The Self-Audit of GPO's Federal Digital System, FDsys

David Walls, United States Government Printing Office (USA)

Introduction

On December 20, 2010, the United States Government Printing Office released version 1 of the Federal Digital System, also known as FDsys. During fall and winter 2010, the Government Printing Office conducted an internal self-audit of FDsys utilizing the metrics of the Trustworthy Repository Audit and Checklist or TRAC, established by the Center for Research Libraries or CRL. FDsys is the first Federal Digital repository to be self-audited according to TRAC. The findings of the internal self-audit team are currently under review by GPO. GPO is also currently planning for an independent audit of FDsys. However, since GPO is part of the legislative branch of the US Government, this audit project is dependent on approval and funding by the Joint Committee on Printing, a subcommittee of the United States Congress. Rather than wait for the completion of the independent audit before sharing our experiences, we felt that our audit was productive and positive and that we should take the opportunity to tell what we've learned so far.

Most information produced by the US Federal Government today is digital. Digital versions of the Federal Register, the Public Papers of the President, and other titles have taken precedence over the paper editions long familiar to government information researchers. Federal agencies from the EPA, NASA, to the Treasury Department, disseminate most of their information to the American public over the World Wide Web. Much of this web-based public information changes through a process of continual updating. Old information is rapidly superseded and websites are frequently taken down, requiring web harvesting to archive web content. In addition to born-digital information, the US Government has been distributing paper-based publications to libraries through the Federal Depository Library Program, FDLP, since 1813. Some FDLP libraries are digitizing rare and endangered collections of government documents, creating a third tier of digital government information.

The US Government Printing Office is the primary information distribution and dissemination agency for the US Federal Government. GPO's mission is to provide permanent access to US Government information for the American People. The objective to provide permanent access to an ever widening scope of digital content requires a robust content management system, preservation repository, and a search engine capable of responding to the queries of a user community as varied as the American public.

The Federal Digital System, FDsys, is GPO's digital repository and asset management system. Each of the varied user communities from the US Congress, Federal publishers, and the American public will rely on it to serve their information needs.

Due to federal agency audit and reporting requirements, the design and development of FDsys is thoroughly documented.

FDsys was pronounced one of the 10 best Federal digital projects.[1] The staff of GPO is excited about FDsys. Yet, we recognize that our own enthusiasm and conviction are not appropriate metrics for fulfilling the level of trust GPO needs to provide its users. GPO's mandate to meet the information needs of these important and varying user communities requires GPO to provide a high level of assurance that FDsys is a trustworthy digital repository; one capable of the long-term preservation of information assets.

An audit is an evaluation. Certification is confirmation of the achievement of certain standards.[2] The Center for Research Libraries, CRL, is a not-for-profit international consortium of college and research libraries. In 2003, the Research Libraries Group and the National Archives and Records Administration, NARA, created a joint task force to develop criteria to identify digital repositories capable of reliably storing, migrating, and providing long-term access to digital collections. The U.S. based Center for Research Libraries, a 501 (c) (3) status educational support organization contributed to the further development of certification criteria through their Auditing and Certification of Digital Archiving Project. The Center for Research Libraries finalized the Trustworthy Repositories Audit and Certification Criteria and Checklist, publishing v. 1.0 of TRAC in February 2007.[3]

These metrics were developed by CRL based upon wide-spread research into generally accepted best practices in the management of digital repositories. CRL requires the repository to provide evidence to demonstrate the repository's commitment to a supportive organizational infrastructure, digital object management, and technologies infrastructure and security. The TRAC checklist criteria consist of 84 metrics that must be answered or defended by citing supporting documentation and describing how that documentation supports the metrics. The resulting body of documentary evidence builds a case toward determining whether a digital repository meets trustworthy status.

The TRAC checklist criteria represent the organizational and technical infrastructure required for a digital repository to be considered trustworthy and capable of certification. It establishes a baseline definition of what a trustworthy digital repository is and lays out the components that must be considered and evaluated as a part of that determination.[4] It is important to remember that while the OAIS model provides the basis for a secure and trustworthy digital repository, TRAC evaluates much more. In addition to the repository's architectural attributes, TRAC also assesses an organization's ability to function as an effective manager of that digital repository.

Audit Methodology

There are three essential attributes for a successful self audit. First, the audit must be an established priority for an

organization. The audit must be scheduled according to the annual fiscal and project planning that an organization completes for any given fiscal year. Second, the audit should be completed by a designated audit team of staff who are professionally knowledgeable about the repository and digital preservation. Third, agency heads and managing directors have to publically sanction and support the audit team. Since the documentation needed to complete the audit may not be immediately available to the audit team, the support of managers may be needed to have staff and agency contractors comply with requests for information, as well as requests for meetings to explain processes and fill in gaps in requested documentation.

The Preservation Librarian of GPO was assigned to be the lead on the self-audit and formed an audit team with GPO's Lead Program Planner for Project Management, the internal agency group responsible for the planning and development of FDsys. Other GPO staff from the Project Management Office and Library Services and Content Management assisted throughout the audit process.

For the audit, the audit team set up regular working meetings. To provide a framework for our documentation, each of the 84 TRAC criteria were copied from the TRAC checklist directly into an Excel spreadsheet. A column in the spreadsheet was created to record the specific documentation that would be used to support a particular TRAC statement. Documentation was reviewed against the description for each metric of TRAC following the examples of supporting documentation found in the text of the TRAC Checklist. Throughout the audit, the audit team met weekly and occasionally twice weekly to review documentation and discuss our conclusions for each TRAC metric.

Documentation

Fortunately for the audit team, there was extensive documentation for FDsys. As a Federal government project, which had to meet regular reporting requirements as well as regular audits by the Inspector General's office, there were extensive reports describing FDsys' concept of operation and repository design as well as plans to meet continuity of access and continuity of operations. In addition to these reports, the audit team reviewed staff organizational charts, financial statements, staff training documentation, and Inspector General Reports. Little if any documentation needed to be created for the audit. However, it should be noted that this is not typical, at least not among organizations the audit team talked with.

Audit Process

To learn more about the audit process, the audit team contacted staff from Portico and MetaArchive, two digital repositories that recently completed their own audits. The audit team also set up several phone conferences with lead audit staff at Portico to learn more about their experiences. Portico recently completed an independent audit by an audit team from the Center for Research Libraries. Since GPO's plan is to complete a similar audit, Portico generously shared examples of their own documentation.

In addition to the 84 metrics, the TRAC checklist includes ranking criteria consisting of five statements that describe a

repository's score for a particular metric. The ranking ranges from 5, "compliant with all metrics fully and consistently" to 1, "compliant with all critical metrics, with a minimum of inconsistencies or deficiencies in areas that might lead to minor defects of a systematic nature."[5]

After consulting with staff at Portico and MetaArchive, the FDsys audit team decided to follow current practice for the self-audit of not providing a numerically specific ranking for each TRAC metric of the internal audit. Both MetaArchive and Portico chose to assign an "in compliance" or "not in compliance" ranking to each metric. This seems especially preferable when planning for an independent audit, as the independent auditors will assign a numerical ranking. The FDsys audit team felt that discrepancies in the scores could potentially undermine the credibility of both the self and independent audit processes. In the case of both a self-audit and independent audit, the interpretation of the documentation and whether TRAC metrics are supported is essentially subjective. It is unlikely that two groups of auditors would rank the same repository in exactly the same way.

The FDsys audit team chose to follow this same course, providing an in compliance score if we felt that FDsys fully met the metric or a not in compliance score if we felt there were metrics where FDsys would likely rank a score of less than 4 or 5 in the TRAC ranking. The decision to point out areas where future growth and development were needed supports one of the original purposes of TRAC, that "digital preservation is an ongoing mission of a digital repository, requiring constant monitoring, planning, and maintenance as well as conscious actions and strategy implementation."[6]

It is important to recognize that the five ranking criteria for each metric are designed to point out strengths as well as areas where additional growth and development are needed. A score of 1 documents areas where future growth is needed and is not designed in process or wording to be punitive.

It is also especially important to stress the supportive features of TRAC when managing the expectation of your user community, management, faculty, constituents; whoever has a stake in the success of the repository. It is pointless to get stressed over why your repository was ranked only a 3 in one metric when everyone was hoping for a 5. It is productive however, to see a score of 3 in a particular metric as pointing out areas for improvement that would result in the better management of digital assets.

Conclusions

After carefully reviewing our documentation, the FDsys self-audit team concluded that FDsys was in compliance with most TRAC metrics. Most of the areas where FDsys is not fully in compliance are concerned with FDsys business planning and cost modeling. Communicating repository services and their costs to users is an important aspect of the working life of a digital repository. Cost modeling for FDsys has proven to be complicated and we are still working on developing internal cost models that show real operational costs as well as modeling cost options for user services. One metric in particular is concerned with designating a successor for the repository in the event that a repository's parent organization ceases to exist. GPO has an

agreement with The National Archives and Records Administration, NARA, to archive GPO publications. However, on reviewing the Memorandum of Understanding between the two agencies, it was clear that the agreement didn't cover taking over the operation of FDsys.

The final conclusion of the FDsys audit team is that the self-audit of FDsys is a time to celebrate the significant work that has gone into creating a trustworthy digital repository. With the creation of FDsys, GPO begins an active commitment toward the long-term preservation of the digital information assets placed in trust for the American people. The self-audit identified proven areas of success, while also documenting areas where future development is needed to ensure the long-term viability and successful growth of FDsys.

---

[1] *Government Computer News*, July 27, 2009

[2] Wikipedia contributors, "Certification," *Wikipedia, The Free Encyclopedia,* www.wikipedia.org

[3] http://www.crl.edu/sites/default/files/attachments/pages/trac_pdf

[4] TRAC Document p.2, 3

[5] TRAC Document p.5

[6] TRAC Document p.3