



# Introduction to Digital Preservation at NARA

Leslie Johnston, Director of Digital Preservation  
U.S. National Archives and Records Administration (NARA)

# Two Preservation Units at NARA

- **Preservation**, with the mandate to care for the physical holdings and perform preservation reformatting digitization of physical items, and to undertake preservation research.
- **Digital Preservation**, with the mandate to support electronic records processing archivists, perform audits of the holdings, and assess the need to perform preservation actions.

# What is Digital Preservation?

- Digital preservation covers the resources, technology, metadata, and the organizational support required to assess risk and preserve and maintain access to digital materials.
- The actions taken under this practice help to maintain access beyond the limits of media failure or technology obsolescence.
- Any digital object is in scope for digital preservation, whether born-digital or digitized.

# Mitigating Digital Preservation Risks

- **Safeguarding the data:** Preservation Planning, Active Object Management, and Bit-level Preservation.
- **Ensuring that the content stays accessible and usable:** File Format Migration.
- **Maintaining the context of the data:** Descriptive, Structural, Administrative, and Preservation Metadata.
- **Ensuring there is ongoing trust in the data:** Auditing and IT Security.
- The keys to all of this are good planning, thorough documentation, capturing and creating as much metadata as is feasible, and ongoing monitoring of the objects and the infrastructure they are stored in.

# Guiding Assumptions

- Electronic record files should conform to the NARA Transfer and Metadata Guidance whenever possible.
- Changes to files must be detectable and repairable. This is accomplished through replicated storage and recorded fixity information. Fixity refers to a file being unchanged.
- Actions taken on files must be recorded and tracked.
- Separate public use copies of files are created.
- At this time, ongoing preservation format transformations are not performed but are planned. Format transformations DO happen when formats are received that we cannot process/preserve.
- Regular audits must be performed.

# Guidance to Agencies are the Beginning of the Digital Preservation Lifecycle

- M-23-7: OMB/NARA Memorandum on the Transition to Electronic Records
- 2015-04: Metadata Guidance for the Transfer of Permanent Electronic Records
- 2015-02: Guidance on Managing Electronic Messages
- 2014-06: Guidance on Managing Email
- 2014-02: Guidance on Managing Social Media Records
- 2010-05: Guidance on Managing Records in Cloud Computing Environments

<https://www.archives.gov/records-mgmt/bulletins>



# Formats for Records Transfers

- An integral part of NARA's digital preservation work is the issuance of guidance on all aspects of Federal electronic records management and transfer to NARA, including media types, file formats, and metadata.

<https://www.archives.gov/records-mgmt/policy/transfer-guidance.html>

- NARA cannot be 100% proscriptive in the formats it accepts. When records are transferred, they are validated to ensure that they are uncorrupted, and, if possible, meet NARA's format guidance. There are "Preferred" and "Acceptable" formats, and NARA negotiates with each agency about what it can provide.

# Digital Preservation Strategy

- NARA has a Digital Preservation Strategy (first published in 2017, updated in 2022) to guide its internal operations.

<https://www.archives.gov/preservation/digital-preservation/strategy>

- This outlines the specific strategies that NARA will use in its digital preservation efforts, and specifically addresses:
  - Infrastructure
  - Format & Media Sustainability and Standards
  - Data Integrity
  - Information Security
- It applies to born-digital agency electronic records, digitized records from agencies, and NARA digitization for access and preservation reformatting.



# Who is working on Digital Preservation?

- The Digital Preservation unit guides internal operations:
  - Advised by an agency-wide Digital Preservation Group
  - Responsible for a Holdings Profile across all custodial units
  - Responsible for the Digital Preservation Framework: the Risk Analysis Matrix, and File Format Preservation Action Plans
  - SOPs for digital preservation and management of files during digitization are submitted to the digital preservation unit
  - Creation of Digital Preservation training content
  - Transparency and replicability are key: processes and documentation are made available to all NARA staff
- This is distributed work across the agency, not just the digital preservation unit:
  - Agency Services for external guidance, creation of Schedules, and records appraisal
  - Processing archivists in the Legislative Archives, Presidential Libraries, and Research Services for Federal records
  - IT Operations and IT Security for monitoring of systems and storage status, and regular emergency system backup restoration tests
  - The Innovation unit provides public access through the National Archives Catalog
  - Agency leadership supports the work through policies and resources

# Digital Preservation Framework: Risk Matrix

- NARA has an extensive Risk Matrix, designed to apply a series of weighted factors related to the preservation sustainability of the file formats in the Collection Format Profile to generate a numeric score.
- Each question has a relative weighting that maps to the level of risk for each question and, to the extent that it can be defined, resource costs (staff time or budget).
- The Matrix also includes high level factors that assess the preservation actions that could be taken vis-à-vis our current environment and capabilities.
- The Matrix calculates numeric scores, which are mapped to High, Moderate, and Low Risk. The risk thresholds are open to review and revision over time.

# Digital Preservation Framework: Preservation Plans for Record Types

NARA developed its Digital Preservation Framework to document and share recommended preservation actions based on its electronic record holdings and current capabilities. There are Preservation Plans for 16 categories of electronic records (or “record types”) which identify “Significant Properties,” the properties that should, if possible, be retained in any format migration:

- Calendars
- Databases
- Digital Audio
- Digital Cinema
- Digital Design/CAD
- Digital Still Image
- Digital Video
- Email
- Geospatial
- Navigational Charts
- Presentation & Publishing
- Software Code
- Spreadsheets
- Structured Data
- Web Records
- Word Processing

# Digital Preservation Framework: Preservation Plans for Formats

- NARA developed Preservation Action Plans for over 700 file formats based on an analysis using the Risk Matrix and the Significant Properties for each associated record type. The Format Plans include links to standards and specifications and proposed preservation actions and tools based on current NARA thinking and capabilities.
- The Digital Preservation Framework Matrix and Preservation Plans are publicly available on the NARA Github account for reuse and adaptation, as well as discussion.

<https://github.com/usnationalarchives/digital-preservation>

- The Framework can be applied across the lifecycle: its documented format sustainability metrics provide critical context for records creator and management decisions, support records selection and appraisal, and guide the selection of formats available for public access.

# Core Digital Preservation Infrastructure

- ***ERA Business Objects***, to initiate and record approvals for federal agency electronic records scheduling and transfers
- ***Processing Tools***:
  - ***AISS***, which runs the automated workflow to read media received from federal agencies to prepare for ingest.
  - ***AMIS***, to log ingest and processing workflows for federal records.
  - ***AERIC***, to validate structured data received as federal records.
  - ***Commercial and Open-Source Software***, to automate processing, review files, and transform formats.
- ***Preservation Systems and Storage***:
  - ***ERA 2.0***, The current cloud-based system to process and preserve federal records.
  - ***ERA EOP***, to preserve Presidential records.
  - ***ERA CRI***, to preserve Legislative records.
  - ***Census Staging Buckets***, to preserve Census records from 2000, 2010, and 2020
- ***DAS***, to create descriptive metadata for records
- ***The National Archives Catalog***, to deliver records to the public



# Updating the NARA Infrastructure

- ERA 2.0 went into production in October 2018 in the AWS GovCloud: it introduces a more flexible and extensible framework for processing and preservation.
- Updated infrastructure and tools to support more efficient transfer, processing, and delivery of born-digital and digitized records.
- Increased capacity for the transfer and processing of records, and the ability to support additional and every-changing file formats more readily.
- A Digital Processing Environment (DPE) system for the records transfer and ingest process that also incorporates virtual machines which run commercial and open source tools needed by archivists to process electronic records.
- A cloud-based Digital Object Repository preservation system.
- Support for ingest, processing, and preservation of digitized records created in-house and by partners.
- An updated Disposition Scheduling and Transfer administration workflow infrastructure went into production in 2023.



# How do we assess our progress?

- In 2019, 2021, and 2023, NARA completed internal self-assessments of its programs and systems using the PTAB (Primary Trustworthy Digital Repository Authorisation Body) instrument based on ISO 16363.
  - <http://www.iso16363.org/iso-certification/preparation/>
  - <https://public.ccsds.org/Pubs/652x0m1.pdf>
- We readily acknowledge that there are gaps in our processes, documentation, and systems. This is the necessary first step in a gap analysis to prioritize investment in the infrastructure, policy revisions, and the normalization and documentation of processes.

# NARA Self-Assessment Outcomes

In 2023 NARA meets 76% of metrics - up 28% from 2021 - and meets or partially meets 98%

**PROGRAM  
IMPROVEMENTS  
2019-2023**

	2019	2021	2023
Meets	32	52 (+20)	83 (+31)
Partially Meets	64	54 (-10)	24 (-30)
Does Not Meet	13	3 (-10)	2 (-1)

**For More Information, Feel Free  
to Contact Us**

Leslie Johnston

[leslie.johnston@nara.gov](mailto:leslie.johnston@nara.gov)