Developing Guidelines for Digitization of US Federal Government Records

Michael Horsley and Kevin L. DeVorsey United States National Archives and Records Administration, College Park, Maryland, USA

Abstract

The United States National Archives and Records Administration (NARA) has issued a new Strategic Plan FY 2018-FY 2022 that reflects NARA's ongoing commitment to support the transition to electronic record keeping, increased access and a fully digital government. The plan provides federal agencies with a deadline: "By December 31, 2022, NARA will, to the fullest extent possible, no longer accept transfers of permanent or temporary records in analog formats and will accept records only in electronic format and with appropriate metadata."

It is likely that Agencies will digitize large volumes of paper records to meet this goal. This paper presents NARA's Records Management Policy and Standard's Team's draft guidance on digitizing records. The paper discusses the evolution of NARA's 2004 digital imaging for access guidelines into the current 2016 FADGI guidelines, and discusses some of the unique issues applying these guidelines in a records management context to comply with NARA's Strategic Plan.

Introduction

The National Archives and Records Administration (NARA) is an independent agency established in 1934 to identify, protect, preserve, and make publicly available the historically valuable records of all three branches of the Federal government. NARA manages the Federal government's archives, administers a system of Presidential Libraries, operates museums, conducts education and public programs, provides oversight of government-wide records management activities, and provides temporary storage of other agencies' records on their behalf. NARA publishes the Federal Register and makes grants through the National Historical Publications and Records Commission. NARA provides for the appropriate declassification of classified national security information, mediating Freedom of Information Act disputes, and overseeing agency actions regarding classified and controlled, unclassified information.

The Federal Records Act requires that NARA issue regulations and provide guidance to agencies to ensure proper records management. Through a NARA reviewed scheduling and appraisal process agencies determine the value of their Federal records, how long temporary records must be retained before destruction, and which records are permanently valuable and must be transferred to NARA for preservation.

The Electronic Records Policy and Standards Team develops records management policies and guidance related to the creation, management, and transfer of electronic records and participates in the development of international records and document management standards. Key records management guidance maintained by the team are: "*2015-04: Metadata Guidance for the Transfer of Permanent Electronic Records*" [1], and "*2014-04: Revised Format Guidance for the Transfer of Permanent Electronic Records*" [2]. The team is currently updating the 2014

format transfer guidance, in addition to developing guidance on digitizing Federal records as a recordkeeping activity.

NARA's Strategic Plan

NARA's strategic plan [3] builds on the goals described in the "*Managing Government Records Directive (M-12-18)*" [4], and the "*Federal Records Act Revision of 2014*" [5]. Directive M-12-18 was issued in 2014 as part of the US Federal government's efforts to modernize records management and policies in order to transition to a digital government. The Directive aims to improve performance and promotion of openness and accountability by better documenting agency actions and decisions; further identify and transfer to NARA permanently valuable historical records; and assist agencies to minimize costs and operate more efficiently. The directive encourages agencies to digitize permanent records created in hard-copy format or other analog formats (e.g., microfiche, microfilm, analog video, analog audio)." The Federal Records Act Revision of 2014 included the requirement that the Archivist develop and promulgate digitization standards: "*3302 – The Archivist shall promulgate regulations, not inconsistent with this chapter, establishing - (3) – standards for the reproduction of records by photographic, microphotographic, or digital processes with a view to the disposal of the original records*".

Transition to electronic recordkeeping across the Federal government

NARA's strategic plan defines the mission, strategic goals, and the long-term objectives the agency hopes to accomplish over the next four years. Once finalized, it is to be published this year to cover FY 2018 through FY 2022. The plan is broken down into four Goals, two of which "*Goal 2 Connect with Customers*" and "*Goal 3 Maximize NARA's Value to the Nation*" are areas of particular interest for records management and digitization of Federal records. Specifically, NARA will by FY 2020, have policies and processes in place to support Federal agencies' transition to fully electronic recordkeeping, and by December 31, 2022, NARA will, to the fullest extent possible, no longer accept transfers of permanent or temporary records in analog formats and will accept records only in electronic format and with appropriate metadata.

NARA's Records Management Program is working with over 200 agencies that have different approaches, budgets, missions and business needs to make a major transition to fully electronic record keeping. To facilitate this transition our strategies for meeting Goal 2 include:

- Conduct an internal business process reengineering to improve the customer-facing functions of the Electronic Records Archive 2.0
- Provide Federal agencies with digitization guidance and "success criteria" for managing electronic records

- Develop Federal records management requirements and working with Federal and commercial vendors to incorporate the requirements into software applications and cloud offerings
- Developing two draft documents *Electronic Records Management Federal Integrated Business Framework,* and the *Use Cases for Electronic Messages* as part of our Federal Electronic Records Modernization Initiative (FERMI).

Goal 3 "*Maximize NARA's Value to the Nation*", contains the ambitious statement that with certain exceptions, NARA will only be accepting records in electronic format. Currently, NARA's Records Management Program has two key Bulletins related to the transfer of permanent electronic records to NARA in our existing body of Guidance for agencies: *File Format Guidance, Bulletin 2014-04* which specifies which file formats are acceptable when transferring permanent electronic records to NARA, and *Metadata Guidance Bulletin 2015-04* which provides the minimum list of metadata terms necessary for describing permanent electronic records. As part of our work to support the Strategic Plan and to meet increasing requests from federal agencies, we are developing a Digitization Guidance Bulletin that describes best practices and requirements.

Developing digitization guidance for Federal agencies

Creating digitization guidance based on recognized standards that can communicate both core record keeping requirements to agency records managers and highly technical requirements to digitization professionals is challenging. Most digitization guidance for cultural heritage and archival materials is written to satisfy either the relatively low quality requirements of access projects or the extremely high quality requirements for digitizing rare or intrinsically valuable materials. There is a dearth of guidance for reformatting paper and analogue records into electronic surrogates suitable for permanently replacing the originals.

To satisfy the goals of the strategic plan, NARA's guidance must identify technical and recordkeeping requirements for digitizing paper records that result in electronic surrogates that can be used for the same purposes as the originals. Since it is common for boxes of records to include a variety of media and for information to be recorded on envelopes, as captions on the backs of photographs, or the covers of bound volumes, the guidance must ensure that *all* original record material are addressed and that any gaps in coverage noted appropriately. While the guidance will be mandatory for records scheduled for permanent retention and transfer to NARA, agencies should adopt them when digitizing records relating to rights and interests and other temporary records to ensure that electronic copies are authentic, reliable, usable, and have the integrity necessary to allow for the disposal of the original records.

The guidance will address areas including:

- Record selection,
- Assessment and preparation,
- Technical, format, and metadata requirements,
- Quality management considerations, and
- Project documentation requirements.

The guidance will be issued as a cover bulletin that defines high-level requirements for digitizing original records as well as appendices that will address the specific requirements of different media types. Initially the guidance will address textual records printed on modern paper since these materials comprise the largest percentage of the records affected by the strategic plan. Additional media types will be addressed as necessary through the release of additional appendices. The guidance incorporates principles from many sources that describe file format and metadata selection, as well as technical standards.

All digitized images should be assigned administrative, descriptive, and rights metadata documenting the original records and their disposition schedule, as well as technical metadata describing the resulting digital objects and the digitization process. File formats must be identified as appropriate for use with scanned text in *NARA Bulletin 2014-04* and metadata should include elements identified in *NARA Bulletin 2015-04*, as well as additional elements from ANSI/NISO Z39.87-2006 (R2017), Data Dictionary – Technical Metadata for Digital Still Images.

The guidance recommends federal agencies follow the Federal Agency Digital Guideline Initiative's (FADGI) *Technical Guidelines for Digitizing Cultural Heritage Materials (2016)* [5] as the basis for digitization projects. [6].

FADGI Standards

The Bulletin incorporates principles from both the 2016 and 2010 FADGI guidelines. The current FADGI guidelines evolved from the NARA 2004 "Technical Guidelines for Digitizing Archival Materials for Electronic Access" [7]. NARA's 2004 guidelines represented internal best practice at the time and were intended to be informative to the digital imaging community at large. They were not intended to form the basis of official records management policy. The 2004 Guidelines were replaced by NARA's Digitization Services' products and services specifications [8]. The Products and Services specifications are consistent with the current digitization recommendations cited in the 2014-04 Transfer Guidance for federal agencies. The FADGI 2010 guidance was a product of efforts by digitization practitioners from several governmental cultural heritage institutions, and greatly expanded the technical basis for digital quality control and performance metrics. The 2016 FADGI Guidelines further expand upon the 2010 work by introducing the DICE (Digital Imaging Conformance Environment) program.

While these digitization best practices are widely adopted beyond Federal agencies they advise that they are intended to be "informative, and not...prescriptive" and avoid explicit language to define preservation reformatting specifications (See NARA 2004 Appendix A for a detailed analysis). The FADGI Still Image Working Group 2010 "Gap Analysis" provides an expanded discussion of the issues not covered in the Technical Guidelines. These documents were originally conceived to present best practices for creating digital surrogates to facilitate reproduction and access master files. NARA 2004 Guidelines, Products and Services, and FADGI Guidance do not explicitly define all the parameters affecting digitization of federal records in the context of records management policy.

However FADGI guidance has very specific image specifications and a quality assurance methodology that is based on ISO standards to ensure that the image quality of the digital products meet certain criteria of fidelity to the original. Fidelity is a major component to authenticity, and authenticity of the digital surrogate is a major component of digital preservation. If the

process is well defined, and the results are monitored and validated, then there is a reasonable assurance that the electronic copies are authentic, reliable, usable, and have the integrity necessary to allow for the disposal of the original records. If records are to be digitized and the originals disposed, then the process needs to perform at a very high standard.

It is important to note that while these guidelines provide parameters for input resolution, bit depth, color space etc., their true value is to provide specific ISO standards based methods to measure image quality. The standards define a set of imaging characteristics that characterize the optical and signal processing performance of an imaging system and compare that performance against ISO standardized criteria. A taxonomy of imaging performance has been developed that measures the Spatial Frequency Response (SFR), Opto-Electronic Conversion Function (OECF), Noise Power Spectrum (NPS), and Spatial Distortion. Reliance upon manufacturer and vendor claims, visual evaluation or adherence to input specifications alone do not guarantee quality.

Records Management Context

Most of the guidance described above has been developed to satisfy internal access or reproduction quality digitization. FADGI uses the term "cultural heritage" as a shorthand to describe the holdings of galleries, libraries, archives and museums. While the holdings of NARA may fit into these categories, the fundamental role of records management is to preserve the informational, evidentiary, and historical value of archival records. FADGI standards represent a high bar for mass digitization projects, and are well suited for records that contain high intrinsic value. The challenge will be implementing these standards on a mass scale.

The bulletin presents digitization as a records management function to reformat and replace records with electronic copies that can be used for the same purposes as the originals. As a records management activity it encompasses a series of processes that result in the creation of provably authentic, reliable, and usable electronic records that are suitable to replace the originals. Digitization includes the planning, selection, assessment, and preparation of records; the technical processes of digitization including metadata capture, quality assurance and quality control; file maintenance and the execution of disposition for the resulting electronic records. When digitizing records agencies must ensure that their digitization processes result in surrogates that meet the requirements for authenticity, reliability, usability, and integrity as described in ISO 15489-1:2016 Information and documentation -- Records management -- Part 1: Concepts and principles.

Technical Specification

Technical specifications for digitization are driven by the characteristics of the original media of the source records and the reformatting approach selected. Decisions regarding the resolution, method of encoding color, and compression selected all impact the usability and reliability of digitized records. Agencies must select technical specifications that are appropriate for the original record type and that take into consideration the following factors:

- The disposition of the original records (either permanent or temporary),
- The retention period of the original records, which may identify the need for a migration plan,
- The media type of the original record (textual paper records, bound volumes, photographic prints, video tapes, etc.),

- The significant characteristics of the original records (wet ink signatures, seals, illustrations, etc.), and
- Program objectives affecting access, use, and preservation.

Equipment

Agencies must ensure that the type of equipment used in any digitization project is appropriate for the source materials selected. Agencies must document all hardware, software, and services used to digitize records. Hardware, software, and services must be selected based on the ability to meet the objectives of the digitization program plan in accordance with the requirements defined in this Bulletin and other relevant standards and regulations. Quality assurance and quality control (QA & QC) procedures must ensure that equipment remains in compliance throughout the project's duration. Agencies must document the QA & QC methods used and retain reports detailing the accuracy of the program. Documenting the QA/QC methods employed, and the results of regular checks and audits, and steps taken in response to failed checks contribute to the authenticity and reliability of digitized records. Technical requirements for specific media types will be provided as appendices to the Bulletin.

File Format Selection

Agencies must select file formats identified as either Preferred or Acceptable in *NARA Bulletin 2014-04*. File formats must be selected based on their ability to preserve identified characteristics. File formats that are standards-based, widely used, and supported by common applications contribute to the usability and reliability of digitized records. File formats appropriate for specific media types will be provided in the appendices.

Metadata

The appendices attached to this Bulletin identify the required Administrative, Descriptive, Rights, Structural, and Technical metadata elements that must be captured during digitization projects. Metadata elements describe different hierarchical levels of records. Elements such as the name of the creating agency or the disposition of a schedule item may apply to all of the records digitized during a project while other elements such as the captions of photographs or titles of reports may be unique to individual records. *NARA Bulletin 2015-04* identifies item-level administrative, descriptive, structural and rights metadata elements required to accompany permanent born-digital records when transferred to the National Archives. To ensure the authenticity and usability of digitized records, it is necessary to capture additional metadata documenting the non-electronic source records as well as metadata describing the electronic files

Quality Management

Quality management is a term that encompasses quality assurance (QA) and quality control (QC). QA seeks to ensure that the products that result from a process meet established quality objectives while QC is the process of statistical sampling of equipment and product to determine how accurate the workflow is performing against the QA targets. QA/QC plans and procedures must be initiated, maintained, and documented throughout all phases of digitization. The plans must address all specifications and reporting requirements associated with each phase of the

digitization project including issues relating to equipment calibration, electronic record validation, metadata capture, and storage (file transfer, data integrity).

QA/QC plans must address accuracy requirements and acceptable error rates for all aspects evaluated. Agencies must validate that digitized records and metadata accurately reflect what was documented by the original record and are usable prior to their being accepted as the record copy. For large digitization projects it may be appropriate to use a statistically valid sampling procedure to inspect files and metadata.

Agencies must review FADGI guidance related to the media type of the original records. FADGI has developed a robust program that defines quality and testing best practices.

Project Documentation

NARA's guidance specifies that agencies must document each phase of their digitization projects and retain that documentation so that it can be transferred to NARA in association with the resulting electronic surrogates. Documentation describing the type of records assessment conducted, the quality assurance measures implemented, the methods used to verify and validate resulting files, the file and folder naming standards, and the results of periodic fixity checks must be maintained and transferred to NARA to support the authenticity and reliability of the electronic surrogates.

Conclusion

NARA's Strategic plan is part of the US Federal government's efforts to modernize records management policies to transition to a digital government. The goal to no longer accept analog records by 2022 will have a major impact on Federal agency records management responsibilities. NARA's Office of the Chief Records Officer has issued several key guidance products to assist agencies meet these challenges. The Records Management Policy and Standards Team has developed digitization guidance based on previous NARA guidelines, and recommends adoption of FADGI and other standards as part of a suite of best practices for the digitization of analogue originals and creation of permanent electronic records. Recommendations based on widely accepted standards allows for easier implementation by agencies and will result in digital products that meet objective criteria.

References

[1] https://www.archives.gov/records-mgmt/bulletins/2015/2015-04.html
[2] https://www.archives.gov/records-mgmt/bulletins/2014/2014-04.html
[3] https://www.archives.gov/about/plans-reports/strategic-plan
[4] https://www.archives.gov/files/records-mgmt/m-12-18.pdf
[5] https://www.congress.gov/bill/113th-congress/house-bill/1233/text
[6] http://www.digitizationguidelines.gov/guidelines/digitize-technical.html
[7] https://www.archives.gov/preservation/technical/guidelines.html
[8] https://www.archives.gov/preservation/products

Author Biography

Michael Horsley is an Electronic File Format Specialist in the Office of the Chief Records Officer, Policy and Planning Program's, Policy and Standards Team, United States National Archives and Records Administration.

Kevin L. DeVorsey is an Electronic Records Policy Analyst in the Office of the Chief Records Officer, Policy and Planning Program's, Policy and Standards Team, United States National Archives and Records Administration.