



Strategy for Preserving Digital Archival Materials

June 8, 2017

Introduction

The National Archives and Records Administration (NARA) identifies, preserves, and provides access to the U.S. Government's vast holdings of archival materials. We preserve these records to protect Citizens' rights, ensure government accountability and document the national experience. Our archival holdings include more than 13 billion pages of unique documents plus electronic material, maps, charts, aerial and still photographs, artifacts, as well as motion picture, sound, and video recordings. The records we hold belong to the public and our mission is to drive openness, cultivate public participation, and strengthen our nation's democracy through public access to high-value government records. Preserving NARA's digital holdings, including public use copies and the digital surrogates created through our digitization efforts, is integral to achieving these goals and our continued success.

Scope

NARA is committed to preserving and maintaining access to the content of all of the born-digital records and digital surrogates in our holdings that are determined by the Archivist to have sufficient historical or other value to warrant continued preservation by the United States Government per 44 U.S.C. §§ 2107 and 2203(g). By access we mean the continued, ongoing usability of records and their content, retaining qualities of authenticity, accuracy, and functionality deemed to be essential and feasible for the purposes the digital materials were created.

Strategies

NARA will employ several key strategies to enable the effective preservation of our digital content, recognizing that our strategies have to be flexible to adapt to ongoing changes in scale, technology, and standards. The goal is to reduce risk and achieve best practices to preserve and maintain access to our digital content.

- (1) *Documentation of Standards and Procedures.* NARA documents our internal standards for the creation of digital surrogates, provides guidance on agency creation of digital surrogates as per 44 USC 3302(3), and provides guidance on minimum metadata and preferred file formats for electronic records to be transferred to NARA; promotes the use of open standards-based formats and accepted voluntary, community-based standards to help facilitate future access

and preservation; and provides guidance to Federal agencies for the management of Federal records and transfer to NARA to support a digital records preservation lifecycle.

- (2) *Prioritization.* NARA will take a risk-based approach to setting digital preservation priorities and perform digital preservation activities on a schedule created by reviewing appropriate priorities for action. Regular assessments of the formats in our holdings will alert us to at-risk formats for which we do not yet have practical preservation strategies or where the necessary actions are technically complex.
- (3) *File Management.* NARA will store our digital content in our trusted Digital Object Repository and provide ongoing management and access to the content throughout its lifecycle. NARA's repository will be based on the concepts embodied in the Reference Model for Open Archival Information Systems (OAIS), ISO 14721:2012 for Trusted Digital Repositories. A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future (OCLC. Trusted Digital Repositories: Attributes and Responsibilities, 2002). NARA will minimize the number of file formats that must be actively managed by aggressively normalizing files into selected formats that retain the significant characteristics of the original format, while retaining the original format files in low-access storage.
- (4) *Authenticity.* Authenticity refers to the trustworthiness of the record as an accurate representation of the original. NARA will ensure authenticity by documenting all digital preservation actions as per OAIS, ISO 14721:2012.
- (5) *Preservation Metadata.* NARA will assign persistent digital identifiers and record preservation metadata about each digital object, data stored as computer files and requiring applications software for viewing, to aid in the preservation of our digital holdings over time through manual and automated preservation processes. Preservation metadata ensures that essential contextual, administrative, descriptive, and technical information are preserved along with the digital object.
- (6) *Organizational Relationships.* NARA will actively engage with the local, national, and international digital preservation communities to share information and experiences, seek guidance, and collaborate to address digital preservation challenges. This engagement will help NARA identify emerging risks, practices, and standards to continually improve our program. We will engage the Information Technology (IT) industry to ensure it has an understanding of the needs of digital preservation as it develops new technical tools and systems.

Digital Preservation Activities

NARA digital preservation activities will undergo ongoing assessment using appropriate voluntary, community-based assessment instruments (e.g., TRAC or DRAMBORA, based on

Trusted Repositories Audit & Certification [ISO 16363:2012], or the National Digital Stewardship Alliance Preservation Levels) that measure program capabilities and maturity. Digital preservation will be achieved through a digital preservation infrastructure that: ensures data integrity, format and media sustainability, and information security.

- (1) *Infrastructure.* NARA's digital preservation infrastructure (hardware, software, networks, storage, related equipment, and facilities used to develop, test, operate, monitor, manage and/or support information technology services) includes:
 - (a) Storage, network capacity, systems, and tools for the ingest or creation, processing, active file management, and preservation of NARA born-digital files and digital surrogates.
 - (b) Processes to regularly review and update systems and tools that may be developed or procured by NARA to meet business needs.
 - (c) Affordable, managed, replicated content storage infrastructure for born-digital files and digital surrogates. Replication includes one preservation copy in a different storage environment, preferably in a remote geographic region, such as the replication that can be provided through NARA Cloud services.
 - (d) Tools to inventory all born-digital files and digital surrogates upon ingest.
 - (e) Tools for forensic identification and format characterization, which includes file format identification (identify the technical file types), format validation (confirm that the files meet documented format specifications), and technical metadata extraction (documenting how the files were created, including the applications and operating systems) which is used to support policy-based assessment of format obsolescence risks and to present the file to users using the appropriate application or viewer in context.
 - (f) Tools for file format transformations to perform file migrations over time as formats become obsolete and at-risk.
 - (g) Standardized workflow processes for associating born-digital and digital surrogate files with record identifiers and metadata and ensuring that files are in appropriate preservation storage and access server locations (on premise or in the cloud).
- (2) *Data Integrity.* NARA will:
 - (a) Inventory all incoming files and log the results of all ingest events, as well as all later lifecycle events such as format transformations, file movement, and audits.
 - (b) Ingest files, a process which must include malware scanning and the

checking of file fixity. File fixity checking refers to the validation that a file has not been altered from a previous state.

- (c) Copy content off physical media, incorporating the use of write-blockers, devices that prevent accidental damage to the content on the physical media, as appropriate.
 - (d) Perform an annual sample audit of all born-digital electronic record files and digital surrogates stored in the preservation repository, including fixity checks.
 - (e) Repair and/or replace files with fixity issues.
 - (f) Perform quarterly audits of logs in order to validate that files in the preservation repository have remained unchanged and uncorrupted over time.
 - (g) Perform an annual sample audit of media containing permanent records that are retained in NARA legal custody (36 CFR 1236.28 (e)).
 - (h) Before media containing permanent records are 10 years old, recopy onto tested and verified new electronic media (36 CFR 1236.28(f)).
- (3) *Format and Media Sustainability.* NARA will:
- (a) Characterize and validate file formats at the point of ingest. Characterization refers to the identification and description of a file's technical characteristics like its production environment. It is usually captured by technical metadata. Validation refers to confirming that the file in hand conforms to the expected characteristics of its type.
 - (b) Create File Format Action Plans that identify file formats and the actions required if those formats are no longer sustainable. e.g., are no longer created by or accessible through current software.
 - (c) Create normalized versions of files that are in an at-risk format as defined in File Format Action Plans. Normalization refers to converting all files of a particular type (e.g. emails, color images, etc.) to a chosen file format that will be sustainable.
 - (d) Analyze file formats and media formats that are received and determine potential obsolescence on an ongoing basis.
 - (e) Perform automated and manual format migrations or other preservation activities based upon File Format Action Plans.
 - (e) Monitor the larger preservation community and technological environment for signs that formats, media and equipment are becoming obsolete and

are no longer sustainable.

- (4) *Information Security.* NARA will:
 - (a) Identify and enforce who has:
 - (i) access to the physical media items;
 - (ii) access to ingest and processing systems and services; and
 - (iii) read, write, and execute authorization to folders and files on servers (on premise or in the cloud).
 - (b) Perform a scheduled review of individuals and groups who have read, write, and execute authorization to folders and files on servers.
 - (c) Ensure that no one person has write access to all files.
 - (d) Maintain a system of record logs of actions on files, including deletions and preservation actions.

Key Enabling Factors

There are many factors that will contribute to the ultimate success of this Digital Preservation Strategy. This section is intended to highlight the critical factors that must be addressed by NARA for its objectives to be met.

- (1) *Staffing Resources.* With this strategy, NARA is acknowledging that digital preservation is a significant business process that crosses multiple business units. NARA will develop a separate human resource plan to support this function.
- (2) *Information Technology Infrastructure.* NARA will require a planning process that identifies infrastructure needs to support digital preservation that includes systems and tools, storage, network capacity, data integrity, and information system security. This should document relevant operational and governance processes, including those for forecasting for storage and network capacity and planning for and implementing additional capacity and technology refreshes.
- (3) *Guidance on Standards to Records Creators.* NARA will continue to develop and promulgate guidance to agencies for technical, format, and metadata standards to ensure the sustainability of born-digital files and digital surrogates.
- (4) *Guidance and Policy for Digital Preservation.* NARA will promulgate further internal guidance and policy as business units begin implementing the strategy.