



Preserving and Providing Access to Digital Collections at the National Archives and Records Administration

Leslie Johnston
Director of Digital Preservation, NARA
National Academies
May 6, 2018

NARA Mission, Visions, Values

- Mission

- Our mission is to provide public access to Federal Government records in our custody and control. Public access to government records strengthens democracy by allowing Americans to claim their rights of citizenship, hold their government accountable, and understand their history so they can participate more effectively in their government.
- We drive openness, cultivate public participation, and strengthen our nation's democracy through public access to high-value government records.

- Vision

- We will be known for cutting-edge access to extraordinary volumes of government information and unprecedented engagement to bring greater meaning to the American experience.
- We collaborate with other Federal agencies, the private sector, and the public to offer information—including records, data, and context—when, where, and how it is needed and transform the American public's relationship with their government.
- We will lead the archival and information professions to ensure archives thrive in a digital world.

- Values

- Collaborate: Create an open, inclusive work environment that is built on respect, communication, integrity, and collaborative teamwork.
- Innovate: Encourage creativity and invest in innovation to build our future.
- Learn: Pursue excellence through continuous learning and become smarter all the time about what we know and what we do in service to others.
- Our Values reflect our shared aspirations that support and encourage our long-standing commitment to public service, openness and transparency, and the government records that we hold in trust.

Authorities for NARA Operations

- [The Federal Register \(44 U.S.C. Chapter 15\)](#)
- [National Archives and Records Administration \(44 U.S.C. Chapter 21\)](#)
- [Presidential Records Act \(44 U.S.C. Chapter 22\)](#)
- [NARA Trust Fund Board \(44 U.S.C. Chapter 23\)](#)
- [National Historical Publications and Records Commission \(44 U.S.C. Chapter 25\)](#)
- [The Records of Congress \(44 U.S.C. Chapter 27\)](#)
- [Records Management by the Archivist of the United States \(44 U.S.C. Chapter 29\)](#)
- [Records Management by Federal Agencies \(44 U.S.C. Chapter 31\)](#)
- [Disposal of Records \(44 U.S.C. Chapter 33\)](#)

Presidential Records Act and Federal Records Act Amendments of 2014

CONGRESS.GOV

Legislation | Congressional Record | Committees | Members | Browse About | Glossary

Current Legislation

[Home](#) > [Legislation](#) > [113th Congress](#) > H.R.1233 [Print](#) [Subscribe](#)

H.R.1233 - Presidential and Federal Records Act Amendments of 2014

113th Congress (2013-2014)

LAW

Sponsor: [Rep. Cummings, Elijah E. \[D-MD-7\]](#) (Introduced 03/18/2013)

Committees: House - Oversight and Government Reform | Senate - Homeland Security and Governmental Affairs

Committee Reports: [H. Rept. 113-127](#); [S. Rept. 113-218](#)

Latest Action: 11/26/2014 Became [Public Law No: 113-187](#).

Major Recorded Votes: 01/14/2014 : [Passed House](#)

Tracker:

Introduced > Passed House > Passed Senate > Resolving Differences > To President > **Became Law**

New Definition of Federal records

RECORDS

includes all **recorded information**, regardless of form or characteristics, made or received by a Federal agency under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the United States Government or because of the informational value of data in them

New Definition of Federal records

- Recorded Information
 - Includes all traditional forms of records, regardless of physical form or characteristics, including information created, manipulated, communicated, or stored in digital or electronic form.
- Does not include:
 - library and museum material made or acquired and preserved solely for reference or exhibition purposes.
 - duplicate copies of records preserved only for convenience.

NARA Provides Guidance for Records Transfers and Preservation

- 2018-01: Revised Format Guidance for the Transfer of Permanent Electronic Records
 - 2015-04: Metadata Guidance for the Transfer of Permanent Electronic Records
 - 2015-03: Guidance on Managing Digital Identity Authentication Records
 - 2015-02: Guidance on Managing Electronic Messages
 - 2014-06: Guidance on Managing Email
 - 2014-02: Guidance on Managing Social Media Records
 - 2013-03: Guidance for Agency Employees on the Management of Federal Records, Including Email Accounts, and the Protection of Federal Records from Unauthorized Removal
 - 2012-02: Guidance on Managing Content on Shared Drives
 - 2010-05: Guidance on Managing Records in Cloud Computing Environments
- <https://www.archives.gov/records-mgmt/bulletins>

Digital Preservation Strategy

- NARA published its first Digital Preservation Strategy in June 2017.

<https://www.archives.gov/preservation/electronic-records.html>

- 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.

What are in the NARA Holdings?

- Permanently valuable electronic records of the federal government.
- The first electronic record transfer was in 1970.
- There are over 200 agencies that are required to transfer records to NARA.
- The holdings comprise over 1.47 billion files with over 600 versions of file formats—textual, email, audio, video, still image, GIS, datasets, databases, system exports, web sites, social media etc.
- The majority of records are Unclassified, but NARA does have Classified records.
- NARA has several large-scale internal digitization projects, as well as mass digitization partnerships with external organizations that produces hundreds of thousands of files per year.
- After December 2022, NARA has proposed that it will no longer accept paper records from federal agencies.

What are the “Disruptors” in Federal Digital Preservation?

Scale

- Even with only 3% of federal materials identified as permanently valuable records, the size of records transfer are increasing. A single transfer can be several hundred million files or 200 Pb of files.
- There are large datasets and a massive amount of observational data that will potentially be identified as permanent records.
- Some types of collections – audio, video, film, email – produce both huge files and huge numbers of files to preserve.

Obsolescence

- Format Obsolescence
 - There are literally thousands of variant versions of file formats over time, and they just keep changing. The tools currently available cannot identify every legacy format with certainty.
- Media Obsolescence
 - There are dozens of carrier formats—floppy disks, hard drives, CDs, DVDs, thumb drives, tapes, etc.—and organizations such as NARA must to be able to read the files off them to preserve them.
 - The issues of working with vintage hardware and media are getting more attention in the larger community:
 - [The National Archives of Australia was interviewed about its practices.](#) And [video preservation continues to be a critical topic.](#) [Born digital media files only 4 years old can be at risk without stewardship.](#) [As is email.](#)
- NARA has recently developed a Risk Assessment framework for file formats, and a set of Preservation Actions Plans that identify risk mitigation options.

The Web as Record

- Web Archiving is not new, but there is greater public awareness of web content as valuable in policy research and in open government initiatives as the transitory nature of site sites becomes a more visible issue.
- While federal web content is sometimes under the purview of the Government Publishing Office and its preservation mission for government publications, in other cases web content comprises permanent federal records as the only copies are online and nowhere else.
- This could lead to a large scale re-appraisal of an entire category of records.

Context, Complexity, and Algorithmic Transparency

- There is new recognition that digital items have a context -- they're not always files sitting in directories on desktops and servers -- they live in systems and in complex web applications.
- Individual items are increasingly complex, comprised of multi-part or containerized files that require all their components.
- Our interactions are increasingly based on algorithms that control how data is presented to us, and in some cases, so personalized that those algorithms are actually mediating what information we encounter.
- Preservation of both the data and process/rules used to interpret the data is necessary for reproducibility.
- Again, this could lead to the identification and appraisal of new record types than have been seen in most federal record schedules.

Preservation of Code

- A corollary to algorithmic transparency is the preservation of code used in research, but also the code necessary to process or provide access to records.
- The [Library of Congress has a symposium on architectural design files and software preservation.](#)
- The [Software Preservation Network](#) and [Software Heritage](#) are testing the boundaries of public policy and legal framework to preserve code.
- Code is not a formal record type currently covered by NARA guidance, but could become so.

The Cloud

- Agency records are already in the Cloud—in hosted email and applications and storage, as well as contract hosted vendor systems that produce and store permanent records.
- NARA makes use of Amazon Web Services for application development, agency records transfer review, record processing, preservation, and public access.
- Taking advantage of several storage tiers as appropriate for the record access requirements, as well as replication within and between regions.
- NARA is designing processes to work with agencies where their records are, enabling Cloud-to-Cloud transfers, including across vendors.

Discussion

Leslie Johnston

Director of Digital Preservation

leslie.johnston@nara.gov