

From Theory to Practice: The Art and Implementation of Trustworthy Digital Repository Certification

JESSICA TIEMAN, Digital Preservation Librarian, Government Publishing Office

Jessica Tieman holds a Master of Science in Library and Information Science with a Specialization in Data Curation from the University of Illinois at Urbana-Champaign. During her one-year residency as part of the National Digital Stewardship Residency program, sponsored by Library of Congress and IMLS, Jessica conducted an internal assessment of the Government Publishing Office's digital repository, FDsys/govinfo against the ISO 16363:2012 standard. Jessica has lead GPO's progress toward formal certification under the ISO standard by a third party auditor for over 3 years.

DAVID WALLS, Preservation Librarian, Government Publishing Office

David Walls is the Preservation Librarian for the US Government Publishing office, where he is responsible for developing and managing strategic initiatives for the preservation and life-cycle management of Government information. He is also a member of the FedLink Advisory Board to the Library of Congress and a founding member of the Federal Web Archive Working Group. Prior to coming to GPO, he was the Preservation Librarian at Yale University for twelve years. David has an MLIS with Endorsement of Specialization in Preservation and Conservation Studies from the University of Texas at Austin.

This short paper shall explore GPO's insight into the audit process as a digital repository currently in the process of becoming ISO 16363:2012 certified and express the ways in which institutions reasonably capture and maintain digital preservation environments according to accepted standards and how it is possible to measure success.

iPRES 2018 Conference Themes: • **Mapping out sustainable digital preservation approaches and communities**; *how much can we reasonably capture, maintain according to accepted standards, and scale up?*; • **What business and economic models can facilitate digital preservation feasibility? How can success be documented and/or measured?** • **How have you implemented digital preservation strategies?**

KEYWORDS

Digital repositories, certification, assessment, digital stewardship, standards, best practices

Abstract

In January 2015, GPO initiated a project to assess FDsys/**govinfo** against the ISO 16363 Trustworthy Digital Repository standard. As part of the assessment, GPO gathered evidence, documented procedures and policies, and confirmed adherence to best practices for digital preservation and sustainability for FDsys/**govinfo**. GPO became a participant in the National Digital Stewardship Residency program sponsored by Library of Congress and IMLS to perform an internal audit of FDsys/**govinfo** and explore the viability of securing certification under the ISO 16363 standard after an ISO 16919 accredited certification body became available.

With the relative youth of the ISO 16919 standard, there have been many uncertainties across the digital preservation community about the viability of ISO 16363 certification processes, the availability of auditors, and even the objective value of obtaining such certification under the ISO standard. GPO has embraced a necessary artfulness in preparing for such a recently developed audit process. As the first Federal institution to pursue ISO certification, GPO has encountered the challenges of the ambiguity surrounding ISO certification with innovative strategies to both stay entirely objective about its practices, and meet the highly intricate needs across the spectrum of its Designated Communities. In this manner, preparing for an ISO 16363 certification, an objective, formal process of evaluation, theoretically should not require an interpretive process; however, we will present a realistic overview of how thinking creatively while working closely with a standard can produce authoritative and trustworthy practices in order to define and meet the needs of the repository's Designated Community.

From Theory to Practice: The Art and Implementation of Trustworthy Digital Repository Certification

The US Government Publishing Office, GPO, is working to become the first Federal agency to be named as a Trustworthy Digital Repository for Government information through the certification of **govinfo.gov**, **GPO's digital repository** under ISO 16363:2012 *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*¹. Certification of **govinfo** from an accredited certifying body shall serve to validate GPO's commitment to standards-based digital preservation practices and activities across the 109 criteria of Organizational Governance, Digital Object Management, and Infrastructure and Security Management of the ISO standard. Trustworthy Digital Repository certification has been a key GPO strategic initiative and a joint effort of GPO's Library Services and Content Management (LSCM) and Programs, Strategy, and Technology (PST) business units since 2015. Certification under ISO 16363 will provide GPO stakeholders, including the Federal Depository Library Program, FDLP community, that GPO's only repository, **govinfo** is a standards-compliant digital archive in which Government information is preserved, accessible, and will continue to be usable well into the future.

With the relative youth of the ISO 16363 and ISO 16919 standards, there have been many uncertainties across the digital preservation community about the viability of ISO 16363 certification processes, the availability of auditors, and even the objective value of obtaining such certification under the ISO standard²³. Across the digital preservation and library communities, professional interest groups across associations such as the Research Data Alliance⁴, the National Digital Stewardship Alliance and Digital Library Federation⁵, and federal organizations, some of which GPO is a member of, such as CENDI⁶, have established interest groups dedicated to focusing on the topic of digital repository standards and assessment practices. Currently, six academic and cultural institutions have been assessed by the Center for Research Libraries (CRL), under the *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)*⁷, only six US-based repositories have been awarded the Data Seal of Approval⁸, and many academic, cultural, and federal institutions have completed or embarked on self assessment and internal audits against the *NDSA Levels of Digital Preservation*⁹, the TRAC checklist, risk registries such as Digital Curation Centre (DCC)'s *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)*¹⁰, and the newly developed RDA/World Data System CoreTrustSeal¹¹ certification process. However, the professional community

¹ (International Organization for Standardization)

² (National Institute of Health)

³ (Tieman)

⁴ (Research Data Alliance)

⁵ (National Digital Stewardship Alliance)

⁶ (CENDI Federal Scientific and Technical Information Managers Group)

⁷ (Center for Research Libraries)

⁸ (Data Seal of Approval General Assembly)

⁹ (National Digital Stewardship Alliance)

¹⁰ (Digital Curation Centre)

¹¹ (CoreTrustSeal)

collectively acknowledges the ISO 16363 standard¹² to be the “highest level” of internationally recognized repository certification¹³. With this in mind, the certification process to accredit certification bodies performing the ISO 16363 audits has been published for only four years and much uncertainty still exists about how repositories can and should prepare for such audits, who will perform the audits, and how a formal certification process can lead to influential outcomes with the repository’s Designated Community and the digital preservation community.

At the onset of preparing for ISO 16363 certification, GPO became aware that the existing language within the standard itself, and the type of institutions which had completed TRAC¹⁴ audits, a predecessor to ISO 16363, created a sense of ambiguity surrounding how GPO, a federal institution, could be evaluated against such criteria. Because of this uncertainty, and because no other institutions had direct experience with this auditing process, GPO recognized that there would be some risks for a federal institution to make formal ISO 16363 certification a top strategic priority for the agency. Despite this, in January 2015, GPO initiated a selfassessment project for **govinfo** against the ISO 16363 standard internally. As part of the assessment, GPO gathered evidence, documented procedures and policies, and confirmed adherence to best practices for digital preservation and sustainability for **govinfo**. GPO became a participant in the National Digital Stewardship Residency program sponsored by Library of Congress and the Institute of Museum and Library Studies, IMLS, to perform an internal audit of **govinfo** and explore the viability of securing certification under the ISO 16363 standard after an ISO 16919 accredited certification body became available.

From its conception in 2004, the Federal Digital System (FDsys), now, **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 material currently in the custody of GPO and Federal depository libraries. In the **govinfo** *Concepts of Operations (ConOps) v2.0*, **govinfo** was to be developed according to standards and national guidelines accepted by the digital preservation and library community at the time, including OAIS framework compliance. Because of this, GPO was eager to pursue the external certification in order to validate that the design and mission of its original conception did truly actualize into trustworthy digital repository architecture, design, and systems. Preparing for an ISO 16363 certification, an objective, formal process of evaluation, theoretically should not require an interpretive process; however, we will present a realistic overview of how thinking creatively while working closely with a standard can produce authoritative and trustworthy practices in order to define and meet the needs of the repository’s Designated User Community.

GPO’s first step in preparing for the ISO 16363 audit was creating a master list of every type of documented practice, procedure, workflow, and policy necessary to meet the requirements of the 109 criteria for the standard. This resulted in a curated collection of over hundreds of policies, procedures, data management documents, organizational workflows and staffing plans, financial planning and budget documents, infrastructure and security risk management documents, backup plans and processes, and evidence of audit logs, fixity checks, system change configuration management, and so on. Additionally, GPO also participated in the ISO 16363 High-Level Learning Course instructed by the organization the Primary Trustworthy Authorization Body, PTAB, a group whose membership includes authors of the ISO 16363 standard itself. Another component of the preparation for an external audit included informal interviews with all six of the TRAC-certified repository managers.

¹² (International Organization for Standardization)

¹³ (CoreTrustSeal)

¹⁴ (Center for Research Libraries)

From the feedback gathered during these interviews, GPO produced an internal white paper emphasizing that one could not understate the significance of the Designated Community when preparing for this audit. The ISO 16363 standard language iterates that a “trustworthy digital repository” must start with a mission to provide reliable, long-term access to managed digital resources to its Designated Community, now and into the future¹⁵.” Necessarily, defining the Designated Community is essential in order to effectively, and objectively, evaluate if the repository is fulfilling its mission and if it is truly meeting the needs and expectations of its Designated Communities in ways that are both verifiably transparent could be validated through evidence. This was an eye-opening part of the process of preparing for the audit for GPO, as prior to this realization; many staff were primarily concerned about the technologies of the repository meeting the standard - a natural concern to have, as the overwhelming majority of the criteria in the standard relate to digital object management and infrastructure. This finding now led GPO to focus most of its attention to the metrics regarding the repository’s organizational infrastructure and governance, including thorough documentation of the repository’s Designated Community.

This new focus on organizational infrastructure motivated GPO to examine its policy framework regarding the repository as a first step. GPO’s digital repository falls under the statutory authority of the Superintendents of Documents (SuDoc) programs. Under SuDoc policies, a detailed collection development plan for the repository needed to be developed within the context of GPO’s National Plan for Access to U.S. Government Information¹⁶, GPO’s strategic long-term initiative to expand access to Government information. Outlined in the National Plan was GPO’s strategic commitments to 1) continuously develop and evolve the Federal Depository Library program 2) continue to enhance the statutory Cataloging and Indexing Program to support a National Bibliography of government publications 3) create a strategic initiative known as the Federal Information Preservation Network (FIPNet) in order to work collaboratively through partnerships with agencies and libraries to preserve digitally reformatted and born-digital content across the FDLP and 4) develop a Preservation Program within GPO’s Library Services and Content Management business unit.

These strategic initiatives, and the early development stages of the collection development plan, required GPO to take a closer look at the repository’s definitions of its Designated Community. Though the mission of GPO is “*keeping America informed*,” we recognized that the “General Public At-Large” is too broad a definition of a user community to be able to usefully apply the ISO 16363 standard in a practical way. It is in this manner that the theoretical concepts of the ISO 16363 standard to contribute to more discussions of where policy and procedure could more effectively ensure that our Designated Community’s needs were the driver of this audit process. This ultimately resulted in the release of GPO’s *System of Online Access Collection Development Plan*¹⁷. Within the *Collection Development Plan*. GPO formally documents its Designated Community as such:

“GPO’s system of online access is accessible to anyone around the globe with Internet access. The mission of GPO, however, is Keeping America Informed. As such, system content is driven by the needs of those who reside in the United States or are Americans working or stationed abroad. Designated Communities have been identified for the system.

Per ISO 14721, which defines the OAIS (open archival information system) reference model, a Designated Community is “an identified group of potential Consumers who

¹⁵ (International Organization for Standardization)

¹⁶ (Government Publishing Office)

¹⁷ (Government Publishing Office)

should be able to understand a particular set of information.” The Designated Community for the system includes staff members in Federal depository libraries, the United States Senate, the House of Representatives, the Administrative Office of the United States Courts, and the Office of the Federal Register. Members of the Designated Community are familiar with the organizations, documents, publications, and processes of the legislative, executive, and judicial branches of the United States Federal Government. The Designated Community is able to access content information from the system and render it electronically. The Designated Communities are a representative subset of the broader groups who provide content to and consume information from the system of online access including but not limited to the stakeholder groups listed below:

- *Academic (education/research)*
- *Data Consumer*
- *Federal depository, and non-depository libraries*
- *Government (federal, state, and local)*
- *Interested Citizen*
- *Internal GPO*
- *Legal News Media*
- *Non-profits*
- *Private Industry*
- *Transparency Organizations*

The stakeholder communities have diverse needs; varying levels of knowledge of the legislative and regulatory processes, resources, or subject of their search; and different approaches and skills to finding Government information. GPO provides extensive introductory, advanced, and specialized training in the use of the system through multiple channels for its communities of users. GPO also conducts user surveys and focus groups to ensure user needs are being met.”

You will note that this definition contains the language, “the stakeholder communities have diverse needs; varying levels of knowledge...and different approaches.” The number one challenge GPO faces as a digital repository is that for the more than forty different specific collections of content the repository makes available to the public, GPO is not always the content producer; rather, GPO has varying levels of influence on the file formats, content, supporting metadata, and methods of displaying content based on where the content originates prior to submission to the **govinfo** repository. GPO has made the determination that not only are the content consumers a primary Designated Community, but so are the content producers, many of whom, find that timeliness of ingest and flexibility in content type and discoverability are of high priority. In this way, GPO believed it was appropriate and necessary to document its Designated Community in a way which was unique and original.

In order to document the needs of the Designated Community, GPO has made the decision to record detailed information for each of its content collections, which are primarily designated by which body produced the publication and their topic matter (e.g. Congressional Hearings, Economic Report of the President) within the Data Management Documentation (DMD) for each collection. In this way, the repository defines a Primary (typically, the content Producer), a Secondary (often, the FDLP and Federal government institutions), and a Designated Community at-large (public library users, the media, non-profit organizations). In this way, the repository is able to document and meet the specific needs of the content

producer, as they defined in cooperation with GPO in order to make efficient ingest and appropriate discoverability possible, and also document the minimum requirements of the content's understandability, renderability, and significant properties to ensure the files can be of utility to the Designated Community at-large. In order to document the needs of the Designated Community with even more detail, in addition to GPO's public *System of Online Access Collection Development Plan*, GPO also maintains Data Management Documents (DMDs) per each collection of content per each content contributor, in most cases a federal agency or Congressional author. Within each DMD, GPO thoroughly documents as much information as possible about the provenance of the content, including how it is produced by its author if such information is made available, and what file formats and significant properties must be retained, at the expectation of the content creator, to meet the primary Designated Community of that content.

GPO consistently works closely with its content producers to develop an understanding of what kinds of users are most frequently using this content and in what medium or ways they are likely to search for and use the content. Because each type of content is different, GPO infrequently implements "one-size-fits-all" approach to how content is accessed or discovered within the repository, and in order to approach this situation with accurate documentation, GPO defines a primary, a secondary, and an "at-large" Designated Community for each collection. The needs of these users allow GPO to document decisions regarding why certain information may be displayed in one way versus another, and to what degree GPO has the flexibility to prioritize the look and feel of content within the repository in order to meet the needs of both the content producer and the public users, both of which are part of the Designated Community. This allows the repository to maintain transparency with its content producers, partnerships, and prioritize and make decisions about digital content when its population of users is so broad and diverse. GPO sees strategic solutions such as these a necessary way for repositories to reimagine their documentation and processes against the ISO 16363 standard when considering the repository's uniqueness in the context of standardized best practices.

With the agility to document and better define roles, responsibilities, and commitments regarding its digital repository and collections, GPO ultimately performed and completed an internal assessment against all of the criteria in June 2016. The challenges and opportunities for GPO from that internal assessment were shared with executive management and the FDLP community and as GPO continued to respond to those identified areas of opportunity, GPO also released a Request for Information (RFI) in August 2016 to identify the availability of potential ISO 16363 certification bodies.

After receiving responses to the RFI, in October 2017, GPO released a solicitation in order to procure an external certification body to perform the formal audit. In January 2018, GPO awarded a contract to PTAB – Primary Trustworthy Authorisation Body to perform a formal audit in order for GPO to receive ISO 16363:2012 Certification. The external audit of GPO's digital repository will include a phased process of initial assessment, opportunities to respond to the initial assessment, and a final assessment by the auditor.

GPO anticipates the formal audit will be completed prior to the end of FY 2020. Ongoing surveillance audits will be necessary for GPO to maintain status as a certified repository. GPO will initiate the first stage of the audit in February/March 2018. This stage includes the evaluation of GPO's self-assessment by an auditing team in order to identify potential areas of concern or nonconformities; GPO will have the opportunity to respond to these nonconformities prior to Stage 2. In Stage 2, GPO will undergo further evaluation, including onsite visits from auditors. GPO will have a set number of months to address any existing concerns from the auditor or nonconformities; if GPO successfully responds to all areas of concern, the auditing team recommends that GPO receive certification.

Sources

- CENDI Federal Scientific and Technical Information Managers Group. *CENDI*. 2018. Web. 12 April 2018. <<https://www.cendi.gov/index.html>>.
- Center for Research Libraries. *TRAC Metrics*. 2018. Web. 12 April 2018. <<https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>>.
- . *TRAC Metrics*. 2018. Web. 30 May 2018. <<https://www.crl.edu/archiving-preservation/digitalarchives/metrics-assessing-and-certifying/trac>>.
- CoreTrustSeal. "An Introduction to the Core Trustworthy Data Repositories Requirements." November 2016. Document. 12 April 2018. <https://www.coretrustseal.org/wpcontent/uploads/2017/01/Intro_To_Core_Trustworthy_Data_Repositories_Requirements_2016-11.pdf>.
- . *CoreTrustSeal*. 2018. Web. 12 April 2018. <<https://www.coretrustseal.org/>>.
- Data Seal of Approval General Assembly. *Data Seal of Approval*. 2018. Web. 12 April 2018. <<https://www.datasealofapproval.org/en/community/>>.
- Digital Curation Centre. *Digital Repository Audit Method Based on Risk Assessment*. 6 August 2015. Web. 12 April 2018. <<http://www.repositoryaudit.eu/>>.
- Government Publishing Office. "GPO's System of Online Access Collection Development Plan." 30 September 2016. *Federal Depository Library Program*. Document. 12 April 2018. <<https://www.fdlp.gov/file-repository/about-the-fdlp/gpo-projects/trustworthy-digitalreports/2812-gpo-s-system-of-online-access-collection-development-plan>>.
- . "National Plan for Access to U.S. Government Information: A Framework for a User-Centric Service Approach to Permanent Public Access." February 2016. *Federal Depository Library Program*. Document. 12 April 2018. <<https://www.fdlp.gov/file-repository/aboutthe-fdlp/gpo-projects/national-plan-for-access-to-u-s-government-information/2700national-plan-for-access-to-u-s-government-information-a-framework-for-a-user-centric-service-approach-to-permanent-public-access>>.
- International Organization for Standardization. *ISO 16363:2012 (CCSDS 652.0-R-1) Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*. 10 July 2017. Web. 12 April 2018. <<https://www.iso.org/standard/56510.html>>.
- National Digital Stewardship Alliance. *Levels of Digital Preservation*. April 2013. Web. 12 April 2018. <<http://ndsa.org/activities/levels-of-digital-preservation/>>.
- . *Standards and Practices Interest Group*. 2018. Web. 12 April 2018. <<http://ndsa.org/working-groups/standards-and-practices/>>.
- National Institute of Health. "A Primer on the Certifications of a Trusted Digital Repository (TDR)." 20 April 2017. *DataScience@NIH*. Web. 12 April 2018. <https://datascience.nih.gov/Trusted_Digital_Repository>.

Research Data Alliance. *RDA/WDS Certification of Digital Repositories IG*. 11 September 2017. Web. 12 April 2018. <<https://www.rd-alliance.org/groups/rdawds-certification-digitalrepositories-ig.html>>.

Tieman, Jessica. "Future for Trust in Digital Repositories." 1 December 2015. *The Future of Information Alliance*. Blog. 12 April 2018. <<https://fia.umd.edu/future-for-trust-in-digitalrepositories/>>.