and Administration together with the Committee on House Administration agreed with the joint recommendation of the Secretary of the Senate and the Clerk of the House that XML should be the primary data standard employed for the exchange of legislative documents and information among the House, Senate and other legislative branch agencies. The committees recognized that the implementation and transition to XML will take several years and will require coordination among all legislative agencies, with the House and Senate continuing their efforts to refine the technical editing tools appropriate to their specific needs. Furthermore, the committees called for joint piloting of programs for the actual exchange of legislative information as each institution completes its own XML evaluations. Development of XML within Congress will facilitate preservation of Congress' electronic records. (See section II.A. for discussion of the National Archives' ERA program.)