

# Auditing a Dark Archive

Alex Kinnaman, Michael Popham

## Abstract

As digital collection-building projects have become a focus for libraries in the digital age, there is a noted lack of uniformity in their method of construction. The standards for digital preservation are just recently beginning to formalize, but this only extends to completed, established projects. In a case study at the University of Oxford, the Trustworthy Repository Audit and Certification checklist is applied to the project *Digital Safe* to determine if existing digital auditing metrics can act as a scoping exercise for developing digital projects. *Digital Safe* is an incomplete dark archive, making it a unique subject to audit. This case study focused on how TRAC aligned with the goals of the project and if it highlighted new aspects of concentration. This approach to auditing an incomplete digital project is challenging because of the lack of material and cohesion of the dark archive, which could be a useful tool for those thinking about or currently developing a digital collection.

**Keywords:** TRAC; dark archives; digital repository audits

**Citation:** Kinnaman, A., & Popham, M. (2017). Auditing a Dark Archive. In *iConference 2017 Proceedings* (pp. 839-842). <https://doi.org/10.9776/17351>

**Copyright:** Copyright is held by the authors.

**Contact:** [kinnama2@illinois.edu](mailto:kinnama2@illinois.edu)

## 1 Introduction

As digital collection projects continue to proliferate, there is a noted lack of uniformity in how they are constructed. Standards do not yet exist in the development stages for such digital projects, highlighting a need for an investigation of current digital assessment tools to determine if and how existing tools help guide and standardize digital project development. The primary research question is: can standard auditing metrics be used as a scoping exercise for designing requirements for a dark archive?

## 2 Case Study

### 2.1 Context

The context of this research is centered around how best to conduct a review of the *Digital Safe* project prior to seeking next-phase funding. *Digital Safe* is the University of Oxford's solution for a collective infrastructure and technology for storing long-term, high-security data. Student records, financial records, political materials, and other personally identifiable data need to be digitally preserved and stored, but with elevated security requirements. *Digital Safe* is the Bodleian Digital Library Systems & Services' (BDLSS) solution to this issue. Due to the high-security content, *Digital Safe* is classified as a dark archive meant for limited access storage. They opted to outsource the technologies for storage and digital preservation and management via a local interface. In 2014, the project went on hiatus after completing two phases of its planned 3-phase project. In summer of 2016 the project began seeking funding for phase 3, requiring a review of the project (N. Jefferies, personal communication, July 19, 2016).

### 2.2 Difficulties in Reviewing *Digital Safe*

The BDLSS opted to outsource each technology for *Digital Safe*, and chose Arkivum for long-term, high-security storage (Arkivum, 2014 & 2016), Archivematica (2015) for a customizable digital preservation workflow tool that functions within Arkivum (Stanbridge, 2016), and their own interface design for local management (Jefferies & Tomkins, 2014). This means that all three components require review.

Furthermore, *Digital Safe* does not yet exist, so there is no complete integration of these components, no content, no formalized policies, and therefore limited user feedback. Ultimately, there is also no standard metric for auditing dark archives, or for auditing dark archives that are incomplete.

### 3 Methodology

#### 3.1 Choosing a Metric

Choosing a digital repository auditing metric began with a review of The Center for Research Libraries Assessment Tools (CRL, 2007), followed by consultations. Consultations included: Michael Popham, Head of Digital Collections & Preservation Services for the BDLSS; Neil Jefferies, R&D Manager for the BDLSS and *Digital Safe* Project Manager; and Rhiannon Bettivia, Postdoctoral Research Associate, specializing in digital preservation, at the School of Information Science at the University of Illinois. The rejected metrics and rationale are as follows:

- The Ten Principles (CRL, 2007), for its intended simplicity;
- Digital Seal of Approval (2013), for not have thorough enough standards to handle three different components;
- Data Asset Framework (Jones, et al., 2008), because it focuses on research data assets which is unrelated content;
- Data Asset Assessment Tool project (Pinsent & Ashley, 2006), because it is outdated and not a universal standard;
- ISO16363 (2012), because it is too complex for the time frame of this initial experiment;
- Digital Repository Audit Method Based on Risk Assessment (DRAMBORA, 2008), because it requires content and use of the service for measurement of risk.

The Trustworthy Repository Audit and Certification (TRAC, 2007) was chosen as the accepted metric for its universal use, balance of complexity, and preference of those involved in the *Digital Safe* project.

##### 3.1.1 Review of TRAC

TRAC was established in 2007 and is hosted by OCLC (2016). It is also the foundation for ISO16363. It is an 84-question checklist with criteria that center around three primary sections measuring Organizational Infrastructure, Digital Object Management, and Technologies, Technical Infrastructure, & Security.

##### 3.1.2 Limitations of TRAC

TRAC is primarily applied to repositories, which by definition are open access, and also focuses on accessibility and transparency. *Digital Safe* is a dark archive, which significantly alters how the metric handles accessibility and transparency. Additionally, auditing an approach to archive development rather than the archive itself raises challenges in measuring how the content is handled.

#### 3.2 Conducting the Informal Audit

The informal review of *Digital Safe* was carried out over a 3.5-week period. The resources used to evaluate *Digital Safe* included the technologies documentation and user guides, project reports, and personal interviews of project personnel.

#### 3.3 Determining Results

The results of this informal audit were reliant on how well a TRAC criterion can be met by using the existing documentation. A criterion might be considered met in one of two ways: *Digital Safe* actually completes the criterion, or if it has a plan of action in place for eventually completing the criterion. A

TRAC criterion is not met if it is non-existent in the resources or if there is no indication of planning for its fulfillment

## 4 Results

### 4.1 Informal Audit Results

The project culminated into three preliminary results. First, using a standard auditing metric on an incomplete dark archive does not work successfully as a formal audit. A true audit requires a complete and working service, content, and user experience and feedback in order to completely understand if the service is meeting the needs of its users and providers. Secondly, despite the difficulty in adapting the TRAC audit to *Digital Safe*, this approach can successfully provide insight and evaluation to a project in progress. The TRAC criteria identified areas of strengths and weaknesses of the current state of the project and its work plan, and offered evidence to support continued funding for *Digital Safe*. Finally, using the TRAC standard on an in-progress project was successful in acting as a guideline for developing requirements for a dark archive. The criteria responses highlighted areas that reflected *Digital Safe's* priorities while also providing a means for focusing on those priorities.

### 4.2 Limitations

These results are preliminary due to several limitations. *Digital Safe* is an archive and this method has not been tested on other in-progress repositories, digital libraries, or similar digital collections. *Digital Safe* is also a dark archive, which narrows its relatability to other digital collection management projects. Furthermore, there may also be discrepancies as to what an auditor considers fully meeting criteria. Other metrics were also not tested on *Digital Safe*, leaving room for additional analysis. Finally, this project was completed in a short amount of time. A second review of *Digital Safe*, when it is completed, is necessary to determine if the previously identified areas of strengths and weaknesses were improved upon, and if the initial review was a factor in that improvement. Another evaluation metric should also be tested on *Digital Safe* to see if the results are similar, and identify any discrepancies. All of these current limitations are possible endeavors for future work.

## 5 References

- Archivematica (2015). *Archivematica 1.5*. Retrieved from <https://www.archivematica.org/en/docs/archivematica-1.5>
- Arkivum Limited (2016). Arkivum. *Arkivum Limited*. Retrieved from <http://arkivum.com/>
- Arkivum Limited (2014). Frequently asked questions – Arkivum. *Arkivum Limited*, 2.2. Retrieved from <http://arkivum.com/wp-content/uploads/2014/04/Arkivum-Frequently-Asked-Questions.pdf>
- Center for Research Libraries (n.d.). Digital Preservation Metrics. *Center for Research Libraries*. Retrieved from <https://www.crl.edu/archiving-preservation/digital-archives/metrics>
- Center for Research Libraries (2007). Ten Principles. *Center for Research Libraries*. Retrieved from <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re>
- Digital Repository Audit Method Based on Risk Assessment (2008). DRAMBORA: About. *Digital Curation Centre*. Retrieved from <http://www.repositoryaudit.eu/about/>
- ISO (2012). ISO 16363:2012: Space data and information transfer systems – Audit and certification of trustworthy digital repositories. *International Organization for Standardization*. Retrieved from [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=56510](http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510)

- Jones, S., Ross, S., and Ruusalepp, R. (2009). Data Audit Framework Methodology (Version 1.8). *Data Asset Framework*. Retrieved from [http://www.data-audit.eu/DAF\\_Methodology.pdf](http://www.data-audit.eu/DAF_Methodology.pdf)
- Data Seal of Approval (2013). Data Seal of Approval Guidelines Version 2. *Data Seal of Approval*. Retrieved from <http://www.citationmachine.net/items/new>
- Jefferies, N., & Tomkins, D. (2014). Digital Safe: Archiving records for the long term. *Bodleian Digital Library*. [PowerPoint slides]. Retrieved from <https://digitalsafe.wordpress.com/2014/08/18/digital-safe-presented-at-the-ictf-conference-10-july-2014/>
- Pinsent, E. & Ashley, K. (2006). Digital Asset Assessment Tool (DAAT) Project Final Report (Version 3.5). *Jisc*. Retrieved from [http://www.webarchive.org.uk/wayback/archive/20140614062318/http://www.jisc.ac.uk/media/documents/programmes/preservation/daat\\_final\\_report\\_2006.pdf](http://www.webarchive.org.uk/wayback/archive/20140614062318/http://www.jisc.ac.uk/media/documents/programmes/preservation/daat_final_report_2006.pdf)
- Stanbridge, N. (2016). New digital preservation solution from Arkivum, shaped to grow with your data. *Arkivum Limited*. Retrieved from <http://arkivum.com/blog/perpetua-digital-preservation/>
- Trusted Repository Audit & Certification (2007). *The Center for Research Libraries*. Retrieved from [https://www.crl.edu/sites/default/files/d6/attachments/pages/trac\\_0.pdf](https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf)