Preservica
Active digital preservation

Essential Guide

Preserving electronic government records

Integrate active digital preservation into government information management practices to ensure electronic records are findable, readable and useable when required.

Contents

Executive summary

The digital age has fundamentally changed the way that information is created, used and managed. In the public sector, millions of government records are now born-digital or transformed to electronic formats daily making traditional paper and microfilm preservation methods increasingly obsolete. Cybersecurity threats, increased citizen demands for transparency and virtual access to government services are transforming the nature of how public sector agencies operate and ensure continuity of essential services.

> "As formats change, software is retired and hardware becomes obsolete, the data that organizations might want to keep can be lost forever."

The proper care and protection of public records is fundamental to our democracy and it is the duty of elected and appointed officials, agencies, departments as well as government employees and contractors. Many types of long-term[1] and permanent records exist exclusively in "machine-readable" formats and will have to survive numerous technology cycles, custodians, and administration changes to remain authentic and readable far into the future.

Government archivists and records managers are sounding the alarm that electronically stored public records are at risk of being lost, irretrievable or unreadable when required due to bit rot, broken links, obsolete hardware and software, missing metadata and retirement of legacy applications. These risks are compounded by shorter technology refresh cycles, long retention periods, resource constraints plus the sheer volume and diversity of digital content. The situation has prompted leading technology analysts like Gartner to warn that "As formats change, software is retired and hardware becomes obsolete, the data that organizations might want to keep can be lost forever."[2]

This Essential Guide explores how forward-thinking public sector organizations are taking steps to mitigate and protect electronic records by integrating digital preservation technology and good practices into their enterprise infrastructure; recognizing that content servers and backup systems provide only "bit level" protection and are insufficient for ensuring authenticity and future readability. Government leaders need to work together to identify and deploy electronic records lifecycle management approaches that will protect citizen rights, faithfully document the decisions and actions of government, and preserve our shared history.

This Guide also describes recognized digital preservation standards and community-based practices for future-proofing long-term digital content and electronic records, and highlights why it is only a matter of time before compliance and cybersecurity auditor—as well as the general public—demand evidence that their government agencies and institutions have essential capabilities in place to ensure permanent electronic records are findable, readable and useable when required by future generations.

1 Long-term is defined in the digital preservation community's de facto standard as a period of time that is "long enough to be concerned with the impacts of changing technologies, including support for new media and data formats or with a changing user community. Long Term may extend indefinitely."
2 https://www.gartner.com/en/documents/3038917/cool-vendors-in-content-management-2015

Transition to electronic government

Government technology environments are complex and constantly evolving. Many types of public sector records that must be retained permanently—such as board meeting minutes, land deeds and plats, court case files, correspondence of elected officials, audio and video recordings, vital statistics, publications and reports—are now managed exclusively in electronic formats.

These public records will have to survive numerous technology cycles, custodians and administration changes to remain authentic and readable far into the future. The preservation and transfer of digital information is widely considered to require more proactive and continuous attention than that of other media types.

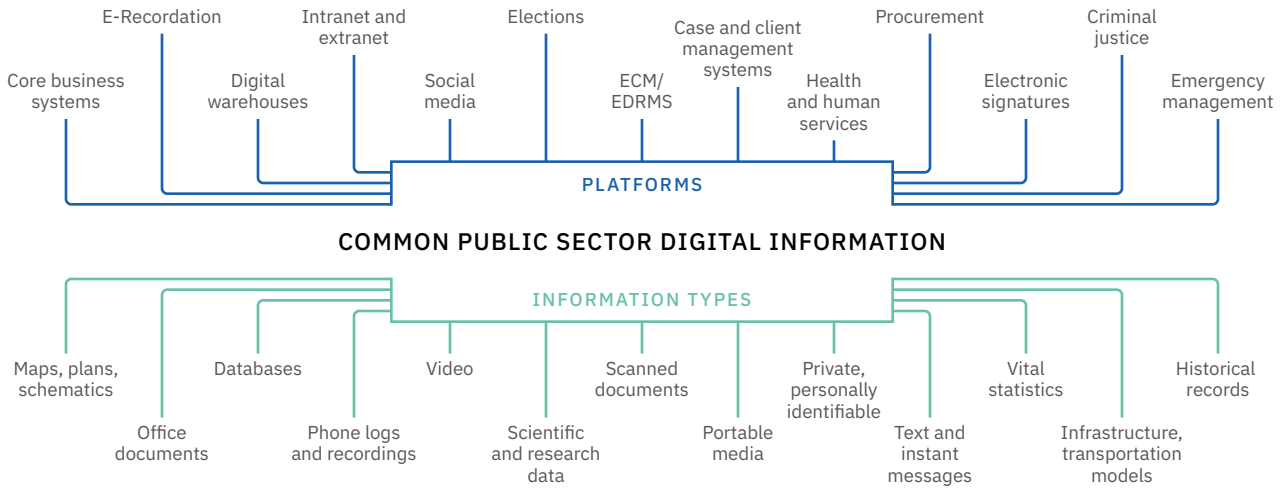
Compounding these realities is the exponential growth and diversity of electronic records that public sector organizations are handling. No longer just traditional text and document formats but increasingly multi-media and interactive assets like audio/visual recordings, images, CAD drawings, spatial datasets, websites, social media posts and emails with embedded links, files and attachments.

The transition to electronic government detailed in a recent series of federal government initiatives and deadlines all point to the reality that 'print to paper' as an official recordkeeping practice is no longer viable. By way of example, as of 12/31/2019[3] federal agencies are expected to manage in electronic formats all their born-digital records appraised as "permanent" for eventual transfer and accessioning to the National Archives and Records Administration (NARA). And after 12/31/2022, NARA will no longer accept legacy paper records in any form. This means that legacy paper records appraised as permanent records still in the custody of agencies will have to be digitized before transfer to NARA.

These developments and similar digital continuity initiatives around the world are testament to the fact that public sector institutions need to actively integrate digital preservation capabilities into their existing electronic records and information management practices to ensure the longevity and viability of "machine-readable records" both while in the custody of producing agencies and after transfer into the custody of historical archive institutions.

As we will explore in the next section, without the proper processes, policies and systems in place for long-term digital preservation, electronic records remain at constant risk of not being findable, readable or useable when required.

Common public sector digital information platforms and types



3 https://www.whitehouse.gov/wp-content/uploads/2019/06/M-19-21.pdf

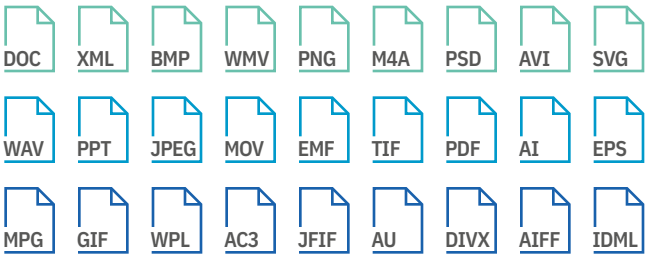
Digital information is fragile and at risk

Researchers, digital preservation coalitions and communities, as well as the real-life experiences of IT professionals concur that digital content is fragile and at risk—susceptible to intentional or accidental loss, bit corruption and media degradation, technology refresh cycles, vendor abandonment, as well as file format and software obsolescence. The fragility of digital content as a significant potential risk to government and business has been identified in recent years by leading analyst firms that include Forrester, Info-Tech Research Group and Gartner. In 2016, Gartner noted:

> **"Governments and associated agencies have important roles to play, not only for the duration of individual lives, as in care, adoption and criminal records, but also in maintaining historical facts and records for generations to come. This, therefore, places additional requirements on government CIOs, architects and archivists to preserve true chronicles as the records' accuracy and future value will need to be maintained through several technology changes."**[4]

Technology refresh cycles are relatively short, in the three- to five-year range. Like hardware and storage media, file formats can become outdated, obsolete or unsupported, often without users realizing it. Think of once popular formats like Lotus 1-2-3 and WordPerfect, which have come and gone.

Widespread adoption of open standard technology neutral ("OS/TN") formats like JPG, PDF/A or SVG by records producers to store electronic records over the past few decades only partially addresses this threat. These formats are often inconsistently used across the enterprise and fall short of a one-time solution for indefinite long-term or permanent records preservation.



Diversity of digital formats

As stewards of America's history and heritage, government archivists preserve and provide access to records in their care regardless of format. The revolution in electronic recordkeeping over the past 50 years combined with the minimal investment[5] made in government records management and archives, has put important digital information about our identity and accomplishments as a nation at great risk of loss.

Both IT units and agencies reported in a 2019 CoSA research report[6] that the state archives are rarely consulted in the many stages of technology planning, refresh and retirement. Add to this the reality that transfer from agencies to archives may take place many decades after the records were first captured, and it is clear there are widespread threats to government records and an urgent need to implement and sustain enterprise-wide electronic records management and digital preservation capabilities.

Any part of the technology stack that suddenly lacks support or backward compatibility will impact an agency's ability to use the application securely while also posing threats to the integrity of the content generated. Documented examples Apple's removal of support for QuickTime for Windows, prompting mass video migration to HTML 5 and MP4 formats. More recently Microsoft announced the end of for Windows 7. **When obsolescence happens, accessibility becomes a challenge.**

4 Neville Cannon, Is Cloud Fit for Government Archiving? Gartner, April 2016.
5 On average, state and territorial governments spend .007% of their budgets on records management and archives according to the Council of State Archivists (CoSA), A National Risk: The State of State Electronic Records Report, 2017.
6 CoSA, Toward a Common Understanding: Insights on Inter-Agency State Electronic Records Transfer (Frankfort, KY: December 2019) p. 6.

Standards for safeguarding digital objects and metadata

Digital preservation is a formal set of processes and activities that maintain long-term information stored in digital formats in order to ensure continued access. Digital objects managed in a purpose-built preservation system are actively migrated over time to newer formats using policies and automated workflows. Preservation actions are captured in the metadata associated with each digital object to demonstrate authenticity and chain of custody. Digital objects are organized into collections and shared in accordance with the organization's unique permissions and rights.

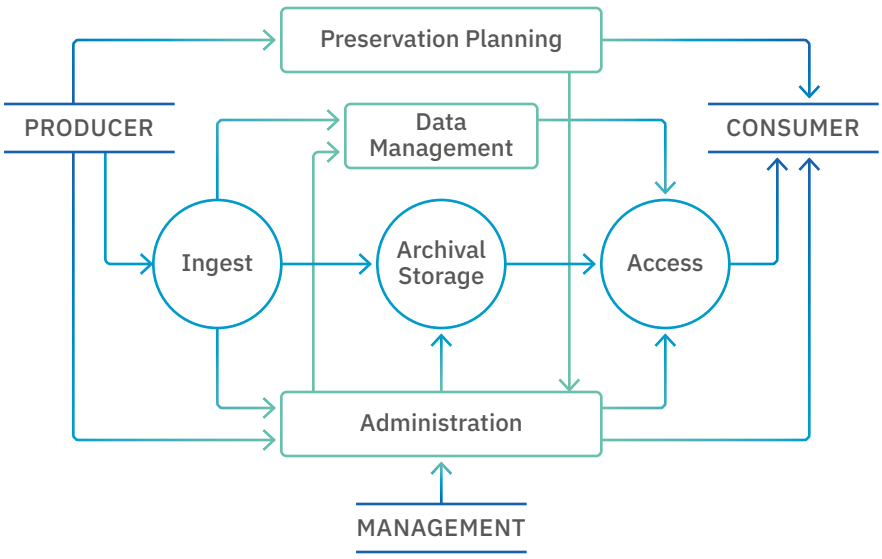
Digital objects require continuous monitoring to maintain their evidentiary status as records. Establishing fixity via checksums or cryptographic hashes is widely relied upon to provide proof of the electronic records' authenticity and reliability. Fixity provides a wrapper, or protective shield,[7] ensuring the content remains unchanged. It must be managed and preserved for the entire lifecycle of the electronic record. ECM/RM and commonly used business applications fall short of preserving and monitoring the content, structure, context and integrity of records over extended periods of time.

Many open source and commercial digital preservation tools are based on the Open Archival Information System (OAIS) reference model which was released in 2003 as the ISO 14721 standard and updated in 2012. OAIS is a conceptual framework for functions and actions that a digital repository must execute to ingest, store, preserve and provide access to digital objects for a community of users.

Common activities for digital archives conforming to the OAIS functional model include file ingest and characterization, integrity validation and protection, collection management, system and data monitoring, migration of assets from obsolete file formats, replication to multiple geographic locations, robust metadata management to facilitate search and retrieval, and secure access.

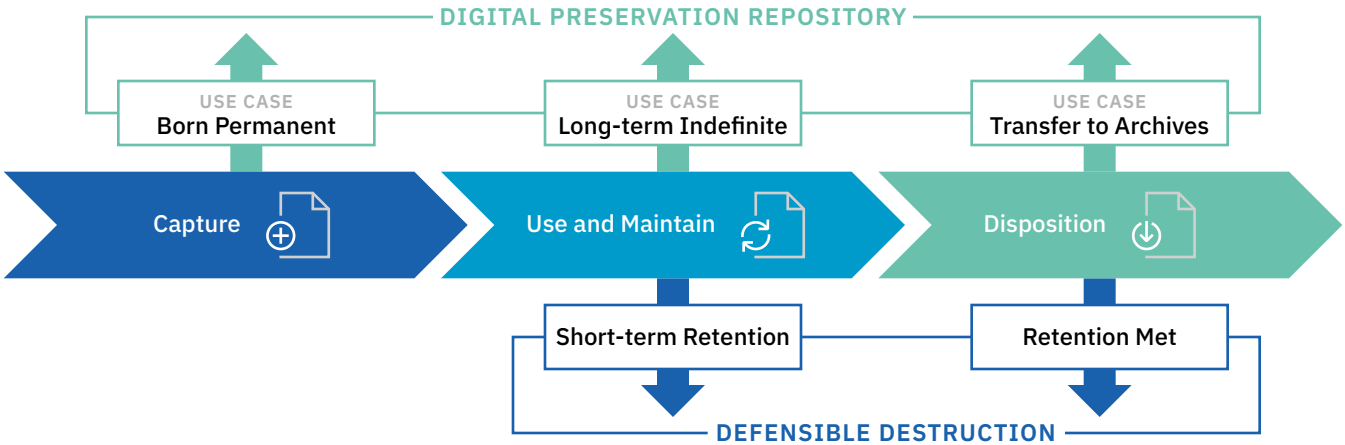
A companion standard, Audit and Certification Criteria for Trustworthy Digital Repositories (ISO 16363:2012), includes OAIS technical functions in addition to identifying a range of organizational and security management capabilities and metrics. Dedicated resources—funding, skilled staff, tools, storage and organizational commitment—are required for a trusted digital repository to persistently monitor risks and adapt to changing conditions.

OAIS Functional Model



7 Consultative Committee for Space Data Systems (2012) 'Reference Model for an Open Archival Information System (OAIS)', CCSDA Secretariat, Washington, D.C., p. 2-7.

Integrate digital preservation into the records management lifecycle



Government information and records managers face many challenges in consistently applying retention and defensible disposition practices to electronic records and data sets. Within any given jurisdiction, government records are managed in many different formats and by hundreds of specialized applications. Record retention periods vary from brief (e.g., 30 days) to indefinite (e.g., termination plus 30 years) to perpetual (e.g., permanent). And because metadata practices across the enterprise are often inconsistent, search and re-use of documents by staff to support public record requests, investigations or litigation can be inefficient and incomplete.

Government recording offices and agencies rely on specialized software solutions as well as common enterprise applications like email and Enterprise Content Management (ECM) systems to capture, share and manage revision control and governance of electronic records. These systems function as official records systems and enable storage and control to varying degrees including the application of retention periods, disposition triggers, search, legal hold and authorized destruction. Enterprise IT has protocols and workflows in place to transfer and restore records systems from archival storage and backup.

It's important to note that records held in either RM systems and archival storage and backup systems are only protected from loss or corruption at a *bit level*—there is no mechanism to ensure future readability or usability. **This is where active digital preservation comes in.**

The primary role of a digital preservation system is to ensure records and their metadata remain accessible, useable and readable over the long-term by providing a proactive way to migrate file formats as they become obsolete or are no longer supported by a vendor or by the agency. Adopting a *preservation approach* to the routine capture and management of permanent electronic government records requires a formalized migration approach, planning and budgeting for hardware and software upgrades, the means to facilitate automated transfer of content, and knowledgeable records personnel. Managing the lifecycle of electronic records is by necessity a cross-boundary group effort.[8]

CIOs are primary stakeholders who need to determine how electronic records management requirements are addressed in IT strategies and enterprise architecture. Other key stakeholders in the lifecycle of electronic records include legal counsel and compliance officers, elected and appointed officials, security and privacy experts as well as records managers and archivists.

8 CoSA and NASCIO, State Archiving in the Digital Era, A Playbook for the Preservation of Electronic Records, 2018, p. 4.

Digital preservation use cases in the US public sector

Federal Agency On December 28, 2018, the Government Publishing Office (GPO) made history by becoming the first organization in the United States and only the second organization in the world to achieve ISO 16363 certification as a trustworthy digital repository. GPO is the Federal Government's official, digital, secure resource for producing, procuring, cataloguing, indexing, authenticating, disseminating and preserving the official information products of all three branches of the U.S. government.

The scope of the audit was GPO's govinfo repository[9] which includes a content management system, preservation repository and public access website. The audit evaluated the operations, development, features, policies, procedures, personnel and digital preservation activities against the ISO 16363 requirements.[10] There are two instances of the repository system, a primary instance and a continuity of operations instance with identical hardware setups and automatic synchronization of databases and files. Current holdings are approximately 65TB.

State Archives and Records Administration The Vermont State Archives and Records Administration (VSARA) is charged with administering the Statewide Records and Information Management (RIM) Program for all public agencies in accordance with generally accepted record-keeping principles and industry standards and best practices. The Statewide RIM Program in defined in 3 V.S.A. § 117. VSARA, a division of the Office of the Secretary of State, implemented a digital preservation system in 2014 to satisfy the storage and preservation needs of its digital state archives called VT Re•tain. More than 300,000 digital objects have been ingested and VSARA recently upgraded its trusted repository to allow for more storage. Collections include gubernatorial and legislative records as well as municipal reports.

County Records and Archives Facing an exponential growth in permanent public records stored electronically, a large county in the Pacific Northwest implemented a cloud-based digital preservation solution to support 200+ county agencies and preserve the records of elected officials, departments, programs, committees, and councils. The Records and Archives team manages the County's electronic document and records management system (EDRMS) and routinely receives electronic records from shared drives, phones, iPads, Google docs and dozens of business applications.

In addition to protecting and proving access to permanent public records, the County plans to use its digital preservation repository to back up the essential records currently created and managed in the EDRMS. This is to ensure that records related to services such as community justice and health and human services, as well as internal support records such as employee benefits and payroll, will be protected and available in the case of emergency or service disruption.

City Archives The City of Boston Archives house the permanent historical and administrative records of the Boston City government. This includes mayoral files, correspondence, public works records, tax records, along with school department, fire department and City Council records. Prior to establishment of the Archives in 1989, the records of the city government were scattered in various city buildings and agencies.

More recently the challenge shifted to not only transitioning paper records into digital form to protect them and enable computerized search and access, but also to preserve the increasing volume of born-digital records including modern information sources such as emails, websites, videos, presentations and word documents.The Boston City Archives implemented a digital preservation system in 2016 which allows them to preserve, flexibly manage and provide wider access to its unique digital records. Using a single cloud-hosted application has enabled the Archives to increase efficiency and focus on their role of curating and sharing these records rather managing local IT servers and individual tools.

9 www.govinfo.gov
10 https://www.fdlp.gov/file-repository/preservation/3910-the-ptab-iso-16363-audit-of-the-gpo-govinfo-repository-system-public-report

Summary and call to action

Public records are vital to citizens, businesses, the legal community as well as staff at all levels and in all branches of government. The right to access records is enshrined in federal and state laws. Consequences of failing to preserve and produce documentary evidence of government decisions, operations and heritage—or to produce an authentic copy on demand, are well documented—loss of public trust, non-compliance, risks to public health and safety, fines and litigation and the inability to re-use institutional knowledge to meet evolving business needs.

While digital preservation has primarily been addressed by archival communities for the past few decades, the world's reliance on computer technologies has moved the discipline into the purview of records management and information governance. Statutory mandates, stringent privacy rules, fast technology refresh rates, shorter response times, the transition to electronic government and the sheer volume and diversity of digital content under management means that all government agencies and archives face the risk that permanent and long-term electronic records will not be findable, readable or useable when required in the future.

Due to the fragility of digital content, preservation actions must be taken over the lifetime of electronically stored information to mitigate the risks associated with changes in software and hardware environments; deterioration of magnetic media, such as CDs, DVDs and computer hard drives; and to keep pace with evolving business, legal and regulatory requirements for access and re-use. Just storing records indefinitely in RMS/ECM systems, on email servers or in archival storage and backup systems provides only "bit level" protection for long-term records and is not a defensible or scalable approach for ensuring future readability.

With the majority of government records now captured and managed for their respective lifetimes in electronic format, CIOs, CTOs and other government leaders are called upon to demonstrate they have the policies, processes and systems in place to ensure digital public records collections under their care will be accessible, readable and useable in the future when required. Failure to act could add to their technical debt[11] and undermine the public's trust in the institution.

While digital preservation has primarily been addressed by archival communities, the world's reliance on computer technologies has moved the discipline into the purview of records management.

Forward-thinking public sector organizations are taking proactive steps to mitigate inherent technology risks and ensure they can continue to meet fiscal, legal and administrative obligations. This requires the modernization of records preservation approaches and practices that conform to standards and best practices for the long-term preservation of digital content like OAIS (ISO 14721) and certification criteria for trustworthy digital repositories (ISO 16363).[12]

Closer integration between electronic records producing and records preservation systems has the potential to expand opportunities for enterprises to retire legacy applications, backup essential records stored in front-end operational systems, free up IT resources and budgets and ensure the secure storage of documentary evidence of government operations and decisions.

11 https://www.techopedia.com/definition/27913/technical-debt
12 https://www.gpo.gov/who-we-are/news-media/news-and-press-releases/gpos-govinfo-makes-history-by-earning-global-certification-for-trustworthiness

Resources

NARA Digital Preservation Framework
Approach to determining risks faced by electronic files and plans for preserving different types of file formats.

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

NARA Universal Electronic Records Management Requirements Version 2
The Universal ERM Requirements identify high level business needs for managing electronic records and are a starting point for agencies to use when developing system requirements.

https://www.archives.gov/records-mgmt/policy/universalermrequirements

NDSA Levels of Preservation (LoP)
Resource for digital preservation practitioners when building or evaluating their digital preservation program.

https://ndsa.org/publications/levels-of-digital-preservation/

Council of State Archivists (CoSA) SERP Framework
The State Electronic Records Preservation (SERP) Framework provides information and guidance in 15 different areas on how to move forward with digital preservation.

https://www.statearchivists.org/electronic-records/serp-framework/

National Association of State CIOs (NASCIO) Electronic Records Preservation Playbook
This document includes eleven plays that state officials should consider when working together toward the preservation of digital archives.

https://www.nascio.org/resource-center/resources/state-archiving-in-the-digital-era-a-playbook-for-the-preservation-of-electronic-records/

ISO 14721 Open Archival Information System (OAIS) Reference Model

http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

ISO 16363 Audit and Certification of a Trustworthy Digital Repository

http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

Primary Trustworthy Digital Repository Authorisation Body (PTAB) Self-Assessment Template
A simple spreadsheet available from PTAB with the ISO 16363 metrics and sub-metrics which a repository can use to record evidence about current capabilities.

http://www.iso16363.org/sdm_downloads/iso-16363-self-assessment-template/

About Preservica

Preservica is changing the way that organizations around the world protect and future-proof critical long-term and permanent digital assets. Available in the cloud (SaaS) or on premise, our award-winning active digital preservation software has been designed from the ground up to tackle the unique challenges of ensuring digital information remains accessible and trustworthy over decades.

Preservica is a proven solution trusted by a growing number of governments, archives, libraries, museums and businesses around the world including the City of Boston, University of Notre Dame, Associated Press, MOMA, Transport for London, World Bank, Yale University and 21 state archives including Texas, California, Kentucky and Massachusetts.

For information, please visit www.preservica.com.