

Public Report: The PTAB ISO 16363 Audit of the GPO govinfo Repository System

Dec 2018



MAY 2019

Primary Trustworthy Digital Repository Authorisation Body
(PTAB)

<http://www.iso16363.org>



Introduction

This document was requested by the U.S. Government Publishing Office (GPO) to describe why and how the GPO obtained ISO 16363 certification, with insights both from GPO and the external auditors.

Digital preservation is not a simple task. For many years there has been a demand for a way to judge whether a repository can be trusted to carry it out properly. Funders of repositories and those who entrust their valuable digitally encoded information to them are entitled to know whether their funding and their faith is well founded. Stakeholders need to know if the repository is worthy of trust.

Third party auditing of digital repositories to widely respected international standards offers the best approach to ensuring that the repository is truly a Trustworthy Digital Repository. The GPO contracted the Primary Trustworthy Digital Repository Authorisation Body Ltd (PTAB) to perform an initial International Organization for Standardization (ISO) 16363 audit of the FDsys/govinfo repository. The govinfo website replaced the FDsys website in December 2018, and at that time, GPO began to refer to the entire system as govinfo, which is used in the rest of this document.

GPO Rationale for the Audit

The mission of the GPO is “Keeping America Informed” as the official, digital, and secure source for producing, preserving, and distributing official Federal Government publications and information products for Congress, Federal agencies, and the American public. Permanent public access to Federal Government information is provided free of charge through the [Federal Depository Library Program \(FDLP\)](#) and [govinfo](#). From its conception in 2004, the Federal Digital System (govinfo) was designed with the intent to ingest, preserve, and provide access to the information produced by the U.S. Government—including information produced by all three branches of Government— and to the digitally imaged content from Federal Depository Libraries and Federal library partners. In the *Concepts of Operations (ConOps) v2.0*, the system was to be developed according to standards and guidelines accepted by the digital preservation and library community at the time, including Open Archival Information System (OAIS) ISO 14721 framework compliance. In January 2015, GPO initiated a project to assess govinfo against the ISO 16363 Trustworthy Digital Repository standard in order to have third-party validation that GPO’s implementation of its repository had fully actualized its original mission of creating an ISO 14721 OAIS repository.

Certification under ISO 16363 would reinforce GPO's commitment to its mission of ensuring permanent public access to U.S. Government information through the preservation of content in digital formats.

To begin the audit process, GPO began the audit process by asking the National Digital Stewardship Residency (NDSR), co-funded by the Library of Congress and the National Institute of Museum and Library Services to provide it one of five residents for one year. Under this program, host institutions are provided one resident for one year to work on a significant project in digital stewardship. GPO's NDSR Resident worked on preparation for the audit and certification of govinfo as an ISO 16363 Trustworthy Digital Repository.

Prior to the audit, GPO reviewed existing repository documentation and performed a gap analysis against the requirements for the ISO 16363 certification. A cross-organizational team comprised of representatives from the Office of Programs, Strategy, and Technology and Library Services and Content Management business unit met bi-weekly to discuss the current status of the documentation that needed to be gathered or developed for the audit. Additionally, GPO conducted outreach with twelve digital preservation repositories which claimed to be OAIS-compliant based on previously being certified under TRAC to gather feedback from the digital preservation community. This feedback regarding standards-based certification of repositories helped identify how GPO could successfully verify the designation of “trustworthiness.” The completion of the formal ISO 16363 audit served as the culmination of many years of preparation and earnest expectation for GPO, GPO stakeholders, and its Designated Community.

ISO Standards and Audit process

Three standards form a closely related family of standards that are important in establishing an internationally recognized and certified set of trustworthy digital repositories. All three were created by the Consultative Committee for Space Data Systems (CCSDS) and subsequently approved through the ISO approval process. The CCSDS Working Groups responsible for developing these standards solicited participation from digital preservation experts not only from space agencies and the aerospace industry, but also from experts from cultural heritage institutions (e.g., libraries, archives, museums, publishers), from the defense sector, from the scientific arena, and from the Information Technology (IT) community. Members of the PTAB both led and participated in the development and approval of the standards.

- ISO 14721:2012, *A Reference Model for an Open Archival Information System (OAIS)* is the reference model for what is required for an archive to provide long-term preservation of digital information.
- ISO 16363:2012, *Audit and certification of trustworthy digital repositories* sets out comprehensive metrics for what an archive must do, based on OAIS, covering Organizational Infrastructure, Digital Object Management including access, and Infrastructure and Security Risk Management.
- ISO 16919:2014, *Requirements for bodies providing audit and certification of candidate trustworthy digital repositories* specifies the competencies and requirements for auditors and serves as an extension to ISO/IEC 17021 *Requirements for bodies providing audit and certification of management systems*, providing additional requirements specific to auditing Trustworthy Digital Repositories.

In order to be able to award ISO certification, the auditors, in this case PTAB, must follow the processes laid out in ISO 16919 and ISO 17021, the latter being the standard which governs audits for all kinds of management systems on which we as individuals depend in our everyday lives in areas ranging from food safety to medical devices to information security to environmental management, and many more. The general process, as used by PTAB, is illustrated in Figure 1.

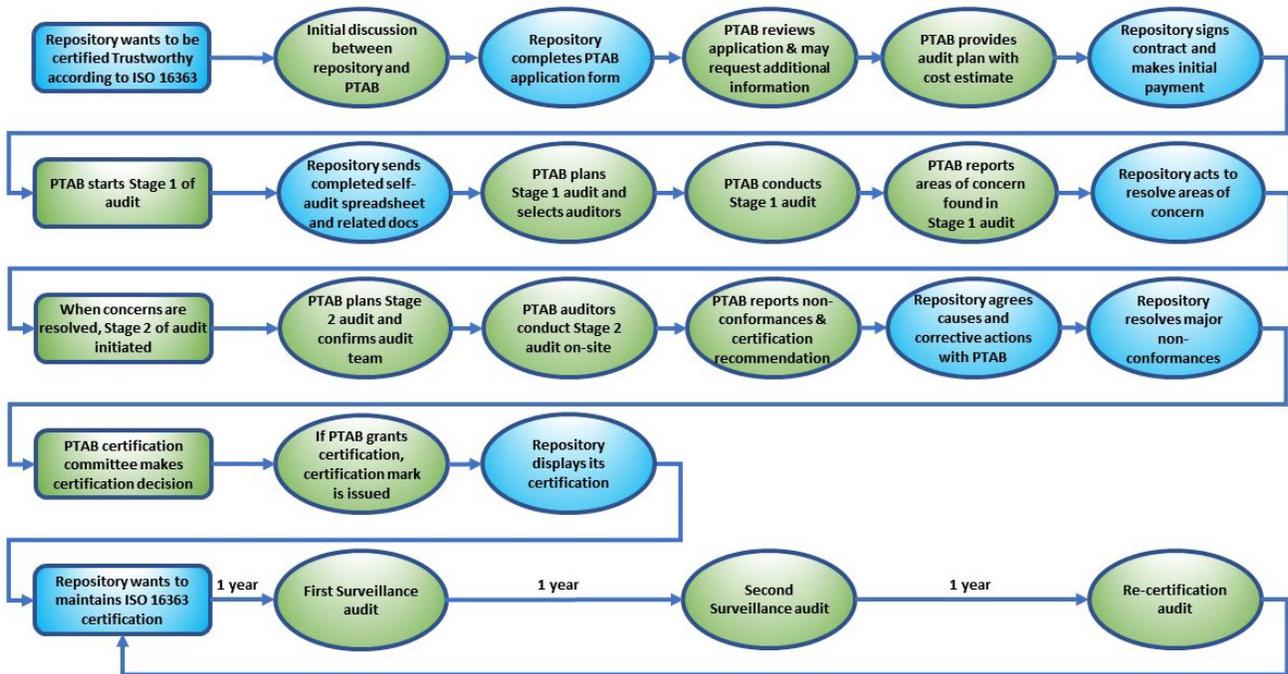


Figure 1 PTAB Audit process

Within the figure, the blue boxes and ovals indicate points where the candidate repository (i.e., GPO's govinfo) is required to provide inputs or make decisions. Once those inputs are provided, PTAB initiates the next task and evaluates the materials supplied by GPO and continues with the mandated auditing and certification process. Further details of the PTAB processes are available on the PTAB website <http://www.iso16363.org>.

ISO has determined that the process of annual audits, in a three-yearly cycle, is appropriate for maintaining certification for any type of management system. Therefore, following the initial certification audit, certified repositories must be subject to subsequent audits if they wish to maintain certification over time.

Processes for the GPO's Audit

Audit Objectives

The objective of the audit was to evaluate GPO's govinfo as a Trustworthy Digital Repository according to all the criteria specified in ISO 16363 and:

- determine the effectiveness of the management system to ensure the client can reasonably expect to achieve its specified objectives;
- identify, as applicable, areas for potential improvement of the management system.

Scope of the Audit

GPO is a large organization with many parts, holding many types of information. The initial discussions with GPO confirmed the scope of the audit to be the govinfo repository, covering its content management system, preservation repository, and public access website. The audit assessed the features, operations, development, procedures, policies, personnel and digital preservation activities against the ISO 16363 requirements.

A successful audit requires the cooperation and participation of many staff members from the organization providing the auditors, as well as staff members from the organization being audited. Key staff members involved in the GPO govinfo system audit from PTAB and from GPO are identified in the subsections below.

PTAB Audit Team

PTAB chose its audit team to be sure that they had the competencies required, as specified in ISO 16919, to perform the audit and understand what must be covered by the audit. The PTAB Audit Team consisted of David Giaretta (Lead Auditor) and John Garrett, who together co-chaired the groups which wrote OAIS, ISO 16363 and 16919, with Bruce Ambacher and J Steven Hughes who were major contributors to the three standards. The PTAB Certification Committee for GPO was made up of Helen Tibbo and Terry Longstreth, both of whom were also major contributors to the three standards. Descriptions of these individuals are available at <http://www.iso16363.org/ptab/>.

GPO Audit Team

Support was provided by many individuals and organizations in GPO including but not limited to the GPO Director, the Chief Technology Officer, the Superintendent of Documents and various organizations such as the Office of Programs, Strategy, and Technology and the Library Services and Content Management business unit. GPO's Digital Preservation Librarian served as Project Manager and PTAB's primary point of contact. GPO's TDR Project Team was a cross-functional team comprised of employees from GPO's Library Services and Content Management business unit and the Office of Programs, Strategy, and Technology.

Stages of the Audit

The ISO process for audits that conform to ISO 17021 and ISO 16919 consists of two stages. Stage 1 involved a detailed review of the GPO's documentation, prepared by GPO, of its systems, policies, and processes. Most of these documents were existing documents of the sort which any good repository should have, for example, its system design and operating procedures; others were created to provide explicit evidence for the audit, for example, the explicit definition of the Designated Community. There were several discussions between PTAB and GPO to clarify issues and obtain more information. Based on this, the PTAB identified areas of concern, which are things that could be found to be nonconformances during the on-site visit. The purpose of

this was to allow the repository staff to correct these issues, for example, by providing missing documentation, so that the on-site visit would run smoothly. GPO therefore had time to resolve the issues which were identified. When this was done, GPO and PTAB agreed that Stage 2 could begin.

Stage 2 involved a two day visit to the repository so the PTAB audit team could witness what actually happens, ask questions, and examine additional evidence such as seeing software systems running, with the aim of identifying major and minor nonconformities. Major nonconformities are those likely to affect whether the repository can be considered trustworthy i.e. meets the requirements of ISO 16363. In general, the repository has 6 months to resolve major nonconformities to the satisfaction of the auditors. During the Stage 2 site visit, PTAB auditors met with GPO to review evidence of conformity to the metrics of Digital Object Management, Information and Access Management, and Technical Infrastructure and Risk Management. The schedule for the Stage 2 site visit is attached as Appendix A.

The audit team delivered a report, as specified in ISO 17021, to the Certification Committee, whose membership excludes the audit team. The Certification Committee reviewed the report to determine if the ISO processes were followed and had the opportunity to review selected evidence, and then made the decision as to whether or not to award certification.

As required by ISO 17021, Surveillance Audits must take place at least annually following certification. A Recertification Audit is required within 3 years following the initial certification and the process repeats. These activities are essential for maintaining ISO 16363 certification and continuous improvement of repository policies and workflows.

Surveillance audits focus primarily on areas related to the non-conformities found or observations noted in previous audits and areas where changes occurred in the system. Typically, a more focused examination of specific ISO 16363 criteria is included. However, the full set of ISO 16363 criteria is subject to review.

The Recertification Audit is more like the initial Certification Audit which GPO just completed. However, a difference would be extra focus on any non-conformities or changes to the information system identified during the prior certification cycle.

It must be pointed out that PTAB is itself assessed on an annual basis in order to be able to perform audits and grant certification. PTAB's assessors are themselves annually assessed. Besides needing to perform the audit following the specified process, ISO 17021 requires that PTAB procedures must fulfill a number of other requirements, including respecting the confidentiality of its clients' information. For example, this document, which contains information about GPO, is published by GPO rather than PTAB.

Summary of govinfo

The United States Government Publishing Office (GPO) is the Federal Government's official, digital, secure resource for producing, procuring, cataloguing, indexing, authenticating, disseminating, and preserving the official information products of the U.S. Government. The GPO is responsible for the production and distribution of information products for all three branches of the Federal Government, including U.S. passports for the Department of State as well as the official publications of Congress, the White House, and other Federal agencies in digital and print formats. GPO provides for permanent public access to Federal Government information at no charge through govinfo (www.govinfo.gov) and has partnerships with approximately 1,100 libraries nationwide participating in the Federal Depository Library Program (FDLP).

The GPO govinfo holdings size is currently 65TB; the repository has 2.2 million Archival Information Packages (AIPs) and 1.5 million GPO specific Access Content Packages (ACPs) which are optimized for access and delivery. The AIPs are made up of 1.1 million United States Courts Opinions collection AIPs, 600,000 Congressional AIPs, and smaller numbers from other sources. In FY17, there were a total of 540,000,000 document retrievals from the public website and an average of 45,000,000 retrievals per month. The system maintains a high level of availability. The staffing is approximately 40 FTEs, almost evenly divided between government employees and contractors. There are two instances of the system, a primary instance and a continuity of operations instance. The two instances have identical hardware setups; the databases and files are synchronized automatically so that if the primary instances fail, or is taken down for some other reason, then the services can be provided by the continuity of operations instance. The govinfo website replaced the FDsys website in December 2018, and at that time, GPO began to refer to the entire system as govinfo.

Audit Findings

The PTAB auditors found that there were no major or minor nonconformities, despite in-depth questioning and examination of evidence, and recommended that the govinfo system could be regarded as trustworthy to preserve the information it held for their specified Designated Community.

The PTAB certification committee examined the reports from the audit team and made the decision to award certification. Certification was granted on 28 December 2018. It is valid until 28 December 2021 subject to successfully completing annual Surveillance Audits.

GPO's govinfo system, in all of its aspects; holdings, system hardware and software, physical and content security, and staff; fulfilled all of ISO 16363's criteria and has been certified as a Trustworthy Digital Repository. This does not, however, mean that the system is perfect. Nor does it mean that this certification could be used to show that the system could be trusted to preserve information outside of the scope of this audit.

Major Factors for GPO's Successful Certification

In the view of the PTAB auditors there were a number of key factors contributing to GPO's successful certification.

The GPO team was very well prepared and had an excellent digital preservation system, both in terms of systems and people. It was clear that they had been planning for ISO 16363 certification for quite some time. GPO had developed comprehensive evidence to support all 100+ metrics and sub-metrics of ISO 16363 which were examined by PTAB.

As a general comment the GPO team had carefully studied ISO 14721 (OAIS) and ISO 16363 and the processes GPO uses are thoroughly documented. The GPO document tree was very thorough and provided a chain of evidence from a summary self-evaluation document of ISO 16363 Compliance through a hierarchy of documents with detailed explanations and referenced evidence for each metric, and finally the detailed GPO documents provided as evidence. Multiple individuals have expertise in each of the repository processes so that no person would be a single point of failure.

All the criteria of ISO 16363 were found to be adequately met, but to give a brief indication of some of the important aspects considered in the audit the following points are will be of interest.

Organizational Infrastructure

- The Designated Community was specified in significant, explicit detail. This makes clear to GPO management as well as depositors and users what the preservation system promises to do while allowing the auditors to properly assess the preservation system. The clarity of the detailed specification was of the utmost importance.
- As mentioned above, the distributions of skills and knowledge, in the mix of government employees and contractors avoids single points of failure.
- GPO has a succession plan which includes a signed agreement with the successor (NARA in this case); this is something most repositories do not explicitly have.
- There must be an adequate funding horizon so that even if GPO as a whole or just govinfo were to cease operations, there would be time to properly transfer the contents

to NARA in accordance with the succession agreement. GPO's legal position means that its funding horizon will satisfy this requirement.

Digital Object Management

- The transfer of information to GPO is clearly defined, covered by its statutory duties.
- The logical design of the Archival Information Package adequately captures the information required to preserve their holdings.
- Preservation plans are adequate for the types of information GPO preserves, given its specific Designated Community.
- The user community is very large and diverse. The way in which access is enabled is tailored to suit a variety of needs. Most importantly, GPO has mechanisms to gauge whether and when changes are needed.

Infrastructure and Risk Management

- The digital objects are stored safely and securely in the hardware at GPO's primary instance, with a copy in GPO's continuity of operations instance. Arrangements are in place to ensure there is sufficient separation of the backup.
- Risks are reviewed on an almost daily basis. Major risks have mitigation plans, and these are verified through testing.

Recommendations for Improvements

As part of the ISO process, the auditors identified potential improvements which could be made. Note that the way in which these improvements might be implemented was not specified. While not identified as major or minor non conformances, the GPO team should:

- Continue to work with Producers to improve standardization of the SIPs submitted.
- Continue to monitor the impact of multiple software packages used by govinfo to facilitate operation of the system for full integration and most efficient operation of the system.
- Current preservation system works well for current holdings, but GPO may need to prepare for new publishing paradigms.

Conclusions

The ISO 16363 audit of GPO's govinfo system was successful. All the objectives of the audit were achieved, following the tried and tested ISO 17021 procedures. The repository was certified to be a Trustworthy Digital Repository but this certification, as with all the auditors and assessors in the ISO system, must be tested annually.

The GPO has wisely chosen to undergo the rigorous process of obtaining ISO certification as a Trusted Digital Repository for their govinfo system through a third-party audit by the Primary Trustworthy Digital Repository Authorisation Body Ltd (PTAB) organization. This provides a number of benefits for GPO stakeholders.

For those who fund repositories or who deposit their valuable resources

- reassurance where it is warranted – and warnings where trust is not warranted;
- comfort that someone besides the repository managers can attest that the repository has (or has not) been doing a good job;
- reassurance that the digitally encoded information will be usable into the future rather than simply identifying how well the bits will be preserved;
- identification of areas which need improvement.

For repository managers

- advice on where improvements are needed

For users

- assurance that the repository's processes are providing, over time, usable, authentic information which was derived from its depositors.

Appendix A

DAY 1

Day_1	6 th Dec 2018		Lead
Time	Activity		
09:00	Arrival		
09:15	Opening meeting: Introductions, confirmation of plans and expectations.		PTAB
09:45	Review of points arising from Stage 1 of audit to ensure that all areas of concern and questions are answered at this point or else will be resolved during remainder of the audit.		PTAB
10:45	Walkthrough by client Overview of the repository operations, and information flow from Ingest to Dissemination and overview of Preservation activities including policies, adherence to Information Model and Mandatory Responsibilities. This should, if possible, consist mainly of demos of actual ingest and access sessions, to provide an overview of the actual GPO processes;		GPO
13:00	Lunch		
13:30	Team 1 Digital Object Management – section 4 of ISO 16363 4.1 Ingestion: Acquisition of Content, 4.2 Ingestion: Creation of AIP, 4.3 Preservation Planning,	Team 2 Organizational Infrastructure - section 3 of ISO 16363	PTAB
17:30	Internal meeting of the team		PTAB
17:45	Feedback to Client on day's findings.		PTAB

DAY 2

Day_2	7 th Dec 2018		
Time	Activity		Lead
09:00	Arrival and start of Audit	Travel to ACF	
09:15	Digital Object Management – section 4 of ISO 16363	ACF Site Inspections + Staff Interviews Infrastructure and security risk management – section 5 of ISO 16363 5.1 Technical Infrastructure and risk management	PTAB
09:45	4.4 AIP Preservation		
12:15	Travel back to GPO HQ		
13:00	Lunch Break		
14:00	Digital Object Management – section 4 of ISO 16363 4.5 Information Management, 4.6 Access Management	5.1 Technical Infrastructure and risk management (continued) 5.2 Security risk management	PTAB
16:00	Internal meeting of the team, Finalization of report, etc		
17:00	Closing meeting		PTAB
17:45	Close of Stage 2 Audit		