

Digital Preservation Strategy 2022-2026

Introduction

The National Archives and Records Administration (NARA) identifies, preserves, and provides access to the U.S. Government's vast holdings of archival records. We preserve these records to protect citizens' rights, ensure government accountability, and document the national experience. The electronic records in NARA's custody include textual materials, email, data files, maps, aerial and still photographs, as well as motion picture, sound, and video recordings. These records 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 born-digital records (those originally created in an electronic format) and digital surrogates—is foundational to achieving these goals. Preserving and ensuring future access to these records also directly supports the policies defined in E.O. 13985, *Advancing Racial Equity and Support for Underserved Communities Through the Federal Government*.

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. In this strategy, access refers to the continued, ongoing usability of records and their content, retaining qualities of authenticity, accuracy, and functionality deemed to be essential to maintain and identify the purposes for which the records were created.

NARA's FY 2022-2026 Strategic Plan identifies digital preservation as a strategic objective for the agency. Goal 3, Maximize NARA's Value to the Nation, states in part "NARA will advance existing physical and intellectual controls for the agency's holdings to enable digital preservation risk planning and risk mitigation in a trustworthy repository, and ongoing access to electronic records."

Strategies

NARA employs several key strategies to enable the effective preservation of our digital holdings, recognizing that our strategies have to be flexible to adapt to ongoing changes in scale, technology, and standards. The goals are to reduce the risk of loss and to implement international best practices and standards to preserve and maintain access to our digital content.

- 1) *Documentation of Standards and Procedures.* NARA documents our internal procedures and standards as appropriate for the lifecycle management of born-digital records, digital surrogates, and public use copies. NARA provides guidance on agency creation of digital surrogates as per 44 USC 3302(3); provides guidance on minimum metadata and preferred file formats for electronic records to be transferred to NARA (Bulletin 2015-04); promotes the use of open standards-based formats and widely-accepted community-based standards to help facilitate preservation and support future access; and provides guidance to Federal agencies for the management of Federal records and transfer to NARA to support a digital records preservation lifecycle.
- 2) *Digital Preservation Program.* NARA's program includes consulting across the agency on digital preservation topics and infrastructure needs; managing documentation of holdings-related procedures across the agency and lifecycle; program self-assessment; the processes and infrastructure to analyze the holdings; identify and manage risks; develop preservation action plans; and publish those plans internally and externally.
- 3) *Prioritization.* NARA takes a risk-based approach for setting digital preservation priorities to perform digital preservation actions. Regular assessments of the formats in our holdings alert us to at-risk formats for which we do not yet have practical preservation strategies or where the necessary actions are technically complex.
- 4) *File Management.* NARA stores our digital content in our trusted Digital Object Repository and provides ongoing management and access to the content throughout its lifecycle. NARA's repository is 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 as per an organization's published Designated Community statement, now and in the future ISO 16363:2012 (Audit and certification of trustworthy digital repositories). NARA's Digital Preservation Designated Community statement is available on its website. NARA minimizes the number of file formats that must be actively managed by transforming files into selected formats that retain the significant properties of the original format, while retaining the original format files in low-access storage.
- 5) *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 ISO 16363:2012.
- 6) *Preservation Metadata.* Preservation metadata ensures that essential contextual, administrative, descriptive, and technical information are preserved along with the record. NARA assigns persistent digital identifiers and records preservation metadata about each record to aid in the preservation of our digital holdings over time.
- 7) *Organizational Relationships.* NARA actively engages with the national and international digital preservation communities to share information and experiences, seek and provide guidance, and collaborate to address digital preservation challenges. This engagement helps NARA identify emerging risks, practices, and standards to continually improve our program. We engage the Information Technology (IT) industry to ensure it

has an understanding of digital preservation needs as the industry develops new technical tools and systems.

- 8) *Staff Training.* NARA ensures staff throughout the agency are provided with appropriate digital preservation training based on staff roles. This is accomplished through a variety of internal and external training modules, which are updated on an ongoing basis.

Digital Preservation Activities

The NARA digital preservation program undertakes ongoing assessment using appropriate community-based assessment instruments that measure program capabilities and maturity (e.g., ISO 16363:2012, or the National Digital Stewardship Alliance Levels of Digital Preservation).

Digital preservation will be achieved through a comprehensive approach that ensures data integrity, format and media sustainability, and information security.

- 1) *Infrastructure.* NARA's digital preservation infrastructure includes:
 - a) Tools for the analysis of the holdings, to identify and manage risks, develop preservation action plans, and publish those plans internally and externally.
 - b) Storage, network capacity, systems, and tools for the ingest, processing, rendering, active file management, preservation, and export between systems of born-digital files and digital surrogates.
 - c) Processes to regularly review and update systems and tools that may be developed or procured by NARA to meet business needs.
 - d) Affordable, managed, replicated content storage infrastructure for born-digital files and digital surrogates. Replication includes one preservation copy in one or more different storage environments, in remote geographic regions, such as the replication that is provided through NARA Cloud services.
 - e) Tools to inventory all born-digital files and digital surrogates upon ingest.
 - f) 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.
 - g) Tools to perform file format preservation transformations over time as formats become obsolete and at increased risk for long-term access. Transformation refers to converting all files of a particular format, or version of a format, to a chosen file format.

- h) 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-premises or in the cloud).
- 2) *Data Integrity*. NARA has a data integrity program to:
- a) Inventory all incoming files in the ERA Digital Object Repositories, log the results of ingest events, and, where possible, later lifecycle events such as format transformations, file movement, and audits.
 - b) Generate fixity information for files in the ERA Digital Object Repositories that are not accompanied by fixity information upon transfer. Fixity refers to a checksum or a “hash” value, an algorithmically-computed numeric value for a file or a set of files used to validate the state and content of the file for the purpose of detecting accidental errors that may have been introduced during its transmission or storage.
 - c) 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.
 - d) Copy content off incoming physical media, incorporating the use of write-blockers, devices that prevent accidental damage to the content on the physical media, as appropriate.
 - e) Perform an annual sample audit of the fixity information for all born-digital files and digital surrogates stored in the ERA Digital Object Repositories in order to validate that files in the repository have remained unchanged and uncorrupted over time.
 - f) Repair and/or replace files with fixity issues.
 - g) Perform an annual sample audit of media containing permanent records that are retained in NARA legal custody (36 CFR 1236.28(e)).
 - h) Recopy any records still stored on media onto tested and verified new electronic media (36 CFR 1236.28(f)) before the media containing permanent records is 10 years old.
- 3) *Format and Media Sustainability*. NARA assesses and acts on risks in several ways to:
- a) Characterize files where possible during processing/ingest into the ERA Digital Object Repositories. Characterization refers to the identification and description of a file’s technical characteristics like its production environment.
 - b) On an ongoing basis, create preservation action plans that identify file formats in NARA’s holdings 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 transformed versions of files that are in at-risk formats as defined in preservation action plans.

- d) On an ongoing basis, analyze file formats and media formats that are received and determine potential obsolescence.
 - e) Migrate holdings onto new preservation storage media over time to mitigate media obsolescence risks.
 - f) 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 is responsible for ongoing security assessments of the ERA Systems Digital Repositories to:
- 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-premises 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 copies of 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 continually be addressed by NARA for its objectives to be met.

- 1) *Organizational Support*. NARA identifies digital preservation as an agency-wide strategic goal, invests in adequate staffing, and ensures that the appropriate infrastructure for digital preservation functions is in place and maintained.
- 2) *Staffing Resources*. With this strategy, NARA acknowledges that digital preservation is a significant business process that crosses multiple business units. NARA will continue to assess its digital preservation staffing and training needs as the program matures.
- 3) *Information Technology Infrastructure*. NARA requires a planning process that identifies infrastructure needs to support digital preservation that includes risk analysis and planning, 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.
- 4) *Guidance on Standards for NARA Staff and Agency Records Creators*. NARA will continue to develop and share guidance with NARA staff and federal agencies for

technical, format, and metadata standards to ensure the sustainability of born-digital files and digital surrogates.

- 5) *Guidance and Policy for Digital Preservation*. NARA will continue to develop and share further internal guidance and policy as technology, best practices, and standards evolve.

Review Process and Version History

The Digital Preservation Strategy will be reviewed and updated on the same schedule as the NARA Strategic Plan. This strategy is owned by the Digital Preservation unit in the Office of the Deputy Archivist of the United States.

The 2022-2026 Strategy is a revision of the initial 2017 Strategy. Changes made are as follows:

- Edited for clarity
- Add reference to strategic goal and align with the time period and strategies in the Scope section
- Add reference to the 2021 NARA Digital Preservation Designated Community Statement
- Add reference to current Transfer Metadata Guidance
- Add a digital preservation program strategy and accompanying references to the program infrastructure section, e.g., tools for holdings analysis, risk analysis, and creating and publishing plans
- Updated the Data Integrity section with additional references to fixity checking tasks
- Add a staff training strategy and updated the enabling factor for staffing
- Add an enabling factor for organizational support