

# Preserving Electronic Records: A History

## Four Decades of Preserving Electronic Records at NARA

By Vivek Navale and Ross Cameron

In late 1968, the Archivist of the United States of America formally established the Data Archives Staff, and the first electronic records arrived in April of 1970. In the early days these "machine-readable" records were main-frame data bases stored on open-reel magnetic computer tapes and occasionally punch-cards. Preservation was performed using punch-card programs to read the agency tape or punch cards and to write a copy of the records to a new tape using leased time on commercial computers. These records were subsequently stored at the Washington National Records Center (WNRC). During this period of time staff also began examining and evaluating new tape products that could be used for archival preservation work, to determine if they met established industry standards.

By the 1980s, electronic records preservation was performed by submitting copy jobs via computer terminals that were interfaced to federal government computer centers, producing two new copies of the records, and then disposing of the original agency tape. Earlier research showed that magnetic tape is a fragile medium with a limited life-span. Therefore to minimize the data loss, two holdings maintenance practices were initiated. First all tapes were replaced when ten-years old by copying the records onto new evaluated tapes. Also, a random statistical sample of tapes was drawn and tested annually to check for any readability problems developing before the tapes were ten-years old. During the mid-1980s the tapes were physically relocated from WNRC to vaults in the National Archives Building in D.C. After a few years they were moved to leased space at Pickett Street in Alexandria, VA.

During the 1990s, the Center for Electronic Records (CER) developed and deployed the Archival Preservation System (APS), to process electronic records received from agencies, enabling preservation on archival magnetic media. CER began archiving electronic records on 3480-class cartridges which were also evaluated before use for preservation work. CD-ROMs and diskettes were accepted from agencies as transfer media of electronic records to NARA.

In 1993, CER received 5,906 backup media containing records of the Executive Office of the President (EOP records from the Ronald Reagan and George H. W. Bush administrations) via a Court order resulting from the *Armstrong, et. al. vs. EOP, et. al.* lawsuit. These backup media were duplicated onto new media starting that year. The new media were ten-year Replacement copied starting in 2003. The Court also ordered an Annual Sample and additional Plaintiff's Inspection of the original media and new media, consisting of a total of 1,039 media each year, which continues to be performed annually.

Over 60,000 backup media from the William J. Clinton administration were received in 2001 and some of these are also sampled. In 2009, over 80,000 media from the George W. Bush administration were received.

To contend with the increasing volume of electronic records from agencies and changing technologies, CER (renamed as Electronic Records and Special Media Division, NWME) during the years of 2000-2002, conducted a scientific study of high density magnetic media (tapes) stability and life expectancies. Based on these studies, NARA started electronic record preservation work on Digital Linear Tape (DLT IV). In 2003, NWME led the development and deployment of the first indigenous DLT evaluator that tests for archival quality of the DLTs (currently still in use) prior to usage of commercially produced tapes for preservation work. Also, frequently asked questions on life expectancies of magnetic and optical media were [addressed](#).

During the same period NWME also implemented a program whereby backup copies of the unclassified tapes were moved to newly renovated space offsite. The classified Title 13 backups were moved there after a cage was constructed to separate them from the other records.

APS continues to be used today for processing, preserving and providing access to NARA e-holdings to customers. Since the early days, the APS system (hardware, software, networks and peripheral devices, databases) has been significantly enhanced with the deployment of the APS tape farm and improvements to software functionality to keep pace with continuous technological changes and advancements.

During this period NARA also began accepting a wide variety of electronic media types and record formats from agencies. In September 2008, electronic records were ingested into the Electronic Records Archive (ERA). Initially these were "legacy" records which NARA had preserved using the APS system. Shortly thereafter new accessions which were compatible with ERA were directly ingested into ERA. APS continues to be utilized to prepare legacy records for ingest into ERA and to initially process and preserve those records which cannot be handled by ERA yet, such as classified, Title 13, donated materials, web records, etc.

The 2011 NARA reorganization resulted in renaming NWME Technical Services as Electronic Records Preservation (RXE). Over the past four decades, the RXE staff preserved over 20 Terabytes of data consisting of over 10 million files. These records are from 200 Record Groups and have a date range from 1819 to the present. A variety of record types and formats are present in the NARA holdings and include data files in EBCDIC (standard and variants), ASCII, TXT, CSV, DBF, DAT, MS Access, MS Excel, Lotus 1-2-3, multi-punch, NIPS, SAS and SPSS; maps and charts in various Shape formats and VSD; moving images in MPEG, CIN, TIFF, WMV, Real Player and Macromedia Flash formats; sound recordings in AVI, WAV and MP3; photographs and other graphic images in TIF, JPEG, PNG, BMP, GIF, PS, GZ, PNM, PPT, PPS, MacPaint, AI, PSD, ICO and early JPL format; textual records in RTF, MS Word, PDF, DOT, WordPerfect, CSS and TXT; email in MSG, Microsoft Exchange/Outlook, Entourage and Text Mail; and web pages in HTML, PHP, CDX, GZ and XML.

The RXE staff will continue to migrate and ingest legacy unclassified electronic records from 9,000 magnetic tapes to network attached disc storage (NAS) for preparation, packaging and ingest into ERA. The migration of legacy e-records will be extended by deployment of Classified ERA (CERA) and Title 13 ERA systems. In addition, staff members contribute to ERA system

improvement, testing and requirements analysis, business process adoption and federal agency wide user training of ERA.

Ten-year Replacement migration to stable media of all records that have not been ingested into and approved in ERA is continued as part of the established preservation procedures, including the annual statistical sampling of holdings. APS is being sustained to handle new media types (e.g. Linear Tape Open cartridges) for processing electronic records received from federal agencies.

RXE staff will continue to provide access to researcher requests, including making reproduction copies of records for researchers on the media type selected by the researcher.