DIGITAL PRESERVATION AUDITING METRICS AS DESIGN TOOLS FOR DIGITAL REPOSITORIES

BY

ALEX KINNAMAN

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Library and Information Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Advisers:

Dean Allen Renear
Professor Rhiannon Bettivia

Abstract

Collaboration via partnership in a consortium and in outsourcing are common aspects in building and maintaining a trusted digital repository. Such collaboration is overlooked in most digital preservation auditing metrics. This not only prevents the possibility of formal certification, but not including third-party participation in the standards implies that there are no standards for negotiating contracts and delineating the roles of partners. This thesis examines the ongoing project Digital Safe, a project in development at Oxford that aims to be a service for storing confidential information. In two case studies, the author employs the Trustworthy Repositories Audit & Certification (TRAC) and the Data Seal of Approval (DSA) to inform the development of Digital Safe and its relationships with third-party vendors. The major goal is to examine how various roles between an institution and third-parties can be delegated based on the necessary standards. This is useful first for helping develop contracts with vendors and understanding exact responsibilities in partnerships. Second, it facilitates a better understanding of the limitations of current auditing metrics.

The case studies reveal that both TRAC and DSA can provide a means for defining roles in partnerships, TRAC being more complex and DSA being more theoretical. Second, the documentation for audit standards is reliant on OAIS reference model, which limits their use in consortia, dark archives, and other specific repositories. The case studies also clarify the type of evidence most appropriate to have and develop in the planning stages for a digital repository. These findings point to future work in a revision of how audit standards are used, specifically indicating their use-value as development tools in addition to assessment tools. The addition of third-party support to these standards could facilitate a better guide to interacting with third parties during planning stages, and ultimately improve digital preservation standards and the trustworthiness of repositories.

Acknowledgements

This work was completed with the assistance of the following organizations and people:

University of Illinois at Urbana-Champaign

The iSchool at Illinois

University of Oxford

Bodleian Libraries

Bodleian Digital Library Systems & Services

Oxford e-Research Center

The Polonsky Digital Preservation Programme

The Center for Research Libraries

The Data Seal of Approval Board

Rhiannon Bettivia, University of Illinois at Urbana-Champaign

Allen Renear, University of Illinois at Urbana-Champaign

Michael Popham, Bodleian Digital Library Systems & Services

Neil Jefferies, Bodleian Digital Library Systems & Services

David Tomkins, Bodleian Digital Library Systems & Services

Susan Thomas, Bodleian Electronic Archives & Manuscripts

David Weigl, Oxford e-Research Center

Pip Wilcox, Oxford e-Research Center

David De Roure, Oxford e-Research

Center Kevin Page, Oxford e-Research Center

Table of Contents

Introduction 1
Literature Review 3
A Review of Digital Preservation Auditing Metrics 3
Environmental Scan 5
Metrics as Assessment Tools 5
Metrics as Development Tools 7
Third Parties in Digital Repositories 8
Outsourcing 8
Consortia 8
Context 11
Project Development 11
Background for Digital Safe 12
Outsourced Technologies 14
Rationale 15
Methodology 17
Evidence Collection 17
The TRAC Assessment 19
The DSA Assessment 20
Results 22
Case Study I: TRAC 22
Case Study II: DSA 26
Limitations 41
Reliance on OAIS Reference Model 41
Documentation & Definitions 42
Recent Updates 44
Additional Metrics 44
Discussion 46
TRAC vs. DSA 46
General Use 46
Use as Development Tools 47
Consortia, Third-Parties, & Metrics 49

Metric Documentation 49
Structure of the Results Section 50
Auditing Parts or the Whole? 51
Use in Building Contracts 52
Documentation & Testimony as Evidence 52
TRAC & the ISO 27000-series 54
Future Work 55
Continuation & Follow-up 55
Inclusion of Outsource Partners 56
Repository versus Service 56
Documentation Revision 57
Conclusion 58
References 59
Appendix A: Acronym Dictionary 64
Appendix B: An Informal TRAC Audit of Digital Safe at the University of Oxford 66
Appendix C: DSA Assessment of Digital Safe – Full Version 127

Introduction

Building a trusted digital repository is no easy task. In a nutshell, it requires juggling governance and infrastructure, assessing needs, prioritizing preservation activities, developing and implementing a preservation plan, and all the while securing the funding and staff to support the project. Many institutions and organizations can provide some but not all of these aspects. It is not surprising, then, that most institutions are outsourcing to third-parties or forming consortia in order to support large digital repository projects. For example, organizations like CLOCKSS[1] contracts-out their operations to Stanford University, and contracts storage spaces at Rice, Indiana, and Stanford Universities (Rosenthal, 2014); the University of Illinois at Urbana-Champaign contracts out file and data storage[2] to Amazon Glacier's cloud storage (Engineering IT) which also acts as support for their Medusa Core[3] repository; the University of Virginia's institutional repository Libra is built on Hydra technology, which is accessible to them as partners in Hydra[4]. Further, organizations like APTrust,[5] Digital Preservation Network (DPN),[6] Hydra,[7] Preservica,[8] and AVPreserve[9] exist to provide a range services for digital preservation activities.

However, the benefit of having partners in digital repositories also hinders any repository aiming to be formally certified as a trusted digital repository. In examining documentation for the digital preservation auditing metrics Trustworthy Repositories Audit & Certification (TRAC) and the Data Seal of Approval (DSA), it is evident that consortia, outsource partners, and other collaborative activities have been largely excluded from these standards. Though the newest 2017-2019 DSA Guidelines now incorporate Outsource Partners as an expectation for current digital repositories, previous versions and additional metrics do not.

---

[1] See CLOCKSS' homepage for additional information: https://www.clockss.org/
[2] Additional contracts held by UIUC Engineering IT Services: https://it.engineering.illinois.edu/services/file-and-data-storage
[3] Additional information on Medusa Core and its contract: http://cms.library.illinois.edu/export/it/helpdesk/service/medusa.html
[4] More information on University of Virginia's partnership and other Hydra partners: https://projecthydra.org/community-2-2/partners-and-more/university-of-virginia-2/
[5] APTrust homepage & mission statement: http://aptrust.org/about
[6] DPN homepage & mission statement: http://dpn.org/
[7] Hyrda's homepage: https://projecthydra.org
[8] Preservica overview: http://preservica.com/preservica-2/
[9] AVPreserve homepage & mission statement: https://www.avpreserve.com/

Digital preservation auditing metrics are built to audit and assess existing digital repositories based on institutionally agreed upon standards for the purpose of certifying trusted digital repositories. They are not built for developing digital repositories or for assessing consortia and vendor partnerships. The first issue is that there are more audit standards than there are development tools, so developers must rely either on collections of preservation planning resources or use audit standards for uses other than their purpose. The second issue is that there are no standards in the digital preservation community for developing contracts between partners and vendors.

Despite their assessment purpose, auditing metrics can serve as scoping tools for designing and negotiating contracts with vendors, and for delegating responsibilities between consortium members. In two case studies that examine the ongoing partnership-based project Digital Safe[10] at the University of Oxford using TRAC Criteria (CRL & OCLC, 2007) and DSA Guidelines (2016), three aspects are explored. First, TRAC and DSA are compared for use as general scoping tools. Second, the metrics are further evaluated for development and evidence collection. Finally, this thesis investigates how audit standards can help guide institutions to work with vendors and partners.

[10] See the Digital Safe blog: Digital Safe website https://digitalsafe.wordpress.com/

Literature Review

A Review of Digital Preservation Auditing Metrics

At its core, the purpose of a standard for trusted digital repositories is to provide evidence that a digital repository is sustainable, that its scope is documented and understood, and that its managerial and technical infrastructure are intact.

Digital preservation auditing metrics provide the structure for assessing and certifying trustworthy digital repositories. Standards such as Audit and Certification of trustworthy digital repositories (TDR/ISO 16363:2012), Trusted Repositories Audit & Certification (TRAC), Data Seal of Approval (DSA), the nestor Seal, and the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) are all tools for maintaining trusted digital repositories. TDR, TRAC, and the DSA are certifiable metrics that can be awarded to digital repositories, and the nestor Seal and DRAMBORA are utilized primarily as self-assessment tools in preparation for a formal audit or internal redesign. These five standards are currently the most utilized of the various audits and tools that exist for measuring digital repositories based on their universality of use and relevance to the digital preservation community. This is also corroborated by lists of certifications provided by digital preservation organizations such as the Inter-university Consortium for Political and Social Research (ICPSR) at Michigan (ICPSR, 2017), the metrics listed by the Center for Research Libraries (CRL), and the "Audit and certification" section in the Digital Preservation Handbook maintained by the Digital Preservation Coalition (DPC, 2017), among other similar organizations.

The two case studies included in this research focus on TRAC and DSA, though the influence of the nestory Seal, TDR, and DRAMBORA is relevant to understanding the chosen metrics for the Digital Safe case studies. The following is a brief overview of the development of these metrics, any related metrics, and the establishment of relevant digital preservation organizations for context:

- **1999**: The Open Archival Information System (OAIS) reference model is developed by the Consultative Committee for Space Data Systems (CCSDS) in the US, and then passed to the International Organization for Standardization (ISO) for approval and control (Lavoie, 2000).
- **2000**: ISO 17799 Information Security Policy is added to ISO from the existing British standard BS 7799. ISO 17799 evolves into ISO 27002.

- **2002**: Digital Preservation Coalition is established in the UK.
- **2003**: OAIS reference model for long-term preservation is officially released by ISO.
- **2004**: The Digital Curation Centre is established in the UK.
- **2005**: ISO 17799-2 becomes ISO 27001 and the ISO 27000 series develops further from ISO 27001 and ISO 27002.
- **2006**: Center for Research Libraries (CRL) is established in the US.
- **2006-2007**: DRAMBORA is developed by the Data Curation Centre (DCC) and DigitalPreservationEurope (DPE) for self-assessed risk management.
- **2007**: TRAC is developed by OCLC and is implemented by CRL auditors.
- **2007**: Ten Principles developed by four digital preservation organizations and published by CRL.
- **2008**: DSA is developed by Data Archiving and Networked Services (DANS) at the Royal Netherlands Academy of Arts and Sciences (DSA About).
- **2009**: Control of DSA is transferred to an international Board of external auditors (DSA About).
- **2012**: ISO 16363 (TDR) is developed as an expanded, better organized version of TRAC and is currently the ultimate certification standard.
- **2012**: DIN 31644 Information and Documentation – Criteria for Trustworthy Digital Archives, which is the expansion of the Ten Principles, is also officially released in Germany.
- **2013**: nestor Seal is developed by the nestor Certification Working Group as verification for DIN 31644 compliance for the purposes of extended certification.
- **2013**: National Digital Stewardship Alliance (NDSA) develops the NDSA Levels of Preservation.
- **2014**: ISO 16919 "Space data and information transfer systems—Requirements for bodies providing audit and certification of candidate trustworthy digital repositories" is created, which is a prerequisite for ISO 16363 certification.

Though the process of digital preservation collaboration and development extends far beyond this timeline, the relevant technologies and organizations still reveal several themes. The incremental building of these tools indicates first, that there are various types of audits that measure different aspects of repositories. DSA and DIN 31644, for example, are primarily for research data and focus on how to preserve for continuous access, whereas DRAMBORA assesses the range of criteria for what level of risk a repository has based on their current technologies and workflows. ISO 16363 and TRAC are the ultimate criteria for digital preservation goals overall, particularly given their relationship to additional ISO certifications, and are therefore the most complex and universally applicable. Second, the timeline

demonstrates how these metrics are built on one another and are therefore still evolving. Finally, though obvious, international collaboration is a major factor in standardizing digital preservation.

Environmental Scan

Metrics as Assessment Tools

Though digital preservation auditing metrics are built to audit repositories, utilizing these metrics for self and peer assessment is not a new concept. Often self-assessments are conducted as preparation for a formal certification, or just for internal use. The purpose of standards is to eventually become universally used as a means of connecting information, streamlining the digital preservation process, and to ensure that data is maintained and usable in the future. Consequently, there are many resources for assisting self-assessments from both digital preservation organizations, and from individual institutions from groups that report to the community on their experiences. Outside of the documentation for the metrics themselves, there are various resources from the major digital preservation institutions that provide basic guides, aides for choosing a metric, guides on how to plan an assessment, and various case studies for the most prevalent auditing metrics. The Digital Preservation Coalition (DPC),[11] the Center for Research Libraries (CRL),[12] the National Digital Stewardship Alliance (NDSA),[13] and the Digital Curation Centre (DCC)[14] are four of the primary organizations for standardized digital preservation practices and guides on using their resources.

Basic tools to guide self-assessments exist both officially and as produced by smaller universities, such as the Ten Principles compiled by the CRL versus the created by the Northeast Document Conservation Center (NDCC).[15] The Ten Principles were officially released in 2007 by the CRL after consulting with three other digital preservation organizations and provide the ten most basic criteria that a digital repository must possess to be a trusted digital repository (2007). Comparatively, organizations like the NDCC have developed short project-structuring documents like "Planning for Digital Preservation: A Self-Assessment Tool" that consists of a four-page list of considerations for developing a repository (2007). More developed resources

[11] DPC: http://www.dpconline.org
[12] Center for Research Libraries: https://www.crl.edu/
[13] National Digital Stewardship Alliance homepage: http://www.digitalpreservation.gov:8081/ndsa/index.html
[14] Data Curation Centre homepage: http://www.dcc.ac.uk/
[15] Northeast Document Conservation Center homepage: https://www.nedcc.org/

are also freely available, such as the Digital Preservation Handbook created and maintained by the DPC. More detailed sources like the POWRR Tool Grid,[16] a comprehensive chart listing digital preservation tools and their attributes against the six aspects of the OAIS reference model, are also useful for evaluating tools based on a standard. In short, the planning and implementation of digital repository self-assessments is manageable based on the larger number of trusted resources that provide the fundamentals.

A notable resource for designing preservation plans are the NDSA Levels of Digital Preservation. The levels are intended to "offer clear, baseline instructions on preserving digital content at four progressive levels of sophistication across five different functional areas" (Phillips, Bailey, Goethals, & Owens, 2013). The levels are described simply and expressed in a succinct table. They are also intended for various institution sizes, resource levels, and without limiting the content type or technologies (Phillips et al. 2013). A specific example of applied archive-building against a standard is Priscilla Caplan's talk outlining the process of creating the software application Dark Archive in the Sunshine State (DAITSS)[17] at the Florida Center for Library Automation (FCLA). Not only does she summarize the timeline for the project, she describes the four theories behind development that included preservation strategies, the OAIS reference model, risk management, and file formats. More discussion on this talk is in the Limitations section.

Beyond basic tools, many libraries also create their own informal audits that are available on their respective websites as examples for references. While these resources are not collected in a single online location, several institutions include a preservation plan for their content that might include an internal assessment. For example, the self-assessment report from the Northern Arizona University's Cline Library in Spring 2014 (Welch & Phillips) is particularly useful for their demonstration of understanding the OAIS reference model, the description for collecting their documentation, and their own internal recommendations for improvement. Other such contributions to the general digital preservation community are continuously created as standards are being tested by users and auditors.

[16] See the POWRR Tool Grid as of 2013: http://digitalpowrr.niu.edu/tool-grid/
[17] DAITSS Digital Preservation Repository Software homepage: http://daitss.fcla.edu/

On the level of formal certification from metrics like ISO 16363, TRAC, DIN 31644, DRAMBORA, and DSA there is official documentation, published reports and certifications, and additional tools from digital preservation organizations. This study focuses on TRAC and DSA, both of which have several published certification reports and resources that are publicly available online. The HathiTrust Audit Report 2011 and corresponding elements included on the HathiTrust website (2011), and the CLOCKSS Audit Report 2014 (CRL) and corresponding blog record from David Rosenthal (2014) are primary examples of TRAC implemented. The HathiTrust report includes further steps for compliance to maintain TRAC certification and includes additional elements for further documentation, offering a useful example. The Controlled Lots Of Copies Keeps Stuff Safe (CLOCKSS) report is also unique in that the LOCKSS creator David Rosenthal also documented the TRAC Audit process on his blog in a three-post series that described the process and lessons learned (Rosenthal 2014). Comparatively, DSA provides a list of every DSA-certified repository coupled with their official certification reports (Seals). From this list, three relevant reports described in the Methodology section were the basis of comparison for the Digital Safe case study. Case studies for DSA are also published, notably the "ADS and the Data Seal of Approval – case study for the DCC" (Mitchan & Hardman, 2011), which offers an outline the process and timeline for attaining DSA certification.

Metrics as Development Tools

Digital preservation auditing metrics are built to assess existing repositories, not as development or scoping tools for ongoing projects. The issue is that there are more auditing metrics than there are development tools, so developers must sort through the multitude of preservation planning guides to find what they need. While preservation planning resources are ubiquitous, there are simply too many options that have not yet been filtered into a comprehensive document that provides the same project planning and designing standards as it does for auditing existing projects. Thus, few resources exist documenting projects that used an auditing metric as a development tool, but many resources exist illustrating the process of an assessment on an existing repository.

Third Parties in Digital Repositories

Outsourcing

Limited resources plague most digital preservation initiatives and outsourcing is a reasonable option for institutions lacking in funding, staff, and IT services. Adopting open-source software for infrastructure or outsourcing storage are common and extensive lists of tools exist from multiple digital preservation organizations. Larger institutions also outsource, one example being the UK Data Archive which lists its tools in its documentation.[18] There are hundreds of tools currently used for digital preservation activities, and this section of the Literature Review is only meant to establish that outsourcing is common.

Though not an exhaustive list, the following are examples of organizations that provide updated resources for digital preservation tools; DCC: Tools and applications,[19] DPC: Technical solutions and tools,[20] and Community Owned digital Preservation Tool Registry (COPTR).[21] Platforms like Fedora, Hydra, Preservica, Digital Commons, and many other resources that are unnecessary to list exist for use by digital repositories. Other organizations provide services rather than tools, which benefits institutions lacking in expertise rather than budget. AVPreserve,[22] for example, is a data management consulting and software development firm that offers recommendation on assessments, planning, software choices, and other aspects in digital preservation activities. Further, APTrust is a consortium where annual fee-paying members have access to long-term storage and preservation (Sites, M., 2013). Cloud services like Amazon Glacier provide basic storage, and organizations like Arkivum offer high-security storage, among many others.

Consortia

Joining a consortium is common and economical decision for digital repositories, especially for smaller institutions, as it alleviates the constant battle for sustaining funds and staffing (Wu, M., 2015). Consortia also increase content resources and allow for a more extensive and customizable technical infrastructure. David Rosenthal even posits that "serious digital archives like CLOCKSS require a distributed implementation, if only to achieve geographic redundancy"

[18] UK Data Archive tools: http://www.data-archive.ac.uk/curate/standards-tools/tools
[19] See DCC: http://www.dcc.ac.uk/resources/tools-and-applications
[20] http://dpconline.org/handbook/technical-solutions-and-tools/tools
[21] COPTR: http://coptr.digipres.org/Main_Page
[22] AVPreserve Services: https://www.avpreserve.com/services/

(2015) even if there is also a central organization leading the consortium. The issue is that TRAC, nestor Seal, and ISO 16363 all stipulate that a consortium cannot be certified as a whole, but each partner can be certified separately in a culminating certification for the consortium (Schwab, F., Tunnat, Y., & Gerdes, T., 2017). The stipulation here is that the DSA 2017-2019 Guidelines can certify repositories with Outsource Partners, which extends to consortia. The previous lack of consortia and their roles in complying with auditing metrics, however, has not prevented some consortia from seeking and achieving DSA certification.

The Digital Repository of Ireland (DRI), for example, is a research consortium compiled of six partners who all contribute to the management and implementation of the repositories policies, guidelines, and training.[23] The DRI is also DSA 2014-2017 certified as of 2015. This is possible because in their Implementation of the Data Seal of Approval report, they state in section 0. Repository Context that the DRI is "built by a research consortium," emphasis on the words 'build by' rather than 'is,' and then later in Requirement 5. that the DRI is "an unincorporated association of six partners" (DSA Board, 2015). Further, rather than list the roles of each partner, the DRI infrastructure relied on a "distributed development team with responsibilities for different Work Packages shared among multiple consortium partners" (DRI, 2015), effectively acting as a single entity rather than six different entities.

In contrast, the Goportis Digital Archive,[24] which is a consortium of three libraries with various roles in contributing to the function of the consortium. The consortium successfully achieved 2014-2017 DSA certification for Goportis by certifying all three of their libraries individually over the course of approximately six months (Schwab, et al., 2017). Their method was to establish the German National Library of Science and Technology (TIB) as the leader of the consortium as it "hosts, operates and administers the Digital Preservation system, and provides Goportis partners with access to the system" (DSA Board, 2015) so that their documentation could be the primary reference. The partners then created their applications simultaneously and collaboratively, referencing each library's policy often to illustrate the interdependence of their roles. This approach is more manageable for three partners, but not necessarily for a larger consortium of more partners.

[23] See the DRI website for additional information: http://www.dri.ie/about
[24] See the Goportis homepage for additional information: http://www.goportis.de/en/home.html

The CLOCKSS Archive is another interesting example of formal certification for a joint archive. While CLOCKSS does not identify as a consortium, it does rely on Stanford University for its technological infrastructure, is a "geographically distributed dark archive," and is supported by its partnership with various other libraries and publishers (CRL, 2014). CLOCKSS' TRAC certification is discussed further in the Limitations section and the Discussion section.

Context

The ultimate purpose of the two case studies is to examine how auditing metrics can provide a framework to aid developing repositories as they negotiate contracts with vendors and delineate the responsibilities of each party. Given that most digital repositories outsource aspects of their digital preservation workflow to some extent, this evaluation could prove useful to both developing projects and established repositories looking to outsource. This initial impetus of the TRAC case study was to provide a general scoping assessment for the ongoing project Digital Safe, described in the following section. The data collected ultimately pointed to the need to explore how the duties of multiple partners are delegated, and ideally audited, in a collaborative project.

Project Development

The original application of a digital preservation auditing metric to a non-traditional repository occurred in the context of the Oxford-Illinois Digital Library Placement Program (OIDLPP) as a short-term project The OIDLPP collaborators chose Digital Safe because there was an increased interest in resuming the project from the Bodleian Library, the Weston Library, and members of the University of Oxford who originally provided feedback in the project's development phase. The goal of the OIDLPP project was to assist in resuming the paused Digital Safe initiative by informally assessing the approach as a whole, the chosen technologies, and the current documentation to ensure that Digital Safe will develop into a trustworthy dark archive. The short-term project highlighted the strengths and areas of improvement for Digital Safe, but also revealed additional purposes for an auditing metric for scoping and planning incomplete and new digital preservation projects.

The Trustworthy Repository Audit & Certification (TRAC) was the original metric chosen to evaluate Digital Safe. In consulting with Michael Popham, Head of Digital Collection and Preservation Services for BDLSS; Neil Jefferies, Research & Development Project Manager for BDLSS and project leader for Digital Safe; and the Center for Research Libraries' assessment tools, TRAC was the agreed upon metric to compare to the plan for building Digital Safe. TRAC is significantly more detailed than the Data Seal of Approval or the Ten Principles, but is less complicated than ISO 16363. This allowed for a compromise between thoroughness and

timeliness. This level of thoroughness was needed to refresh the project team and stakeholders on Digital Safe after an official hiatus of nearly two years.

In order to understand how the use of the metrics can be applied to the proposed plan for Digital Safe, the following is a detailed review of Digital Safe, its evolution and current state, and a review of the technologies chosen for storage and digital preservation workflow.

Background for Digital Safe

The University of Oxford does not yet provide a service for the storage of high-security information. Additionally, among the independent colleges, departments, and universities within the University of Oxford there is no shared infrastructure, technology platform, or methodology for long-term storage. Student records, financial records, patient records, non-anonymized data, and other data with any personally identifiable information is considered high-security and high-priority, and increases every day. Digital Safe is the Bodleian Digital Library Systems & Services' (BDLSS) proposed solution for a universal, long-term, high-security digital preservation workflow and storage system. Originally called the Electronic Archives Pilot Project (EAPP), Digital Safe was initiated in 2012 and has completed two of the three planned phases that have been developed thus far. The ultimate goal of Digital Safe is to "deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis" (Jefferies, Hicks, & Rendell, 2016) The purpose is to have a single infrastructure that is managed locally and customizable by individual colleges and departments that is also economical and easily managed.

Phase 1 (2012-2013) focused on defining the scope of the project. The primary purpose was to determine if there was a need or want for a digital preservation service for the storage of administrative material at the University of Oxford. The project team interviewed various staff members in the colleges, IT Services, the Oxford Colleges Librarians Group, Oxford Archivists Consortium, and other related groups, committees, and departments. Phase 1 established a desire for such a service and gathered information on the range of technical, security, and infrastructure requirements that these institutions might need. Phase 1 also verified that most of the University is relying on the library to provide a solution, which the BDLSS supports.

Phase 2 (January 2013-June 2014) investigated service and infrastructure models for a long-term digital preservation service. The project team examined the infrastructure that the Bodleian

Electronic Archives and Manuscripts (BEAM)[25] currently uses, Oxford's Data Archiving infrastructure for Oxford-produced research (ORA-Data)[26], and services such as DataBank[27] for possible reuse or outsourcing. BEAM infrastructure, developed in 2005, is held on a stand-alone server that has recently not been able to keep up with the increase in acquisitions and the level of organization and security that the BEAM would prefer (Thomas, S. Personal communication, 2016, July 19). ORA-Data was found incompatible for the type and amount of security measures that would need to be implemented. Building an entirely new infrastructure was also investigated but would not have been time or cost-efficient. DataBank was also ruled out for not including all the aspects necessary for the project in one platform, requiring additional outsourcing or increased time and money for the BDLSS.

Phase 2 determined that the Designated Community would likely be the following: College Archivists; University Archive; Central Administrative Records Management; Departmental Research Records Management; and Personal Material held by BEAM (Jefferies, et al., 2016). This will not limit the type data that can be added. Given that these users may not be advanced technical users, the interface design will need to be user-friendly and technical support is necessary. The ultimate decision was to outsource the technologies to third-party vendors and manage the service on a local interface designed at the BDLSS. The name of the project also changed in Phase 2 from Electronic Archives Pilot Project to Digital Safe. The technologies chosen were Arkivum for long-term, high-security storage, and Archivematica for digital preservation activities. A Digital Safe blog[28] was also developed in 2013 by David Tomkins and provides public information on Phases 1 and 2. The blog will be continued in Phase 3 (Jefferies, N., Personal communication, 2016, July 19).

Phase 3 is in development but has not yet received funding to continue, hence the project hiatus. Funding is requested for three major activities. First, to further investigate and fully develop an ideal contract with the outsourced technologies, Arkivum and Archivematica. Second, to design and implement the business and service models within IT Services. Finally, to cover the start-up, training, and storage costs of one year of operating the service. Once the service is deployed to

[25] BEAM homepage: http://www.bodleian.ox.ac.uk/beam
[26] ORA-Data homepage: http://www.bodleian.ox.ac.uk/bdlss/digital-services/data-archiving
[27] DataBank Homepage: http://www.databank.com/
[28] See the Digital Safe blog for additional information: https://digitalsafe.wordpress.com

early adopters, the project team will track user feedback and service function for one year before deciding to launch a service available to the entire University (Jefferies, et al., 2016). A successful Phase 3 will result in a test-run for a service providing long-term, largescale, high-security storage for data produced and held by the University of Oxford. This service is planned to provide training sessions and materials available between training sessions. The Oxford brand and the large number of clients that the University will potentially be bringing to Arkivum is a motivation for Arkivum to work with Digital Safe and develop less expensive start-up and training fees (Jefferies, N. Personal communication, 2016, July 19). Additionally, Archivematica is a built-in tool in Arkivum. In signing with Arkivum the University will only need to purchase a single Archivematica license fee rather than over forty individual fees. Finally, the success of Digital Safe would will also ease the burden of data managers at the University of Oxford as the service is designed to be user-friendly, the technology will be supported by their developers, and minimal stress will be put on the IT Services and the future Steering Committee by handling small issues and outsourcing larger issues.

Outsourced Technologies

In brief, Arkivum is a "long-term, large-scale managed data storage" (Arkivum, 2017) that provides high-security processing and storage with strict accessibility processes. As of July 2016, they have two products that may be utilized by Digital Safe: Arkivum/1+1 and Arkivum/100. Arkivum/1+1 (2015) has one digital copy of the data held in a secure location, and one physical copy held on LTO data tape held in Escrow at a separate location. Arkivum/100 (2015) has two digital copies of the data held in two geographically separate, secure locations, and one physical copy held on LTO data tape held in Escrow at a third separate location. Arkivum/100 also offers the 100% integrity guarantee by ensuring three copies are being managed and preserved. Arkivum services are dictated by the number of pipes being used. Pipes are ingest workflows that can host multiple archives by one client or multiple archives from multiple clients. Each client login can have customized workflows, though only one login can be active at one time. Audit trails are available, and data is only accessible by the administrator login for that specific archive via an encryption key. Without the master encryption key the data cannot be retrieved, even by University staff or IT Services, which ensures the security of the data. Arkivum also contains both digital and physical copies, a contingency plan for lost data, and is ISO 27001 Information security management certified (Arkivum, 2017).

Archivematica is an open source digital preservation workflow tool developed by Artefactual, a company that develops open-source tools for libraries and archives, that has recently been built in to Arkivum (Stanbridge, N., 2016). Arkivum specifically only "provides safe and secure data archiving," (Arkivum, 2014) not digital preservation activities. Archivematica can offer a customizable digital workflow tool, including SIP creation, normalization, AIP packaging, and DIP uploading. This is then ingested into Arkivum.

Finally, the interface for integrating these technologies is planned to be filtered through a local website managed by the Digital Safe Steering Committee and eventually transferred to BDLSS as a provided service. The purpose is to ensure that there is both customization and local technical support. See Figure 1 below for a visualization of the technologies and interface combined.

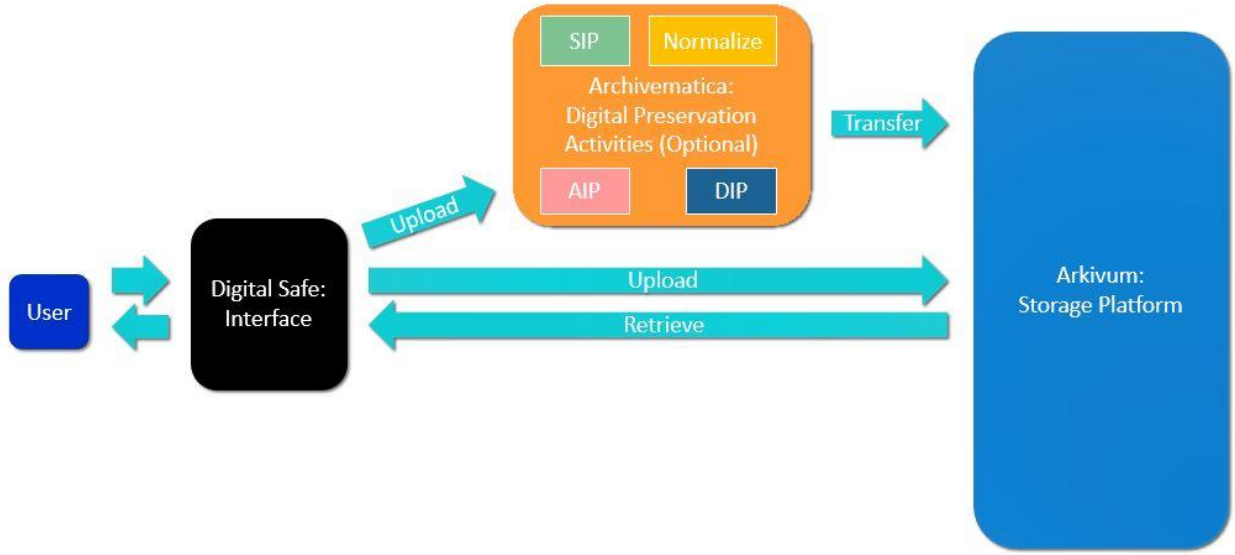

Figure 1: Visualization of the ideal workflow between the local interface, Archivematica preservation activities, and Arkivum storage.

Rationale

In addition to Digital Safe being the focus in the OIDLPP project, it also serves as an example of an ongoing project, meaning that the project requires assessment for the purposes of development and planning rather than for certification. Further, Digital Safe must become a dark archive because its aim is to store confidential information. This in and of itself means the project will eventually produce a non-traditional archive: audit metrics by and large define

digital repositories in terms of requirements for transparency and accessibility to users, while dark archives, by definition, have very limited transparency and access mechanisms. Further, Digital Safe is ongoing and is designed around the culmination of three different technologies.

These facets of Digital Safe create a unique subject for non-traditional auditing and the potential use of metrics as a scoping technique. These case studies point to future research into how standards handle dark archives that are not meant to be fully transparent and an investigation into the idea of certifying a body without a body. This thesis focuses on the more immediate concern of how audit metrics provide frameworks for developing services built upon multiple partners and delineating their roles in the service.

Methodology

Evidence Collection

Before any internal or external assessment can occur, the appropriate documentation and other evidence need to be collected. Digital Safe as a project, the two Digital Safe technologies, and resources on the chosen auditing metrics all required time dedicated to collecting comprehensive documents and guides.

Digital Safe

Data for evaluating the Digital Safe project according to audit criteria are collected from multiple sources. First, Digital Safe consisted of various unorganized documents that lacked a single storage space. These documents included: meetings minutes from the Steering Committee and the Oxford Colleges Librarians Group; Project Initiation Documents; letters of support from various Oxford Colleges; Programme and Project Highlight Reports; conference presentation materials; Project Request Forms; Request for Change forms; End Project Reports; various newsletters and copies of email communications; and various other presentation and documentation materials. These materials were provided by David Tomkins, the Curator of Digital Research Data for the BDLSS and Project Manager for Phase 2 of Digital Safe. David Tomkins also developed the blog to publicly track the progress of Digital Safe and has presented the project at multiple conferences. These documents are not publicly available, with the exception of the published newsletters and blog posts in the Digital Safe blog.[29]

Given that the project has been conducted in phases with different project members, the documentation proved to be a challenge to compile. Individuals often maintained various documents that were not necessarily related to each other. Some information was stored in a Google Drive, and other information is hosted on the blog. This lack of cohesive documentation hindered a comprehensive perspective on Digital Safe, which further complicated the application of the metric. Documentation collection is a common challenge even for established repositories and can take months of preparation just to begin a formal audit process. In the case of Digital Safe and other ongoing projects, organizing documentation in the development stages was

[29] See the Digital Safe blog for additional information: https://digitalsafe.wordpress.com

challenging but will be useful once the project reaches the point where it could be formally audited and searching for documentation becomes less of a challenge.

Furthermore, interviews with various stakeholders were conducted to compile a background on Digital Safe to both gauge interest in continuing the project and to establish how stable the governance and infrastructure is for the project. In addition to David Tomkins, who provided futher explanation and context on the EAPP documents, Susan Thomas and Neil Jefferies were informally interviewed. Susan Thomas, Head of Archives & Modern Manuscripts for the Weston Library in Oxford and previously the Project Manager/Digital Archivist for Bodleian Electronic Archives and Manuscripts (BEAM), is also a stakeholder in Digital Safe. Because BEAM contained a possible infrastructure for Digital Safe to be modeled, Susan Thomas was the primary contact for providing information on BEAM and relaying the needs of BEAM to the Steering Committee. Specifically, she was asked to describe the status of BEAM, to retrieve some background information on Digital Safe, and to review BEAM's current needs from the proposed Digital Safe service.

The second primary interview was with Neil Jefferies, the head of Innovation and a Phase 3 Digital Safe project lead. Neil Jefferies had been involved with Digital Safe since Phase 2, working with David Tomkins in creating presentation materials and acting as the primary investigator into the technologies chosen for Digital Safe. In addition to creating the Project Initiation Document for Phase 3 and having in-depth information on what benchmarks needed to be reached to establish funding for Phase 3, Neil Jefferies also offered information on Arkivum's contract with the University of Oxford, the status of Phase 3, and other information that was not publicly available.

Given that the OIDLPP project was conducted in July of 2016, many current and previous members of the project and stakeholders were not available as they were traveling. This limited the extent of personal information that could be incorporated.

Digital Safe Outsource Technologies: Arkivum & Archivematica

Arkivum and Archivematica both contain extensive documentation illustrating their services. Arkivum provides program documentation in the form of webpages, a detailed Frequently Asked Questions document, and information brochures for each available storage product. Specific contract information with the University of Oxford was provided by Neil Jefferies.

Archivematica also included program documentation in the form of webpages, Storage Service documentation illustrating how Archivematica is built into Arkivum, a Wiki page, and a Format Policy Registry. Both services also provide user training materials and courses for reference.

TRAC & DSA

Further sources for TRAC were also consulted for examples of acceptable evidence for meeting their respective criteria. In addition to the most current official TRAC criteria, the 2011 HathiTrust Trustworthy Repository Audit and Certification was the primary example. The full report (CRL, 2011) was accessed on the Center for Research Libraries (CRL) website, and additional elements were published on the HathiTrust website (2011) Additional CRL metric materials and sources were also examined. DSA Guidelines were accessed from their website and additional published certification reports, community-created records of working with DSA, and previous guideline versions were used for reference.

The TRAC Assessment

The informal assessment originally took place in July 2016 for the BDLSS at the University of Oxford based on the 2007 Trustworthy Repositories Audit & Certification: Criteria and Checklist (CRL & OCLC). The results were presented as a formal report to the Oxford e-Research Center (OeRC) in August 2016 and organized per the TRAC Criteria structure. Each response includes the criteria, the response, and the example evidence provided by TRAC. A full version of this can be found in Appendix B and is referenced in the Results section of this thesis. The responsibilities of Digital Safe as a service, Arkivum as a storage platform, and Archivematica as a digital preservation workflow are all considered for each criterion. If they do not have a responsibility for the criteria it is acknowledged as such. An example of this is criteria B2.2 "Repository has a definition of each AIP (or class) that is adequate to fit long-term preservation needs," in which Digital Safe would rely on the technologies, and Arkivum "provides safe and secure data archiving" (Arkivum Ltd., 2014), not digital preservation activities, and therefore responsibility is solely on Archivematica.

As an added measure for the BDLSS, the audit also includes a rating system that indicates the completeness of each criteria and section. Nancy McGovern of MIT Libraries developed a TRAC review tool in Drupal that allows institutions to self-review themselves. The tool was built in 2013 but is currently only hosted on Archivematica, also requires a DRUPAL

installation, and does not currently have an independent functioning website. This installation and the lack of documentation or examples of the tool omitted its use in the evidence collection and methodology. However, one of the features is a rating system for TRAC compliance, which is utilized in this informal audit. These ratings were also added up and averaged for the completeness of the three major TRAC sections.

The rating system is as follows (McGovern, N., 2013):

- 4 = fully compliant - the repository can demonstrate that has comprehensively addressed the requirement
- 3 = mostly compliant - the repository can demonstrate that it has mostly addressed the requirement and is on working on full compliance
- 2 = half compliant - the repository has partially addressed the requirement and has significant work remaining to fully address the requirement
- 1 = slightly compliant - the repository has something in place, but has a lot of work to do in addressing the requirement
- 0 = non-compliant or not started - the repository has not yet addressed the requirement or has not started the review of the requirement

The DSA Assessment

The DSA assessment took place in March 2017 using the Core Trustworthy Data Repository Requirements per the 2017-2019 Data Seal of Approval Guidelines released in November 2016. Rather than treating the DSA assessment as a formal report like the TRAC assessment needed to be, the assessment criteria is listed as succinctly as possible to demonstrate the level of implementation or theory development and mention any existing documentation. A more complete report can be found in Appendix C, which is also referenced in the Results section of this thesis. This assessment is meant to indicate if it is possible to assess an ongoing project with multiple partners using the DSA requirements, and if so, its effectiveness. Further analysis will determine which sections, if any, are most useful in the design phase for a digital project.

In addition to the 16 requirements there is a section for Context that is also mandatory but not measured on the following Statement of Compliance scale. The Context section includes information on designated communities, repository type, level of curation, and any information on outsource partners. This section is included in the results to ensure the full scope of the DSA guidelines are applied to the limited developmental documentation for Digital Safe.

DSA also provides a Statement of Compliance on a scale from 0 to 4 that is mandatory for each guideline. Similarly to TRAC's rating system, this scale also indicates the level of completeness and is included in the results with a brief statement of explanation that is further explored in the response to the guidelines. The Statement of Compliance scale, as per the 2017-2019 DSA guidelines (2016), are as follows:

Statement of Compliance	Means	Comments and/or URLs
0	N/A: Not Applicable	Provide an explanation.
1	No: We have not considered this yet.	Provide an explanation.
2	Theoretical: We have a theoretical concept	Provide a URL for the initiation document.
3	In progress: We are in the implementation phase	Provide a URL for the supporting document.
4	Implemented: This guideline has been fully implemented for the needs of our repository.	Provide a URL for the supporting document.

Results

Case Study I: TRAC

TRAC provided both obvious and understated successes and gaps in the current stability and development of Digital Safe. For the purpose of organization these results are paralleled with TRAC structure. In order to avoid repetition and otherwise excessive results, the following constitutes a consolidated response for each subsection of criteria. Several subsections have been combined because the same evidence applies to both criteria and the results are a condensed version of the initial informal audit. The full report given to the OeRC with each criterion can be referenced in Appendix B.

A. Organizational Infrastructure

A.1 Governance and organizational viability

TRAC revealed that the governance for the service is not yet in place and policy building will require the most attention during Phase 3. The rating system averaged a 1.45/4 completeness overall. As Digital Safe becomes more cohesive its project goals have altered through each phase. It is anticipated that once it is complete, a mission statement will be derived from the combination of the Electronic Archive Pilot Project's project goal (Wilson, J., 2012) and the statement in the Phase 3 Project Initiation Document to "deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis once in production" (Jefferies, Hicks, & Rendell, 2016). The project team identifies Digital Safe as a service, which also indicates that the mission statement will likely be service-oriented rather than describing repository goals. There is no additional policy formally produced that describes the current state and goals of the project outside of the Phase 3 Project Initiation Document.

A.2 Structure and Staffing

Currently there is a project group and Steering Committee leading the movement to resuscitate the Digital Safe initiative. The hope is that the Steering Committee will continue, or develop into a similar governance committee (Jefferies, N. Personal communication, 2016, July 19). The primary responsibilities for such a committee include reviewing the contracts, funding, and any updates from IT Services and the technologies as needed. IT Services play a vital role in Digital Safe, as the ultimate goal is to develop and business and service model so that Digital Safe will

be integrated with IT Services. This model will better determine staffing needs and training. IT Services will update the University web space for Digital Safe and assist in training, troubleshooting, and communicating with the Steering Committee. The Steering Committee will review and update the service as needed. Because the technologies are outsourced there is little technical training needed to maintain the interface, and the University will rely on Arkivum and Archivematica for maintaining their product. Arkivum provides training virtually, on-site, and in workshops. The plan is to bring Arkivum in for training and to use and develop their training materials to do local training in the University. These materials will be broad, as it is up to the client to determine how much they want out of the service, and will be added to the Bodleian Library's current collection of training materials. These materials will then be reviewed an updated on a 4-5 year cycle by the governance committee.

A.3 Procedural accountability & policy framework (documentation)

Digital Safe does not yet articulate the key users, basic policy, and contact information. The nature of a dark archive is not transparency and much information cannot be made public, however as a service provided to the University, general information and documentation is necessary.

Phase 2 determined that the key users for this service have been identified as: College Archivists; University Archive; Central Administrative Records Management; Departmental Research Records Management; and Bodleian Electronic Archives and Manuscripts (Jefferies, et al., 2016) These users have materials that require high-security and low-accessibility, including administrative records, student records, financial records, personal communication, medical reports, non-anonymized case studies, and other material that has personal, identifiable information. Users are not limited to only these categories, however, as the service is open to all who want to use the service and are affiliated with the University of Oxford. As described above, policies have not yet been developed. The Steering Committee is currently directing the project, but much of the policy will be directed by the contract agreements between the University and Arkivum before they can be documented. Policies for the technologies, however, are well-detailed and located on their respective websites.

Legal permissions will largely be the responsibility of the client as much of the content will be produced by the client. Other materials that have been acquisitioned by the library may require

additional policy measurements, but this is also primarily up to the library and BEAM to maintain. The contract between the University and Arkivum may need to determine if there is a need to build in a deposit agreement concerning the permissions for migration copies. It should be noted that Arkivum employees do not have access to any material as it is only accessible with an encryption key held by the client.

Outside of Digital Safe, the technologies have extensive documentation and are responsible for maintaining their certification and upholding their own policies. Arkivum has automated annual data integrity checks, five-year hardware and software migration, and developed an LTO road map to prevent technology and data obsolescence. More information is on their website. Arkivum also maintains ISO 27001 certification, is audited every six months, and welcomes client audits.

A.4 Financial sustainability

Digital Safe does not have a full cost model developed for the long-term. Phase 3 has outlined the short-term business plan that includes University staff training, Arkivum training costs, startup costs, 1 year of Arkivum service and storage space, and 1 year of maintenance fees. All are outlined in the Phase 3 PID (Jefferies, et al., 2016). Developing this model is a priority for Phase 3 and will ultimately determine what funding is provided after the first year of service to maintain the website and any license and storage fees. Once the service is launched and the clients choose the service they prefer, they will be responsible for their own funding. Phase 3 will help determine the payment method agreed upon between the University and Arkivum. If Arkivum handles the billing, they do direct invoicing for each client. If the University handles the billing it will be managed as a library service, similarly to the process for handling services like catalog use, access to electronic journals, and IT Services. The Steering Committee will be responsible for reviewing and securing funding once it is determined if Phase 3 has been successful.

A.5 Contracts, Licenses, & Liabilities

Digital Safe and Arkivum have been contact since 2014 and are still in discussion over contract specifications. Once funding has been secured for Phase 3, this will be developed on a beta level for the first year of early adopters, and then re-evaluated after the first year is completed. Deposit

agreements and copyright issues will need to be evaluated to ensure legal inclusion of any acquisitioned material that might be included, such as from BEAM.

Section A5 was not completed in this audit because there is no existing documentation yet. However, this does raise several questions how Digital Safe will develop their contracts and licenses between its users and the outsource technologies because this section can be useful in informing how Digital Safe's vendor contracts are written, which is further explored in the Discussion section.

B. Digital Object Management

Digital object management is the most developed section for Digital Safe per TRAC metrics, likely because the majority of the documentation is contributed by the outsourced technologies rather than developed by the Digital Safe Steering Committee. On average, this section is rated 3.5/4.

B.1 Ingest: acquisition of content

All materials that are ingested into Arkivum will be provided by the client. Digital Safe will not be providing any content, only suggesting the type of content that might be ingested. Digital Safe also has recommendations on what metadata properties might be preserved, file format choices, and digital preservation workflow activities, but the content is strictly the responsibility of the client. Arkivum only requires a file to ingest, and depending on the complexity of the digital preservation workflow the client chooses, it can also ingest additional associated files.

B2 Ingest: creation of the archival package & B.3 Preservation Planning

Archivematica offers a customizable workflow tool to clients wanting to digitally preserve in addition to secure storage. Should the client require an OAIS model, Archivematica can create SIPs, normalize data, package AIPs, and normalize data, among other activities. A wide spectrum of processes is available, but ultimately up to the client to choose. Digital Safe may also be able to offer recommendations on best practices.

B.4 Archival storage & preservation/maintenance of AIPs & B.5 Information Management

Arkivum offers multi-location, high-security storage within their own data centers and at an additional Escrow location. During the ingest process, the client can follow the process of ingest and is given a green light once the data is fully ingested and secure so that the client may delete

their own copies. Arkivum also offers audit trails as a service. These must be turned on and specified by the client upon contract agreement. Audit trails will log every event affecting the content, such as an integrity check or migration, and the employee, time, reason, and any changes made. More information can be found on Arkivum's Audit Trails page.

C. Technology, Technologies Infrastructure, and Security

This final section meets most TRAC criteria in theory, again because the documentation for the technologies is extensive. Once the three entities are integrated with each other into a working service it will be more efficient for testing the functionality of the service and this section would need to be re-evaluated.

C.1 System infrastructure & C.2 Appropriate technologies

Arkivum and Archivematica were chosen because they built for institutions without the technical abilities, time, or funding to develop their own infrastructure for digital preservation and storage, and with needs for high-security and low-access. Their systems can be operated from standard operating systems and do not require high levels of technical ability, fitting the needs of potential users identified in Phase 1.

The decision to utilize these technologies was determined after several years of investigation by the Steering Committee. However, the success of combining these technologies will only be made evident by live testing of the service as a whole and therefore cannot be officially established as a success or otherwise.

C.3 Security

Although a mock interface has been created, the system combining Digital Safe, Arkivum, and Archivematica has not yet been completed. Once the service is deployed to early adopters, the security will be tested and become more concrete. The technologies, infrastructure, and security for Digital Safe are largely provided by and well-documented by Arkivum. The service that the University will receive will vary based on client choices and any specific agreements made between the University and Arkivum, and by the client and Arkivum.

Case Study II: DSA

The results for the DSA assessment are organized by each Requirement. The Requirements are related to each other so there is some overlap, such as between 15. "Technical infrastructure,"

16. “Security,” and 9. “Documented storage procedures.” However, the documentation states that “all Requirements are mandatory and are equally weighted, standalone items” (DSA, 2016) so none of the Requirements are consolidated. Rather their responses are condensed and the full version of the DSA assessment can be referenced in Appendix C.

0. Context

Repository type: Institutional repository, (Other: Dark archive still in development)

It should be noted that DSA provides a set list for Repository types and a write-in option, so both are included here.

Brief Description of the Repository’s Designated Community:

The key users for this service have been identified in the Electronic Archives Pilot Project (renamed Digital Safe) Phase 1 as: College Archivists; University Archive; Central Administrative Records Management; Departmental Research Records Management; and Bodleian Electronic Archives and Manuscripts. This community is part of the larger University of Oxford community. These users have materials that require high-security and low-accessibility, including administrative records, student records, financial records, personal communication, medical reports, non-anonymized case studies, and other material that has personal, identifiable information. These users were identified after interviewing various colleges and departments on campus and determining a need for a universal storage system (Jefferies, Hicks, & Rendell, 2016) and are internally documented with letters of support from various colleges. Users are not limited to only these categories, however, as the service is open to all who want to use the service and are affiliated with the University of Oxford.

Level of Curation Performed

A. ***Content distributed as deposited***
B. ***B. Basic curation – e.g., brief checking, addition of basic metadata or documentation***
C. ***Enhanced curation – e.g., conversion to new formats, enhancement of documentation***
D. ***Data-level curation – as in C above, but with additional editing of deposited data for accuracy***

Digital Safe as a service is being designed to provide a wide range of digital preservation activities that suit the needs of the Designated Communities. Digital Safe itself will not offer curatorial services, therefore the lowest possible level of curation ingested by Digital Safe will

fall under “A. Content distributed as deposited.” As discussed in Requirement 7. “Data integrity and authenticity” and in Requirement 8. “Appraisal,” the storage platform Arkivum can ingest any file and will disseminate the file in the exact same condition it was in upon ingestion. Additional digital preservation workflows are customized by the client, who will also determine the relevance and authenticity of any content they opt to store in Digital Safe.

Outsource Partners

Arkivum is a proposed contractual partner that will provide the means of digital storage in the form of ingest pipes for users of the Digital Safe service. According to their website, “Our operations at all sites, including our business offices, is certified to ISO 27001 information security standards” (Arkivum Ltd., 2014). Arkivum is also audited every six months and welcomes client audits (Arkivum Ltd., 2014). In regards to service agreements between Arkivum and clients, there will be a basic contract between the University of Oxford and Arkivum, and individual client preferences will build upon that contract and articulate the clients’ product choice, workflow preferences, and additional storage space options (Jefferies, N., Personal communication, 2016, July 19).

Archivematica is a proposed contractual partner that is built into Arkivum and will provide the digital preservation workflows for users of the Digital Safe service. Ideally the University of Oxford will be able to purchase one license for Archivematica via their contract with Arkivum, rather than each individual client purchasing a license (Jefferies, N., Personal communication, 2016, July 19), though these discussions with Arkivum are ongoing. Archivematica does not as yet hold any previous certifications.

1. Mission/Scope

Statement of Compliance: 2 The scope has been determined by the designated community but the mission statement is still in development as the project evolves. The chosen outsource technologies have fully developed and implemented mission statements.

Self-assessment statement:

The original Electronic Archive Pilot Project’s mission statement is as follows: “The Electronic Archive Pilot Project will establish the feasibility of a working electronic archive for the use of the whole of the Collegiate University. The archive will support the safe and secure storage of all

classifications of non-public record data that individual departments, colleges and associated units are required to keep legally or would like to keep for historic reasons. The pilot project aims to develop a cost recovered service" (Wilson, J., 2012). The Digital Safe service will likely draw on the phrase from the Phase 3 Project Initiation Document to "deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis once in production" (Jefferies, et al., 2016).

2. Licenses

Statement of Compliance: 1 This section is designed in theory and implementation has begun, and is the focus of the next project phase.

Self-assessment statement:

This entire section relies on a fully developed contract with Arkivum. A priority of Phase 3 of Digital Safe is finalizing the contract with Arkivum and in determining any preservation rights and copyrights. In regards to transferring control of data from a client to Arkivum, any services that are part of the library may come under the Heritage Institution exception for the right to change objects and make copies, which would occur in the regular migration of data in Arkivum and in any Archivematica workflow the data is pushed through. (Jefferies, N., Personal communication, 2016, July 19). Other licenses will be at the discretion of the clients.

3. Continuity of access

Statement of Compliance: 2 This is a theoretical concept that will be developed should Digital Safe regain funding. As an Outsource Partner, Arkivum will be responsible for maintaining continuity of access as per their mission statement.

Self-assessment statement:

As any future contract with Arkivum will dictate, Digital Safe will rely on the technologies to remain updated on and implement any evolving best practices in the field. Arkivum has policies and automated processes in place for running integrity checks on the data, software, and hardware. The data is retrieved annually and given an integrity test based on checksums. Arkivum's policy is that data is migrated to new media following the LTO roadmap. The LTO data tapes in Escrow are also migrated every 5 years.

Digital Safe is labeled as a service by the project team, and therefore its only responsibility is to maintain the service and contracts with technologies, and has no influence on the amount of time the data is held. Ultimately those using the Digital Safe service will be responsible for maintaining their encryption key, restricting or releasing access to materials, and providing their own funding to ensure their space and digital preservation workflows are maintained. There is no formal documentation on policies in place for changes in circumstances from Digital Safe. Ideally the Steering Committee will develop these policies in Phase 3 and eventually transfer responsibility to technical support.

4. Confidentiality/Ethics

Statement of Compliance: 2 As per the goal of Digital Safe to "deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis once in production," Digital Safe will store confidential information and will need to consider their documentation carefully.

Self-assessment statement:

Digital Safe will need to be a dark archive because its purpose is to store confidential information. First, Digital Safe has four different levels of access described in their Phase 3 Project Initiation Document that also corresponds to features offered by Arkivum. See Appendix C for a detailed description of the access model.

Arkivum employees do not have access to any client data. Additionally, "there is no direct customer access to data in the Arkivum data centres, e.g. through a web interface or cloud API. This ensures all ingest and access is properly managed through Arkivum appliances" (Arkivum Ltd., 2014) Arkivum is build on ingest pipes that are matched with an encryption key that is unique to each client. Without the encryption key there is no access to any of the data in its original form or a copy. Arkivum does have a detailed contingency plan described in Requirement 16 (Arkivum Ltd.). Arkivum also adheres to UK Government Information Levels IL2 and IL3.

Archivematica is a digital preservation workflow tool to which only the client has access. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process

to complete before it is stored in Arkivum. Archivematica is integrated with Arkivum via Arkivum's A-Stor, which will be built into the contract between the University and Arkivum.

5. Organizational infrastructure

Statement of Compliance: 2 As Digital Safe project team members investigate contracts with Arkivum they are also in the process of developing a stable organization infrastructure.

Self-assessment statement:

Digital Safe is currently in the process of acquiring funding for Phase 3. There is a cost model and projected expense report developed. The projected expense report includes: IT Services internal staff, Non-IT Services staff, Hardware, software, training and equipment/storage (Arkivum), and included a .5% charge for Prime Minister's Office charge, an 85% charge for contingency per the Monte Carlo Simulation, and a forecast on on-going charges per year (Jefferies, et al., 2016). See Appendix C for more information on the proposed expense reports for Digital Safe.

The project team for Phase 3 of Digital Safe is established, but the ultimate short and long-term staffing duties will rely on the contract with Arkivum and collaboration with the IT department. The plan is that the service will be built into the University of Oxford's IT department. Further, a business and service model is a priority for Phase 3 of Digital Safe (Jefferies, N., Personal communication, 2016, July 19).

Digital Safe will on the technologies to maintain their own staffing. According to Arkivum, "In addition to technical change in the archive system, managing staff transitions of those who run the system, for example support staff and administrators, is required"[30] Archivematica is created and staffed by Artefactual Inc.

Digital Safe will update any announcements and training material for the service, and the local governance committee will have little public documentation aside from the aspects described above that will need regularly updated due to the sensitive information and privacy of the dark archives Digital Safe is providing.

6. Expert guidance

[30] Arkivum "Data Integrity:" http://arkivum.com/data_integrity/

Statement of Compliance: 3 The decision to outsource the technologies for Digital Safe indicates that the project team wants the expertise of established digital preservation organizations, with which contract development is ongoing.

Self-assessment statement:

Digital Safe is the result of feedback collected during Phase 1 and will be built based on designated community needs. In the future, the Digital Safe service will be built into the University of Oxford's IT department. Ideally the webpage would have contact information for the governance committee or other local managing team who can assist in troubleshooting smaller issues and directing to training and Arkivum and Archivematica help pages (Jefferies, N., Personal communication, 2016, July 19).

Arkivum seeks feedback from their users.[31] Expertise sources for Arkivum are not in public documentation, but their transparent and detailed descriptions of their preservation workflows and storage methods, additional case studies[32] on their website, and list of current clients offered under the Industries tab on their website offer community support and proof of successful methods.

Archivematica was originally a project use case for OAIS to "process analysis to synthesize the specific, concrete steps that must be carried out to comply with the OAIS functional model from Ingest to Access." This project expanded beyond OAIS into its current state as an open-source digital preservation workflow tool based on user feedback (Artefactual Inc.).

7. Data integrity and authenticity

Statement of Compliance: 3 Arkivum and Archivematica both maintain detailed documentation, which will be absorbed into Digital Safe documentation by the contract between Arkivum and Digital Safe.

Self-assessment statement:

[31] Arkivum Chain of Custody: http://arkivum.com/chain-custody-audit-trails/
[32] Arkivum Case Studies: http://arkivum.com/resources/#

Digital Safe is a service. Authentication of the original data will be the entire responsibility of the client, and each client may also have their own policies on permissions and permissions workflows.

In regards to service agreements between Arkivum and clients, there will be a contract between individual clients and Arkivum that articulate the clients' product choice, workflow preferences, and additional storage space options. Arkivum offers audit trails as a service. These must be turned on and specified by the client upon contract agreement. Audit trails are "accessible in machine-readable XML format through Arkivum REST API calls, either on a per-file or per-folder basis. Audit trails are also accessible in human readable PDF/A format through the Arkivum Service web interface or through an API call. For data integrity Archivematica utilizes fixity checks before AIP storage by generating and verifying checksums using a separate command-line app developed by Artefactual called Fixity (Artefactual Systems Inc.).

8. Appraisal

Statement of Compliance: 3 Digital Safe as a service is planned to offer both basic and extensive digital preservation activities in addition to storage, though the specific workflows are the responsibility of the client.

Self-assessment statement:

The Designated Communities that were identified after interviewing various colleges and departments on campus determined a need for a universal storage system (Jefferies, et al., 2016). The Designated Communities determine what information is included and will have their own documentation dictating the appropriate data. Digital Safe, Arkivum, and Archivematica have no influence over what data is considered appropriate by the Designated Communities. Arkivum can ingest any file format, but does not record their representation information. If the client chooses to utilize Archivematica, Archivematica can determine file formats and normalize using their Format Policy Registry, as well as creating SIPs and AIPs for record this process.

9. Documented storage procedures

Statement of Compliance: 3 Arkivum's long term-preservation strategies are publicly documented and will be absorbed into Digital Safe documentation upon the completion of a contract.

Self-assessment statement:

Digital Safe will rely on the technologies for storage and documented storage procedures as is outlined in their future contract.

Arkivum's documentation is publicly available on their website and specific client preferences are dictated in the final contract. Each file is encrypted with a unique symmetric key using AES256 encryption. The symmetric key for the file is then encrypted using a public key from a public-private key pair using RSA2048 encryption. Both AES256 and RSA2048 are industry standard encryption algorithms and widely used in high security applications, e.g. electronic commerce and for sensitive government information" (Arkivum Ltd., 2014). Many Arkivum clients have strict compliance policies and government documents and require an even higher level of security. Arkivum adheres to UK Government Information Levels by maintaining ISO 27001 certification.

Arkivum has policies and automated processes in place for running integrity checks on the data, software, and hardware. The data is retrieved annually and given an integrity test based on checksums. Arkivum's policy is that data is migrated to new media following the LTO roadmap. The LTO data tapes in Escrow are also migrated every 5 years.[33]

Arkivum maintains at least two copies of the data, one in a secure data center and one on LTO data tape held in Escrow. The Escrow copy is the backup for any data loss or corruption. If the digital copies are corrupted or the clients need to remove content, the LTO tapes are delivered to the client. If there is complete data loss, Arkivum provides a "financial guarantee underwritten by an Information and Communication Technology Professional Liability Insurance Policy," which provides coverage for direct loss relating to data loss (Arkivum Ltd., 2014). Arkivum also provides multiple client sites, so if one site is compromised the data may be retrieved at another site (Arkivum Ltd., 2014).

Archivematica is the digital preservation workflow that occurs before ingest into Arkivum storage and therefore does not have data back-up. Clients are notified of any failed actions and are responsible for managing these issues.

[33] Arkivum Data Integrity: http://arkivum.com/data_integrity/

10. Preservation plan

Statement of Compliance: 2 Arkivum's long term-preservation strategies are publicly documented and Archivematica's options are publicly documented, and individual contracts between clients and Arkivum will dictate specific preservation plans.

Self-assessment statement:

The preservation plan for each user of the service Digital Safe will be unique to their context. They will specify their preferences for storage and transfer custody to Arkivum in the contract with Arkivum, and they will design their own digital preservation workflow in Archivematica based on recommendations from both Digital Safe and Archivematica. They will determine their own preservation level and length of time the data is to be held, and communicate with Arkivum directly. Arkivum follows a strict chain of custody that allows for minimal contact with client data. According to their chain of custody authenticity begins " with the customer to ensure data has been correctly copied into the service. Once ingested, files become read only and cannot be updated or overwritten. Deletion of files is through a strictly controlled process that requires a request to be made to Arkivum. The default is that once a file is in the service then it remains in the service and does not change when it is within the service."[34] Any access to ingested data is restricted to individuals with the encryption key.

Digital preservation activities are solely the responsibility of the client, who will design them using Archivematica. Archivematica allows users to do various archival activities, including adding metadata in Dublin Core, adding rights in PREMIS, data normalization, AIP storage, DIP storage, communication with other tools (e.g. Archivist's Toolkit, ArchiveSpace, Arkivum), among other options. A SIP begins as a transfer. "In Archivematica, Transfer is the process of transforming any set of digital objects and/or directories into a SIP. Transformation may include appraisal, arrangement, description and identification of donor restricted, private or confidential contents."[35] A transfer can be created with submission documentation, existing checksums, or an existing METS structmap. The transfer will be processed through several micro-services. This is then ingested into Archivematica after the green light is given to the client. The completion of all

[34] Arkivum Chain of Custody: http://arkivum.com/chain-custody-audit-trails/
[35] Archivematica Transfer: https://www.archivematica.org/en/docs/archivematica-1.5/user-manual/transfer/transfer/#transfer-checksums

processes in Archivematica are indicated either green to indicate that a process has been completed successfully, and red to indicate that the process has not been completed successfully. A client can search for content by its name. Archivematica's naming system will retain the original name of the transfer unless a new name has been assigned to the SIP upon creation. This name will be combined with a Universal Unique Identifier that is generated and assigned during SIP formation.

11. Data quality

Statement of Compliance: 2 Arkivum and Archivematica maintain best practices, and Digital Safe intends to provide recommendations, but it will be the responsibility of the client to determine what level of quality their data maintains.

Self-assessment statement:

The metadata fields that Digital Safe recommends the clients' preserve are: Title, Description, Creator(s), ID Type, ID Value, Retention review date, Retention rationale, Resource type, Technical description(s), Covering date(s), Finding aid(s), Rights & Licensing information (Jefferies & Tomkins, 2014). It is ultimately the client's decision what properties are preserved.

Arkivum only requires a file for anything to be ingested. Arkivum has no responsibility for anything additional included in the ingest and relies on the client to upload any additional information. Arkivum can ingest any file format and will maintain a copy of the original file alongside a normalized file.

12. Workflows

Statement of Compliance: 1 While extensive documentation for Arkivum is well-established, the integration into Digital Safe documentation is still in progress and are a major focus for Phase 3.

Self-assessment statement:

The ongoing nature of Digital Safe means that documentation of processes is also developing. Project member Neil Jefferies has begun developing a contract with Arkivum, though this is not yet available (Jefferies, N. Personal communication, 2016, July 19). Contracts with Arkivum will

also help Digital Safe to establish documentation on deposits, security, and best practices for digital preservation workflow.

Digital Safe has also determined their Designated Community and a corresponding access matrix, seen in detail in Appendix C, Requirement 4. "Confidentiality/Ethics," will guide documentation evolution.

Arkivum and Archivematica both maintain extensive documentation that is publicly available. See the response to Requirement 9. "Documented storage procedures" for more specific information.

13. Data discovery and identification

Statement of Compliance: 2 The Designated Communities will only have access to their own data, for which they will have provided the identifier that will be maintained by Arkivum.

Self-assessment statement:

Files ingested into Arkivum as the storage platform Digital Safe will retain its original filename, and checksums are provided for incorporation in the Preservation Description Information (PDI) for the AIP, which Arkivum can maintain. The service follows the OAIS model for Archive Storage through use of replication, fixity monitoring and repair, disaster recovery, migration, and tiered storage to deliver a specified level of performance, availability and integrity of storage."[36] If the client decides to utilize the Archivematica tool they may choose to create and package an AIP. The client must determine the processing configuration, which can be left at Archivematica's default setting, or can be created by the client.

After configuring the process as desired, the SIP can be normalized and stored in an AIP. AIP reingest is also an option if the client wishes to add information (e.g. metadata and data normalization) after the SIP process, which could include producing different identifiers.

[36] Arkivum Integration with other systems: http://arkivum.com/integration-with-other-systems/

14. Data reuse

Statement of Compliance: 1 While documentation for Arkivum is well-established, the projected Digital Safe Designated Communities complicate the process of establishing licenses with vendors.

Self-assessment statement:

Digital Safe is designed to hold confidential data with high access restrictions. The Digital Safe Steering Committee has opted to utilize Arkivum as the storage technology and will rely on Arkivum to implement their responsibilities in data reuse that is outlined in their contract.

Ideally, Digital Safe will offer best practices based on recommendations from their Outsource Partners and other experts in the BDLSS based on Designated Community needs. Based on feedback from the Designated Communities, Digital Safe recommends basic metadata that is described in Requirement 9. It is ultimately the client's decision what properties are preserved.

Arkivum only requires a file for anything to be ingested. Arkivum has no responsibility for producing anything additional included in the ingest and relies on the client to upload any additional information. Arkivum can ingest any file format and will maintain a copy of the original file alongside a normalized file. Digital preservation activities are solely the responsibility of the client, who will design them using Archivematica. Archivematica allows users to do various archival activities, including adding metadata in Dublin Core, which is also ingestible by Arkivum, and further explained in Requirement 10. "Preservation plan."

15. Technical infrastructure

Statement of Compliance: 2 Outsourcing the technologies allows Digital Safe to customize their infrastructure based on pre-existing infrastructure, rather than building their own, which is being developed between the project team and Arkivum.

Self-assessment statement:

During Phase 2 of the Digital Safe project, the team investigated service and infrastructure models that included BEAM and ORA-Data systems at the University of Oxford, and also investigated DataBank as an Outsource Partner. All were abandoned, as explained in Appendix C, Requirement 15. "Technical infrastructure." Additionally, given that the Designated

Communities span across a range of departments, colleges, and expertise, the interface design will need to be user-friendly and technical support is necessary. This investigation lead the project team to opt for outsourcing storage to Arkivum, outsourcing optional digital preservation activities to Archivematica as a built-in tool in Arkivum, and create a local interface to be maintained long-term by the University IT Services. The process of considering multiple options and choosing Arkivum is evidence that Digital Safe is maintaining its mission statement to deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis once in production" even in the development stages. This does not fulfill this infrastructure, but it establishes a record of effort to build stable technical infrastructure.

Arkivum adheres to UK Government Information Levels by maintaining ISO 27001 "Information security standards" certification and can maintain IL2 and IL3 UK Government Information Levels. Should Digital Safe be built on a contract with Arkivum, it would allow Digital Safe to absorb their certification and partially fulfill this Requirement. Arkivum and its applications is also constructed to be used from most common operating systems (Arkivum Ltd., 2014). Additional technical support is available for unique operating systems. Updates are automatically provided to clients as they develop (Arkivum Ltd., 2014).

16. Security

Statement of Compliance: 3 The decision to outsource to Arkivum is heavily influenced by the security levels maintained by Arkivum, which will be absorbed into Digital Safe documentation upon the completion of a contract.

Self-assessment statement:

If Digital Safe were to establish a contract with Arkivum, security would largely be the responsibility of Arkivum. According to project member Neil Jefferies, a primary reason Arkivum is the choice for building Digital Safe is their contingency plan (Personal communication, 2016 July 19). First, the data is retrieved annually and given an integrity test based on checksums and preventative data migrations where "each copy has its integrity actively monitored and managed and any corruption or loss is automatically repaired to make the system self-healing" (Arkivum Ltd.). Arkivum's procedures, hardware, and locations are all certified to ISO 27001 standards and are audited every six months. "[Arkivum's] secure storage locations

are based in highly secure facilities, with our operations at all sites certified to ISO 27001 standards. Our locations are manned at all hours and access is strictly restricted to a list of named, trained and vetted members of the Arkivum Operations team. Each site is protected by best of breed firewall technology ensuring that our locations are protected from the latest advanced evasion techniques utilised by sophisticated hackers and intelligence organisations" (Arkivum Ltd., 2014). Data is secured based on "the ability to separately encrypt each file stored in our service. Only encrypted data is ever stored in [Arkivum's] service. Each file is encrypted with a unique symmetric key using AES256 encryption. The symmetric key for the file is then encrypted using a public key from a public-private key pair using RSA2048 encryption. Both AES256 and RSA2048 are industry standard encryption algorithms and widely used in high security applications, e.g. electronic commerce and for sensitive government information," (Arkivum Ltd., 2014).

Arkivum contains a strict chain-of-custody system, audit trails, and a highly detailed security model. "The security and audit model has been developed in partnership with Arkivum customers who have confirmed that the model meets their regulatory requirements as part of a due-diligence/audit process that they have conducted on Arkivum. This includes due-diligence by customers in clinical and financial sectors where regulation is strict" (Arkivum Ltd., 2014).

Finally, Arkivum provides a workflow and safety measures for integrity checks, including a secure data center and LTO tape in Escrow. If there is complete data loss, Arkivum provides a "financial guarantee underwritten by an Information and Communication Technology Professional Liability Insurance Policy," which provides coverage for direct loss relating to data loss (Arkivum Ltd., 2014). Arkivum also provides multiple client sites, so if one site is compromised the data may be retrieved at another site (Arkivum Ltd., 2014).

Archivematica is a digital preservation workflow tool to which only the client has access. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. Archivematica is integrated with Arkivum via Arkivum's A-Stor, which will be built into the contract between the University and Arkivum, and is therefore protected by the security measures of Arkivum.

Limitations

Implementing the case studies highlighted several limitations of this thesis that range from documentational to time limits. The reliance on previously developed standards, the auditing metric's documentation, and the lack of time to conduct additional case studies are all considered in the Discussion section.

Reliance on OAIS Reference Model

An immediate limitation of these case studies is that both of the audit models are built on the previously-established standards described in the Open Archival Information System (OAIS) reference model. The OAIS reference model, also known as ISO 14721:2012 was originally developed in the 1990's by NASA's CCSDS and provides a "conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term" for purpose of developing standards (Lavoie, 2000). It later became an ISO Standard in 2011. In brief, the reference model emphasizes understanding the Designated Community and its needs, controlling the data, and ensuring its availability and future use. A major function OAIS serves is to define universal digital preservation terms, which include but are not limited to the concepts of "digital archive," "Designated Community," "transparency," and "digital repository." Despite the universal purpose of the OAIS reference model, it is well-established that the reference model is "one built on OAIS concepts, not an OAIS suite of standards" (Lavoie, 2014) which has caused some issue with determining what "OAIS-compliant" actually means.

Though it is widely used, the question of relying so heavily on OAIS has been addressed by digital preservation community members. Priscilla Caplan, Director of the FCLA, articulated this issue in her talk on the process of building the dark archive DAITSS. She states:

> "It isn't easy to find a preservation repository that doesn't claim to be compliant with the OAIS reference model. This is a big of a bugaboo of mine because I haven't really seen too many OAIS-compliant applications. Part of the problem may be that OAIS itself doesn't provide much help in defining compliance" (2004).

The context for her presentation is an outline of how the structure of DAITSS is based on OAIS and the process by which DAITSS was designed to do so. Even so, the fact that she provides a

step-by-step description of OAIS compared to DAITSS and still recognizes that OAIS can blur its own responsibilities is enough to encourage developers to consider the language of audit metrics carefully, especially in the context of their own goals in comparison.

From a more critical perspective, David Rosenthal wrote "The case for a revision of OAIS" (2014) after the CLOCKSS Archive was audited per TRAC metrics by the CRL. Rosenthal briefly touched on several issues, the three most relevant here being that CLOCKSS is a dark archive and therefore cannot establish a future Designated Community; second, that CLOCKSS has a "distributed implementation," which OAIS fails to address; and finally that CLOCKSS "contracts-out its operations," which is also not covered at all by OAIS (2014). Without undermining the usefulness and impact of the OAIS reference model, the reliance on other standards does limit how dark archives, consortia, and project developers approach the use of existing metrics.

Reinforcing Caplan's comment above, TRAC and DSA both establish OAIS as their source for terminology and framework. TRAC more heavily relies on OAIS than DSA, and directly states that "key terms in this document have been adopted from the OAIS Reference Model" (CRL & OCLC 2007). The entire section B. Digital Object Management is grouped based on "well-known OAIS functional entities" and is cited throughout the section's criteria descriptions as a resource. DSA references OAIS noticeably less and offers it as a resource for guideline 9. "Documented storage procedures," and in guideline 15. "Technical infrastructure."

Documentation & Definitions

Building on the issue of primary reliance on the OAIS reference model is the issue of how terminology is defined in auditing metric documentation. Defining major digital preservation concepts is useful both for providing a foundation for developing projects, but also to evolve these concepts so that they align with what developers need and with how existing repositories define them based on their purpose. This does not imply that such definitions do not exist, but they are not included in the documentation of leading digital preservation standards.

For example, basic auditing metric documentation like TRAC and DSA tend to outline the foundations and process of developing the metric and then define the key terms and concepts. Though audit criteria are the bulk of a metric's documentation, the section describing the purpose and scope of the metric contains definitions for concepts such as "digital preservation,"

"digital repository," and other key terms in digital preservation. As previously discussed, most of these definitions are based on the OAIS reference model. However, most of the terms are often only defined in the context of the metric's documentation so that they are understood when reading the metric's documentation. The issue is that only defining a term in the context of that particular documentation limits the metric as a planning tool because it is not all-inclusive. An example of a problematic term in the context of audit standard documentation is "Service." This is a concept that could appear obvious in terms of the purpose of a digital repository, but metric documentation often overlooks the term. Neither DSA or TRAC directly define a "service" and it is only partially defined by nestor and ISO.

Arguably, the definition of a service could be implied by the existence of a mission statement that is supported by establishing a designated community. Both a mission statement and the designated community(ies) are required by TRAC and DSA, but there is a disconnect between the two because they provide paradoxical support to one another. TRAC and DSA requires evidence of mission statements similarly. TRAC states in criterion A1.1 "Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information" (CRL & OCLC, 2007). DSA Requirement 1. "Mission/Scope" states, "The repository has an explicit mission to provide access to and preserve data in its domain" (DSA, 2016). Neither definition include providing a service to a designated community, and neither cite a description of the designated community as evidence of a mission statement. Similarly, TRAC section A3.1 "Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met" (CRL & OCLC, 2007), which implies that the designated community dictates the service but does not explicitly define a service. DSA only refers to the concept when implying, and in some ways confirming, that the repository is providing a service.

The discussion of "service" as a concept is relevant here because the project team for Digital Safe refer to it in terms of becoming a service, rather than a repository. This identification has evolved, as originally the Electronic Archives Pilot Project mission statement was to "establish the feasibility of a working electronic archive for the use of the whole of the Collegiate University" (Wilson, J., 2012). When the project named the itself and the final product "Digital

Safe," the mission statement changed to a project goal to "deliver a secure, long-term records archiving service" (Jefferies, et al., 2016). This service-oriented goal raises a question on the difference between a repository and a digital preservation service, and is further discussed in the Future Work section.

The notion of a service is just one example of how metric standards only provide partial definitions for common terms, particularly those that are newer and that lack universal use. Digital Safe will need to be a dark archive in order to maintain restricted access and high security, and dark archives are neither defined nor explicitly included in metric documentation. While conducting these case studies it was necessary to overlook the definition discrepancies and rely on provided terminology.

Recent Updates

Arkivum and DSA both issued updates after the first case study in July 2016 and before the second case study in March 2017. Arkivum altered their documentation and available products in January of 2017. For consistency between each case study the previous products and documentation are referred to in this assessment with the understanding that it does not represent Arkivum's current products. DSA also released their newest 2017-2019 version in November 2016. Assessing based on the previous version when the newest version is available would not be an accurate representation of DSA, thus the current Requirements are utilized in this assessment. Due to the recentness of these guidelines, all of the published certifications are in the format of the previous 2014-2017 guidelines. Responses in published certifications are valuable as references, but the newest version has regrouped and renamed the guidelines so that matching up the previous and current guidelines is difficult.

Additional Metrics

Extending the case studies to additional metrics is not possible with the given timeframe and manpower. Building on the DSA assessment with nestor guidelines would have been useful and provided more insight on trending metrics and their incremental relationship with one another. It would also be useful to test a simple metric like the Ten Principles developed by the Center for Research Libraries (2007) as a preliminary assessment for scoping and planning. Conversely, applying an ISO to a non-traditional repository would also yield valuable results. ISO 16363 is ideal, but even assessing Digital Safe based on the ISO 27000 series on information security

would be useful for dark archives that are low-access and high-security. See the Discussion section for brief comparison of TRAC criteria and the ISO 27000-series. Also see the Future Work section for additional discussion on potential case studies.

Discussion

TRAC vs. DSA

General Use

In a general context, using TRAC and DSA provide different experiences. First, TRAC and DSA are clearly delineated by their complexity. TRAC contains 84 guidelines whereas DSA contains 16 guidelines. Both have the potential to be repetitive, though TRAC is more so given its tiered structure. A tiered structure in this case refers primarily to the relationships between subsections of TRAC section B. Digital Object Management. The first grouping is B1 and B2, where B1 describes the acquisition of ingests and B2 describes the actual ingest process. Similarly in the second grouping, B3 describes the documentation of preservation strategies, B4 and B5 describes the "minimal conditions for performing long-term preservation" and the metadata necessary for performing the preservation strategies, and B6 describes the documentation and evidence of disseminating information (CRL & OCLC, 2007). The relationships here are the similar evidence needed for each subsection in their respective groupings. Documentation is necessary for the subsections requiring a description, but is also considered partial evidence for subsections requiring evidence of successful implementation.

Furthermore, there are several reasons why DSA is currently more popular, and thus has been used to certify more repositories, than TRAC; some are obvious just by the definitions provided in each audit metric. According to the European Framework for Audit and Certification of Digital Repositories, there are three levels of audit certification: Basic Certification is a self-assessment; Extended Certification is an externally peer-reviewed self-assessment that is more comprehensive and builds on the Basic Certification, such as the nestor Seal;, and Formal Certification is an official certification from ISO 16363, DIN 31644, TRAC, or other equivalent audit (2008). That the levels build upon themselves is enough to understand that formal certification begins with basic DSA certification and advances as the repository evolves. At the time of the original TRAC report for OIDLPP for the purpose of assessment, TRAC appeared to be the optimal metric because it provides a thorough checklist for what exactly Digital Safe has implemented, has in development, and what is essential to prioritize in the next project phase. Given the limitations previously described in the Methodology section, DSA Requirements would have been a more suitable place to begin assessing a project that is in-progress and has

restricted documentation like Digital Safe. The current DSA guidelines provide a straightforward method for tracking the level of documentation development, the level of implementation strategies and their development, and what guidelines are theoretical versus in progress or already established.

Time is also an obvious factor. Usually a basic DSA certification requires approximately six months to complete internally. Conversely, in his blog series recording the TRAC Audit of CLOCKSS, David Rosenthal recorded that though the official audit was conducted from September 2013 to May 2014, they actually signed the contract in July 2013 and began collecting documentation six month previous in January 2013 (Rosenthal, 2014). Further, the HathiTrust Audit required 11 months of official audit time to complete, though as a condition of certification, HathiTrust agreed to address any issues revealed by the audit after the official report was released. Additional time spent remediating TRAC's issues was not documented, but would have pushed audit completion and certification to over a year. Internal audits can also consume several months from a project group, as noted by informal audit report generated by the Northern Arizona University's Cline Library, which indicated that it took the full spring and summer semesters of 2014 to complete an internal report (Welch & Phillips, 2014).

Use as Development Tools

Because the current standards for digital preservation exist only as assessment and audit tools rather than scoping and development tools, employing them as development tools offers a different user experience. Digital Safe is an ongoing project with three different entities contributing to the final service that need to be considered. TRAC was especially useful in delineating the responsibilities of the Digital Safe Steering Committee, Arkivum, and Archivematica. All the entities were included in each subsection regardless of whether they had a direct responsibility. In general, section A. Organizational Infrastructure is largely the responsibility of the Digital Safe Steering Committee to develop, especially in terms of finalizing a cost model and business plan for themselves as a guide for the Designated Communities, and in developing internal policies, as well as further in the future as they plan to transfer control of Digital Safe from the Steering Committee to BDLSS IT Services. Further, Archivematica is primarily responsible for section B. Digital Object Management because any preservation strategies that are not simple ingest and dissemination will take place in Archivematica interface,

though Arkivum does provide some support in this section in regards to DIPs and fixity checks. Finally, Arkivum is almost exclusively responsible for section C. Technologies, Technical Infrastructure, & Security as they provide the storage space, security measures, and contingency plan. The justification for choosing these technologies is determined by the Steering Committee, but even this decision is reinforced by Arkivum's services and documentation, and eventually will be evident in the final contract. See Appendix B for more detail on the specific responsibilities in each subsection.

TRAC also has the benefit of extreme detail. The 3 major sections have 14 topics and 84 subsections between them that all contain a description of and evidence for each subsection. The example evidence is invaluable because it informs Digital Safe what types of documentation they will need to consider creating, as well as how extensive the documentation will need to be. TRAC documentation also contains more information on additional ISO certifications that are related or can also act as evidence for TRAC criteria that might be starting points for future TRAC assessment plans.

Where TRAC provides a structure for delineating specific responsibilities for each aspect of Digital Safe and describing types of evidence and documentation, DSA provides a theoretical structure of digital preservation aspects, an official compliance scale, and flexibility for outsource partners and their responsibilities. DSA requirements are more descriptive and inform the user exactly what the requirement is while also including any related requirements. Each section contains a short description, long description, evidence, and the statement of compliance. DSA as a scoping tool is comparable to the NDSA Levels of Preservation (Phillips, Bailey, Goethals, & Owens, 2013) in that the guidelines offers broad categories without specific subsections and with a level of completion. This explanatory approach is useful for Digital Safe as an ongoing effort in that it ensures the Steering Committee is considering all of the major areas of digital preservation.

The added benefit of requiring the statement of compliance for each guideline is also useful for scoping and development in that it offers a means for creating a brief overview. In addition to stating the level of compliance from 0-4, it also requires a sentence for rationale. If a guideline is fully or partially implemented (levels 3 and 4), then DSA asks for a URL to documentation or other evidence. If a guideline is theoretically designed, not designed yet, or not applicable (levels

2, 1, and 0, respectively), DSA asks for a brief explanation (DSA, 2016). A brief overview of each requirement produced by the statements of compliance is extremely useful for presenting to funders and stakeholders and can serve as future documentation.

Consortia, Third-Parties, & Metrics

Metric Documentation

The impact of multiple collaborators on a single repository has, up until recently, been omitted from the leading digital preservation auditing metrics. However, the newest 2017-2019 DSA Guidelines are significantly more flexible, explicitly allowing for the inclusion of outsource partners and other third-parties. The Guidelines state: "If part of a requirement is 'Outsourced' to a third party (where applicable) identify the partner and provide evidence for the parts of the process you are responsible for and for those the Outsource Partner is responsible for" (DSA, 2016). This is a recent addition to the DSA Guidelines. The previous 2014-2015 Guidelines included a short section describing the stipulations for Outsourcing. The Outsourcing section states:

> "In the original version of the DSA outsourcing to third parties was permitted for Guidelines 4, 6, 7, 8 and 13 as long as the outsource partner had a DSA or better level of trust certification. To take account of the increasingly distributed and service-based nature of modern repositories, the DSA Board expanded the possibility of outsourcing to all Guidelines. This decision will be monitored over time and may be amended in future in cooperation with the DSA community. Applicant information relevant to outsourcing is requested in the 'Repository Context' section and must form part of the evidence for each applicable Guideline" (DSA 2014).

The newest guidelines have expanded on the original guidelines to include Outsource Partners as a section in the repository's Context section. In brief, this section asks for a list of any Partners, a description of the Partner's services, and copies of contracts and agreements. The newest caveat is that while the guidelines ask for a list of any certifications maintained by the Partner, the guidelines do not require that the Partners are certified to provide sufficient evidence. The 2017-2019 Guidelines state:

> "Because outsourcing will almost always be partial, you will still need to provide appropriate evidence for certification requirements that are not outsourced and for the parts of the data lifecycle that you control. […] We understand that this can be a complex

> area to define and describe, but such details are essential to ensure a comprehensive review process" (DSA 2016).

This is a significant move forward in the realm of auditing metrics. The inclusion of Outsource Partners' roles as acceptable evidence indicate that governing institutions that maintain digital preservation standards are acknowledging how most digital preservation initiatives need to outsource. In general, this opens the door for standardizing the audit and assessment of consortia. Including Partners' responsibilities forces the repository to fully understand their relationship with a vendor, and also allows the repository to absorb any certifications the Partner maintains. For Digital Safe and similar ongoing projects, it first encourages developers to not be hesitant to outsource when necessary, and second, it offers a framework for how to develop contracts with third parties. This framework is further described below.

Structure of the Results Section

In the case studies, the description for Outsource Partners provided in the DSA Guidelines proves more helpful than TRAC in establishing the roles of the vendors simply because DSA outlines how to describe a Partner and their role. Thus, in the above Results section there is a noticeable difference in how the assessment summaries are formatted. In the original TRAC Assessment report (see Appendix B), each criterion response is separated by the three potential partners Digital Safe, Arkivum, and Archivematica. There is nothing in the TRAC documentation or guidelines that indicates Outsource Partners or other third parties contributing to the building of a repository is acceptable as evidence, which makes it necessary to include all of the potential Digital Safe partners in the Results. As seen in Appendix B, it is clear which partners have a responsibility for each separate criterion, but each section and subsection as a whole vary in balance between each potential partner. This means that it is difficult to summarize each subsection of TRAC by each partner responsibility. Consequently, the results are first, described under the assumption that Digital Safe will be a single entity with three partners, and second, in terms of what Digital Safe as a single entity has and what it would need to meet TRAC criteria. Conversely, DSA requirements are all "equally weighted, standalone items" to prevent "duplication of evidence" where possible (DSA Board, 2016). Therefore, condensing the Requirements would not be accurate or balanced the way in which they were meant. This left a little more space for describing the roles of each partner on a broader level without the detail of TRAC. This is beneficial because it is both more exhaustive of different aspects of digital

repository management, such as the separation of "Documented storage procedures" (requirement 9) from "Preservation plan" (requirement 10) and the inclusion of where "Expertise" (requirement 6) is derived, but not so exhaustive that it is repetitive.

Auditing Parts or the Whole?

The culmination of this discussion about the case study results, metric documentation, and the role of third parties leads to the ultimate question of what the end goal is: should each respective partner of a repository be audited, as in the case for Goportis (Schwab, et al., 2017), or should the entire entity as a consortium or collaborative repository, such as CLOCKSS or the DRI, be audited? Given the results and previous discussion, the answer points to auditing the entity as a whole for several reasons. First, it is unreasonable to audit the entirety of a partner institution or organization when they only a play a role in some criteria. For example, Arkivum would not comply with criteria involving digital preservation workflow activities because its purpose is storage and long-term preservation, not in creating Archival Information Packages. On a different level, even if a third party does comply with certain criteria but in a way that is not relevant to the main repository being audited, then it has no place in the audit. Further, auditing each partner and ensuring compliance for every criteria only makes sense if each partner is involved with every aspect of the creation and management of the audited repository. This is not common even in a consortium environment because the purpose of a consortium is to increase access to resources by sharing, and as DSA Guidelines state, "outsourcing will almost always be partial" (DSA Board, 2016). Finally, the parts are intended to make a whole, meaning that the combined resources and services of several partners create a single repository, not multiple parts of a repository.

The purpose of an audit is to ensure a repository is trusted based on agreed upon standards and will remain trusted based on the repository's governance, policy, digital object management, combination of technologies, access to data, storage, and security. Standards exist to enforce best practices. If combining resources from multiple parties creates the optimum possible repository based on those standards, then the repository itself is of primary importance and the contribution of the third parties should be considered part of that repository and not a separate entity. The questions then change from how a collaborative repository should be audited, to where is line in utilizing a service, such as downloading and implementing the Fedora infrastructure, versus

outsourcing storage space, such as UIUC's contract with Amazon Glacier. This is an area that requires further exploration to determine, but the issues are apparent and will increase and more repositories outsource.

Use in Building Contracts

The argument against third parties in building and maintaining digital repositories is that relying on outside resources can increase the risk of losing sustainability should one or many of the third parties become an issue. Sharing control of a single repository with other institutions or relying on outsourced services is not a decision to be taken lightly by developers. This is the point where auditing metrics becomes invaluable because they dictate what documentation is necessary and what minimal expectations are in place for maintaining a trusted digital repository. Take TRAC section A5. "Contracts, licenses, & liabilities," where the subsections cover contracts and deposit agreement with the Designated Community that also include maintenance, access, withdrawal, and ownership rights. The evidence examples provided for this section are the basics features of building contracts. They include obvious documentation, such as copies of license and deposit agreements, copies of any contracts, and more specific documentations, such as "examples of legal advice sought and received," definitions of service levels, "citations for relevant laws and requirements," "policy on responding to challenges," among others (CRL & OCLC, 2007). This provides the framework for developing contracts and negotiating with vendors.

Comparably in DSA is their new section on Outsource Partners (see the Discussion subsection above on TRAC versus DSA) which directly asks the repository to "list the certification requirements for which the Partner provides all, or part of, the relevant functionality/service, including any contracts or Service Level Agreements in place" (DSA Board, 2016). This requires existing repositories to fully understand their relationship with any Partners, and also forces developing projects to decide exactly what they want from a third party.

Documentation & Testimony as Evidence

Without evidence, compliance to any metric cannot be proven. In the case of an on-going project or a project working towards maintaining confidential information like Digital Safe, evidence in the form of test cases and public policy has not yet occurred. However, documentation of various strategies and descriptions can serve as sufficient documentation, which is especially useful for dark archives, but also useful for communicating needs from vendors. A project undertaken by

Seamus Ross and Andrew McHugh at the Digital Curation Centre conducted six pilot audits on various types of archives using TRAC identified the various forms of evidence that are acceptable for meeting audit criteria, which included primary and secondary documentation. They state that "there are very few examples of check-list criteria compliance that can't be demonstrated at least to some extent with the provision of primary documentation" because it is broad and can cover physical and digital records as well as object metadata (2006). Ross and McHugh describe documentation, commitments, capacity, resource, and planning information can all be described to a certain extent in written documentation, which could include areas such as accounting documentation, infrastructure descriptions, various preservation and governance policies, and business plans can all be recorded in documentation and provide partial to full compliance (2006). Documentation is clearly limited by the lack of demonstration; describing the process does not guarantee success in practice, but it does offer insight into the level of planning and implementation that has been reached by the repository. For TRAC and DSA's 0-4 compliance scales, most documentation can provide enough evidence for criteria compliance to meet a 2, which indicates that a criterion has been theoretically designed (DSA, 2016).

In the case of Digital Safe, documentation was the primary evidence. Various project reports, projected business plans, standard license contracts, and a public blog are currently all that exist as stable evidence of Digital Safe as a whole, and all but the blog are still private. Additional documentation for the technologies Arkivum and Archivematica provided the entirety of evidence for criteria dedicated to digital object management, and partial evidence for criteria on governance and infrastructure, and for security and technologies. Without the technologies' public documentation, Digital Safe would not be able to meet most criteria of any audit at any level.

Further, Ross and McHugh also posit that stakeholder testimony can also provide a level of audit compliance, particularly if it is corroborated by other stakeholders because it offers both a "degree of credibility" when it also is reflected in the documentation (2006). Much of the contractual information with Arkivum, background on the project, and stakeholder interest was collected for Digital Safe through the interviews with current Phase 2 Project Lead Neil Jefferies, stakeholder Susan Thomas, and Phase 2 Project Manager David Tomkins, as described in the Methodology section. Jefferies and Tomkins have both produced documentation for Digital Safe,

and Susan Thomas has participated in several recorded meetings, which granted them credibility and full knowledge of the project. Jefferies is also the primary contact for Arkivum and was in the process of developing a customized contract with Arkvium that provided enhanced understanding of the relationship between Arkivum, and eventually University of Oxford, clients (N. Jefferies, personal communication, July 25, 2016).

Documentation and testimony cannot replace a supervised practical demonstration or interaction with an auditor, but in the case of the Digital Safe project, they are the primary tools for understanding the project's goals and how those goals influence their contract with Arkivum.

TRAC & the ISO 27000-series

One interesting result of examining Arkivum as an outsource partner is how to handle their ISO 27001 certification. In TRAC, an example of evidence for complying with the security criteria found in section C. Technologies, Technical Infrastructure, & Security, is to also maintain ISO 17799 "Information Technology – Code of practice for information security management" certification (CRL 2007). ISO 17799 was revised in 2007 and became ISO 27002:2013 as part of the ISO 27000 series on information security (ISO 27000 Directory, 2008). ISO 27002 provides the options for implementing the requirements in ISO 27001 and are meant to complement each other. Each subsection in TRAC section C3. Security can be met with ISO 27000 certifications. TRAC documentation also indicates that subsections C1. System infrastructure and C2. Appropriate technologies are also likely to meet compliance with partial support from ISO 27000 certification (OCLC, 2007). Because Arkivum maintains ISO 27001 certification, Digital Safe will have less to consider for section C3.

Future Work

Continuation & Follow-up

Two case studies assessing a single project is just a starting point for work in how digital preservation standards can also be used as scoping tools for outlining responsibilities. Applying other oft-used standards and guidelines to Digital Safe as an ongoing project is the clear next step. Beginning with guidelines like the Ten Principles (CRL) and the NDSA Levels of Preservation (2013) would establish a broad theoretical foundation. Once basic standards were examined, building on the DSA assessment to assess nestor's 37 guidelines for extended self-assessment would follow the European Framework for Audit and Certification of Digital Repositories' levels of certification (2008, July 8), and eventually lead to scoping based on ISO 16363:2012. Including all of these metrics offers a range of scoping experiences with the same data.

In the context of the Digital Safe project, the two case studies and their results all lead up to the point where Digital Safe is live and functioning and capable of being assessed or audited for certification. Reviewing the documentation from these cases studies will raise new questions: Did scoping out the project help develop a contract with Arkivum? How did the projected responsibilities of each party reflect the actual contract? How did the project evolve after examining the strengths and gaps revealed by the case studies? Are the case studies referred to in official documentation? Did the formal certification process (or self-assessment) draw from these case studies? The same questions may be raised by other projects documenting their use of standards for planning. Community efforts will likely be key as institutions and organizations report on their experiences planning repositories, working with vendors, conducting self-assessments, and undergoing formal certification processes.

On another level, communicating with vendors on their experiences in developing contracts with institutional repositories would provide insight on their perspective. Collecting information on their process and what they expect from clients could offer potential standards to incorporate into digital preservation metrics. Comparatively, interviewing consortia partners on their process for communicating with each other and with vendors would offer similar insight. This is relevant because while DSA requires contracts and licenses to be included for any Outsource Partner, the extent of how those contracts are created based on standards is not illustrated.

Inclusion of Outsource Partners

Given how common outsourcing and consortial collaboration is in building and maintaining digital preservation repositories, digital preservation metrics must adapt to include them. As previously discussed, the newest 2017-2019 DSA Guidelines are the only current metric that include outsource partners as expected for digital repositories, and view their roles and support as acceptable evidence. Other metrics do not include collaborators, partially due to the reliance on OAIS as discussed in the Limitations section, and partially because digital preservation standards are still evolving. Examining the next two year's DSA certifications will provide evidence supporting the inclusion of collaborators and third-parties into other metrics. As of April 2017 there have been no case studies or certifications per DSA 2017-2019 guidelines. However, adapting the three individual applications from the Goportis Digital Archive DSA certification into a single application based on the 2017-2019 Guidelines would be an interesting starting point for comparison. Additional certifications and re-certifications will also provide a collection of examples to offer support for other metrics to evolve.

Repository versus Service

An interesting aspect of the Digital Safe project is that it is consistently identified by the project team as a proposed service (Jefferies, et al., 2016). The discussion on what defines the notion of a service in the realm of digital repositories in the Limitation section introduces this, but further exploration into documentation for metrics and into the line between service and repository, if any exist, are necessary. The issue is that not all collaborative projects identify or should be identified as services. Europeana[37], for example, is a digital platform for cultural heritage focused on access, but it does not control the data; is it a repository because it holds over 3,000 institutions' material, or a service because its mission statement is that "We transform the world with culture! We want to build on Europe's rich heritage and make it easier for people to use, whether for work, for learning or just for fun," which makes its primary service discovery and access? In the context of Digital Safe, is it proposed to become a service because it is outsourcing its major technical infrastructure to third-party technologies and therefore not in complete control of the data, or is it a service because that is what motivated the project to initially develop? Further, does it matter that it is proposed to be a service and not a repository

[37] For more information on Eurpoeana's vision, see their About us page: http://pro.europeana.eu/about-us/

since it could be argued that a repository provides services? The only certainty here is that the Digital Safe project began because the University of Oxford needs universal storage infrastructure for confidential records, and the project team is working to produce a solution that works best for their allocation of resources. This does not mean that they should have to develop without standards for guidance. This also broaches on the question of whether new standards are needed to fully expand on the impact of Outsource Partners or consortium members on a digital repository, or if the current standards can be updated to include them. DSA has already incorporated them, but future certifications will determine if it was successful and what issue may arise.

Documentation Revision

In addition to incorporating collaborators as an equal partner in maintaining a repository, the observations on definition discrepancies in the Discussion section call for a revision of the language in digital preservation metrics. As a potential dark archive, the Digital Safe project already does not completely comply with digital repository metrics. For example, a digital repository, also commonly interchanged with digital archive, is broadly defined an "infrastructure through which to store, manage, re-use and curate digital materials" in which the primary responsibility is to provide "easy, simultaneous and remote access to deposits" (Semple, 2006). While current auditing metrics tend to focus on transparency in documentation and direct access to data, they overlook other types of repositories, namely dark archives and consortia. As defined by the Digital Preservation Coalition, a dark archive is an archive that cannot be accessed by any current users but may be accessible at future dates subject to the occurrence of specific pre-defined events. Access to the data is either limited to a few set individuals or completely restricted to all" (DCC). By definition, the storage infrastructure, access to deposits, and even access to documentation for an existing dark archive is not open-access and complicates certification. Further, as discussed, a consortium as an entity cannot achieve certification, but members of a consortium can be audited and certified individually (Schwab, Tunnat, & Gerdes, 2017). These gaps in digital preservation auditing metrics hinder the progression of digital preservation and require a revision of current standards. Providing definitions for additional repository types would also motivate revision of digital repository concepts, continue the evolution of standards.

Conclusion

The realm of digital repositories is complex. In the words of Dr. Sandra Collins, Director of the Digital Repository of Ireland, "a digital repository is a living, growing organism, that will face new challenges as the huge quantities of data grow even faster, as the complexity of the data increases, and as the requirements for sophisticated data visualisation, open data and research data management grow" (DRI, Grant, O'Neill, & Webbs, 2013). The Digital Safe case studies exemplify this complexity, but also help to offer a solution towards that complexity. The use of digital preservation auditing metrics can be used for scoping and development of digital repository projects and should not be limited to only certification of existing repositories. Building a repository based on established standards can help produce the most relevant documentation, inform projects on how to write contracts with vendors, and better prepare the future repository for formal audits. The growing number of collaborative digital repositories calls for a update in standards on how to negotiate contracts with vendors and third parties, and to delineate roles and responsibilities within consortia. Collaboration is key to development, and our digital preservation standards need to reflect that collaboration in order for digital repositories to evolve.

References

Arkivum Ltd. (2017). *About us*. Retrieved from http://arkivum.com/about-us/

Arkivum Ltd. (2015, June). *Arkivum/1+1*. Retrieved from http://arkivum.com/arkivum1plus1/

Arkivum Ltd. (2015, June). *Arkivum/100*. Retrieved from http://arkivum.com/a100/

Arkivum Ltd. (2014, Feb. 27). *Frequently Asked Questions (FAQs)*. Retrieved from http://arkivum.com/wp-content/uploads/2014/04/Arkivum-Frequently-Asked-Questions.pdf

Arkivum Ltd. (n.d.). *Higher Education*. Retrieved from http://arkivum.com/he/

Artefactual Systems Inc. (n.d.). Archivematica documentation 1.6. *Artefactual Systems Inc*. Retrieved from https://www.archivematica.org/en/docs/archivematica-1.6/

Artefactual Systems Inc. (n.d.). Archivematica storage 0.10. *Artefactual Systems Inc.* Retrieved from https://www.archivematica.org/en/docs/storage-service-0.10/

Artefactual Systems Inc. (n.d.). Format Policy Registry 1.1.0. *Artefactual Systems Inc.* Retrieved from https://www.archivematica.org/en/docs/fpr/

Artefactual Systems Inc. (n.d.). FAQ. *Artefactual Systems Inc.* Retrieved from https://www.archivematica.org/en/docs/archivematica-1.5/user-manual/troubleshooting/faq/#faq

Artefactual Systems Inc. (n.d.). *Team*. Retrieved from https://www.artefactual.com/team/

Bishoff, L., The Bishoff Group, & Rhodes, E. (2007). *Planning for digital preservation: A self-assessment tool*. Retrieved from https://www.nedcc.org/assets/media/documents/DigitalPreservationSelfAssessmentfinal.pdf

Caplan, P. (2004, Nov.) How to build your own dark archive (in your spare time). [PDF document]. Retrieved from https://libraries.flvc.org/documents/181844/502298/How+to+Build+Your+Own+Dark+Archive/f3cd50ef-1fe4-48c8-85e3-5874ba1feab3

Center for Research Libraries (2014, July 1). CLOCKSS Audit Report 2014. Retrieved from https://www.crl.edu/sites/default/files/reports/CLOCKSS_Report_2014.pdf

Center for Research Libraries (n.d.). *Digital preservation metrics.* Retrieved from https://www.crl.edu/archiving-preservation/digital-archives/metrics

Center for Research Libraries (2011, Mar. 1). HathiTrust Audit Report 2011. Retrieved from

http://www.crl.edu/sites/default/files/reports/CRL%20HathiTrust%202011.pdf

Center for Research Libraries (n.d.). *Other assessment tools*. Retrieved from https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/other-assessment-tools

Center for Research Libraries (2007, Jan.) *Ten Principles*. Retrieved from https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re

Center for Research Libraries & Online Computer Library Center, Inc. (2007, Feb.). *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Retrieved from https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf

CLOCKSS (n.d.). How CLOCKSS works. *CLOCKSS*. Retrieved from https://it.engineering.illinois.edu/services/file-and-data-storage

Data Curation Centre (2008). *Objectives of the DRAMBORA process*. Retrieved from http://www.repositoryaudit.eu/objectives/

Data Curation Centre (n.d.) *History of the DCC*. Retrieved from http://www.dcc.ac.uk/about-us/history-dcc/history-dcc

Data Seal of Approval (n.d.) *About.* Retrieved from https://www.datasealofapproval.org/en/information/about/

Data Seal of Approval (2016, Nov. 10). *Guidelines version 2017-2019*. Retrieved from https://assessment.datasealofapproval.org/guidelines_54/pdf/

Data Seal of Approval (2013, July 19). *Guidelines version 2*. Retrieved from https://www.datasealofapproval.org/media/filer_public/2013/09/27/guidelines_2014-2015.pdf

Data Seal of Approval (n.d.) *List of repositories that have acquired the Data Seal of Approval*. Retrieved from https://assessment.datasealofapproval.org/

Data Seal of Approval Board (2015, June 30.). Implementation of the Data Seal of Approval for the Digital Repository of Ireland. Retrieved from https://assessment.datasealofapproval.org/assessment_146/seal/pdf/

Data Seal of Approval Board (2015, Sep. 15). Implementation of the Data Seal of Approval for the Goportis Digital Archive – German National Library of Economics (ZBW). Retrieved from https://assessment.datasealofapproval.org/assessment_158/seal/pdf/

Digital Preservation Coalition (2015). *Digital preservation handbook, 2nd Edition.* Retrieved from http://www.dpconline.org/handbook/institutional-strategies/audit-and-certification

Digital Repository of Ireland, Grant., R., O'Neill, J., & Webb, S. (Eds.), (2015). *Building the Digital Repository of Ireland infrastructure*. Digital Repository of Ireland. DOI: http://dx.doi.org/10.3318/DRI.2015.5

Engineering IT (n.d.). File and data storage. *Engineering IT Shared Services.* Retrieved from https://it.engineering.illinois.edu/services/file-and-data-storage

European Framework for Audit and Certification of Digital Repositories (2008, July 8). Trusted Digital Repositories. Retrieved from http://www.trusteddigitalrepository.eu/Trusted%20Digital%20Repository.html

HathiTrust (2011, March). *HathiTrust Trustworthy Repository Audit and Certification (TRAC).* Retrieved from https://www.hathitrust.org/trac

Houghton, B. (2015). Trustworthiness: Self-assessment of an institutional repository against ISO 16363-2012. *D-Lib Magazine, 21(3/4).* DOI: 10.1045/march2015-houghton

ISO 27000 Directory (2008). *Introduction to ISO 27002 (ISO27002).* Retrieved from http://www.27000.org/iso-27002.htm

Jefferies, N., Hicks, B., & Rendell, S. (2016, Feb. 26). *Digital Safe project initiation documentation* (Rep. 0.7). Oxford: Bodleian Digital Library Systems & Services.

Jefferies, N. & Tomkins, D. (2014, July 10). *Digital Safe: Archiving digital records for the long term* [PowerPoint slides]. Retrieved from https://digitalsafe.wordpress.com/2014/08/18/digital-safe-presented-at-the-ictf-conference-10-july-2014/

Lavoie, B. (2000). Meeting the challenges of digital preservation: The OAIS reference model. *OCLC Newsletter, No. 243:26-30.* Retrieved from http://www.oclc.org/research/publications/library/2000/lavoie-oais.html

Lavoie, B. (2014, Oct.). The open archival information system (OAIS) reference model: Introductory guide (2nd edition). *DPC Technology Watch Report*. Retrieved from http://www.dpconline.org/docman/technology-watch-reports/1359-dpctw14-02/file

McGovern, N. (2013). TRAC Review in Drupal. *MIT Libraries*. [PowerPoint slides]. Retrieved from http://ils.unc.edu/digccurr/curategear2013-talks/mcgovern-curategear2013.pdf

Mitchan, J. & Hardman, C. (2011). *ADS and the Data Seal of Approval – case study for the*

DCC. Retrieved from http://www.dcc.ac.uk/resources/case-studies/ads-dsa

Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013, April 5). *The NDSA levels of digital preservation: An explanation and uses*. National Digital Stewardship Alliance. Retrieved from http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf

Rosenthal, D. (2014, Aug. 5). TRAC audit: Process [Web log comment]. Retrieved from http://blog.dshr.org/2014/08/trac-audit-process.html

Rosenthal, D. (2014, Aug. 12). TRAC audit: Lessons [Web log comment]. Retrieved from http://blog.dshr.org/2014/08/trac-audit-lessons.html

Rosenthal, D. (2015, Oct.). The case for a revision of OAIS. *Digital Preservation Coalition*. Retrieved from the DPC Wiki: http://wiki.dpconline.org/index.php?title=The_case_for_a_revision_of_OAIS

Rosenthal, D., Robertson, T., Lipkis, T., Reich, V., & Morabito, S. (2005, Nov.). Requirements for digital preservation systems: A bottom-up approach. *D-Lib Magazine, 11(11)*. Retrieved from http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html

Ross, S. & McHugh, A. (2006, Nov. 27). Preservation pressure points: Evaluating evidence for risk management. (A paper presented at iPRES 2006, 8-10 October 2006, New York, USA). Retrieved from http://www.repositoryaudit.eu/images/PreservationPressurePoints.pdf

Schwab, F., Tunnat, Y., & Gerdes, T. (2017, Feb. 7) Consortial certification processes: the Gorportis Digital Archive—A case study. [Web log comment]. Retrieved from https://saaers.wordpress.com/2017/02/07/consortial-certification-processes-the-goportis-digital-archive-a-case-study/

Semple, N. (2006). "Digital Repositories". DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/digital-repositories

Sites, M. (2013, Apr. 22). The APTrust Story: A collaborative model for digital preservation. Retrieved from http://aptrust.org/aptrust-admin/resources/the-aptrust-story.pdf

Stanbridge, N. (2016, Jan. 21). New digital preservation solution from Arkivum, shaped to grow with your data. *Arkivum Limited*. Retrieved from http://arkivum.com/blog/perpetua-digital-preservation/

Welch, T. & Phillips, K. (2014). *Trustworthy repositories: Audit and certification (TRAC) Cline*

Library internal audit, Spring 2014. Retrieved from http://library.nau.edu/speccoll/pdf/TRAC_report_draft_Final.pdf

Wilson, J. (2012, Apr. 17). Electronic archive pilot project. Retrieved from http://archivepilot.oucs.ox.ac.uk/index.xml

Wu, M. (2015, Sept./Oct.). The future of insistutional repositories at small academic institution: Analysis and insights. *D-Lib Magazine*, 21(9/10). DOI: 10.1045/september2015-wu

Appendices

Appendix A: Acronym Dictionary

AIP: Archival Information Package

APTrust: Academic Preservation Trust

BDLSS: Bodleian Digital Library Systems & Services

BEAM: Bodleian Electronic Archives and Manuscripts

CCSDS: Consultative Committee for Space Data Systems

CLOCKSS: Closed Lots of Copies Keeps Stuff Safe

CRL: Center for Research Libraries

COPTR: Community Owned digital Preservation Tool Registry

DAITSS: Dark Archive in the Sunshine State

DHOxSS: Digital Humanities Oxford Summer School

DANS: Data Archiving and Networked Services

DCC: Data Curation Centre

DIN: Deutsches Institut für Normung (trans. German Institute for Standardization)

DIP: Dissemination Information Package

DPC: Digital Preservation Coalition

DPE: DigitalPreservationEurope

DPN: Digital Preservation Network

DRAMBORA: Digital Repository Audit Method Based on Risk Assessment

DRI: Digital Repository of Ireland

DSA: Data Seal of Approval

FCLA: Florida Center for Library Automation

HTRC: HathiTrust Research Center

ICSPR: Inter-university Consortium for Political and Social Research

ISO: International Standards Organization

NDCC: Northeast Document Conservation Center

NDSA: National Digital Stewardship Alliance

OAIS: Open Archival Information System

OCLC: Online Computer Library Center

OeRC: Oxford e-Research Center

OIDLPP: Oxford-Illinois Digital Libraries Placement Program

ORA-Data: Oxford Research Archive Data

PID: Project Initiation Documentation (UK)

POWRR: Preserving (Digital) Objects With Restricted Resources

SIP: Submission Information Package

TDR: Trusted Digital Repository

TIB: German National Library of Science and Technology

TRAC: Trustworthy Repositories Audit & Certification

UIUC: University of Illinois at Urbana-Champaign

Appendix B: An Informal TRAC Audit of Digital Safe at the University of Oxford

August 2016, Version 1.0

The following is the full official report submitted to the OeRC in August 2016. This appendix is included to serve as evidence of observations on TRAC and as a reference to any specific points made. The report includes the original abstract, background and methodology, TRAC assessment results, and recommendations. The TRAC assessment itself includes a response to each criterion of the TRAC checklist, recommendations based on those responses, and is based on the documentation that was available as of August 2016. It has not been further edited since the original report was submitted and does not include updated information from any new documentation from Digital Safe or the technologies. The References listed at the end are the original references for the original OIDLPP project and are not References for this thesis. To avoid confusion between sections of the same name between this thesis and this appendix, the major headings in this appendix are italicized.

Abstract

In addition to the treasures and data held by the Bodleian libraries that are in the process of being digitally preserved, there is an urgent need for unified, long-term preservation of University of Oxford records. Administrative, financial, medical, and personal records are increasing rapidly on a daily basis. Digital Safe, created by the BDLSS, has sought to solve this problem by outsourcing the data management technology to Arkivum and managing it locally. This will allow all 38 colleges and other various departments to have a single technology for secure storage. Priority has recently been placed on completing Digital Safe, which calls for a final review of its trustworthiness, sustainability, and infrastructure. Auditing a Dark Archive focused on conducting a comprehensive, informal audit for Digital Safe. TRAC was used as the guidelines for its comprehensiveness, but is not fully effective for a dark archive. This endeavor is challenging as there is no metric to evaluate a dark archive or a project approach, and the audit included the service as well as its technologies.

Introduction

While traditionally it is looked to the library to manage the cultural content worth preserving and storing, there is an excess of data that is often overlooked and in desperate need of infrastructure and security. Student records, financial records, patient records, non-anonymized data, and other data with any personally identifiable information is considered high-security, high-priority, and increases every day. The University of Oxford does not yet provide a service for the storage of high-security information. Additionally, among the independent colleges, departments, and universities within the University of Oxford there is no shared infrastructure, technology platform, or methodology.

Often this data has requirements to be maintained for a certain amount of time. Financial records, for example, might need to be kept for compliance reasons for a certain number of years before being permanently destroyed. Comparatively, student records may eventually be opened to the public for research, but as they contain information about still-living persons and their families, they must be stored without access for a longer period of time. This data may also consist of acquisitions of the library that does not yet have the funding to review and remove personally identifiable information. These materials require a short to long-term storage space with strict access management until they can be handled. The solution at the University of Oxford for this ever-increasing data is Digital Safe.

Digital Safe Background

Digital Safe was originally the Electronic Archives Pilot Project. Initiated in 2012, the project consists of two completed phases and a third phase currently on hold but still in development. The ultimate goal of Digital Safe is to "deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis"[38]. The purpose is to have a single technology that is managed locally and customizable by the college that also is cost-effective and easily managed.

Phase 1

Phase 1 (2012-2013) focused on defining the scope of the project. The primary purpose was to determine if there was a need or want for a digital preservation service for the storage of administrative material. The project interviewed various staff members in the colleges, IT Services, the Oxford Colleges Librarians Group, Oxford Archivists Consortium, and other related groups, committees, and departments. Phase 1 established a desire for such a service and gathered some of the technical, security, and infrastructure requirements that these institutions might have. Phase 1 also verified that most of the University is relying on the library to provide a solution. More detailed information may be found in the Phase 1 Project Initiation Document.

Phase 2

Phase 2 (January 2013-June 2014) investigated service and infrastructure models for a long-term digital preservation service. The project team examined the infrastructure that the Bodleian Electronic Archives and Manuscripts (BEAM) currently uses, the infrastructure for ORA-Data, and services such as DataBank for possible reuse or outsourcing. BEAM infrastructure, developed in 2005, is held on a stand-alone server that has recently not been able to keep up with the increase in acquisitions and the preferred level of organization and security that the BEAM would prefer[39]. ORA-Data was found to not be compatible for the type and amount of security measure that would need to be implemented. Building an entirely new

[38] Digital Safe Phase 3 PID, page 2
[39] ST interview

infrastructure was also investigated but would not have been time or cost-efficient. DataBank was also ruled out for not including all of the aspects necessary for the project in one platform.

Phase 2 determined that the users would likely be the following: College Archivists; University Archive; Central Administrative Records Management; Departmental Research Records Management; and Personal Material held by BEAM[40]. This should not limit the data that can be added. Given that these users may not be advanced technical users, the interface design needs to be user-friendly and technical support is necessary. The ultimate decision was to outsource the technologies and manage the service locally. The name also changed in Phase 2 from Electronic Archives Pilot Project to Digital Safe. The technologies chosen were Arkivum for long-term, high-security storage, and Archivematica for digital preservation activities. More detailed information can be found in the Phase 2 Project Initiation Document and in the Phase 2 End Project Report. A Digital Safe blog[41] was also developed in 2013 by David Tomkins and provides public information on Phases 1 and 2. The blog will be continued in Phase 3.[42]

Phase 3

Phase 3 has been developed but has not yet received funding to continue. Funding is requested for three major activities. First, to further investigate and fully develop an ideal contract with the outsourced technologies. Second, to design and implement the business and service models within IT Services. Finally, to cover the start-up, training, and storage costs of one year of operating the service. Once the service has been deployed to early adopters, the project team will track user feedback and service function for one year before deciding to launch a service available to the entire University.[43] The original goal was to launch the service for early adopters by September 2016, and while funding may extend this, the goal is to launch the service as soon as possible after funding is granted. More detailed information on the projected budget and project roles can be found in the Phase 3 Project Initiation Document.

A successful Phase 3 will result in a universal technology and infrastructure for long-term, large-scale, high-security data produced and held by the University of Oxford. This service will have training built in and materials available between training sessions, will be customizable by each client, and is based on a cost-recovery model. The Oxford brand and the large number of clients that the University would be bringing to Arkivum is a motivation for Arkivum to work with Digital Safe and develop less expensive start-up and training fees.[44] Additionally, Archivematica is a built-in tool in Arkivum. In signing with Arkivum the University would only be paying for a single Archivematica license fee rather than over forty individual fees. Finally, the success of Digital Safe would also ease the burden of data managers at the University of Oxford as the service would be user-friendly, the technology would be supported by their developers, and minimal stress will be put on the IT Services and the future Steering Committee by handling small issues and outsourcing larger issues.

Technologies

In brief, Arkivum is a "long-term, large-scale managed data storage"[45] that provides high-security processing and storage with strict accessibility processes. They have two products that may be utilized by

[40] Digital Safe Phase 3 PID, page 2
[41] Digital Safe blog
[42] NJ interview
[43] Digital Safe Phase 3 PID
[44] NJ interview
[45] Arkivum "About Us"

Digital Safe: Arkivum/1+1 and Arkivum/100. Arkivum/1+1[46] has one digital copy of the data held in a secure location, and one physical copy held on LTO data tape held in Escrow at a separate location. Arkivum/100[47] has two digital copies of the data held in two geographically separate, secure locations, and one physical copy held on LTO data tape held in Escrow at a third separate location. Arkivum/100 also offers the 100% integrity guarantee by ensuring three copies are being managed and preserved. Arkivum services revolve around the number of pipes being used. Pipes are ingest workflows that can host multiple archives by one client and multiple clients. Each login can have customized workflows, though only one login can be active at one time. Audit trails are available, and all of the data is only accessible by the administrator login via an encryption key. If the master encryption the data cannot be retrieved, even by University staff or IT Services, which ensures the security of the data. More detailed information may be found in the Phase 3 PID[48], Arkivum's website[49], and Section B of the informal TRAC Audit.

Archivematica is an open source digital preservation workflow tool developed by Artefactual that has recently been built in to Arkivum.[50] Arkivum specifically only "provides safe and secure data archiving,"[51] not digital preservation activities. Archivematica can offer a customizable digital workflow tool, including SIP creation, normalization, AIP packaging, and DIP uploading. This is then ingested into Arkivum. More information can be found in Archivematica's documentation[52] and in the informal TRAC Audit.

Audit Purpose

Recently there has been an increased interest in the project resuming, both within the library and from members of the University. This calls for a review of the current goals of the project, its documentation, and the technologies chosen. The goal of the Oxford Illinois Digital Library Placement Program (OIDLPP) is to assist in promoting Phase 3 by informally auditing the project approach as a whole and the chosen technologies to ensure that they will produce a trustworthy dark archive. This audit will highlight the strengths of the project to assist in securing funding, as well as locating areas of improvement to focus on developing in Phase 3, simultaneously reviewing and scoping Digital Safe.

Project Limitations

It is important to note that this project was completed during a limited time frame, which leads to subsequent restrictions. First, the time spent on this project was roughly 3.5 weeks. This time consisted of understanding the project, reviewing documentation, understanding the technologies and their documentation, interviewing project team members and stakeholders, and learning how to use a repository metric efficiently. The time constraint also means that the project is not as detailed as is ideal, but it is comprehensive and easily built upon. Rather than performing and true, formal audit on Digital Safe, an informal audit was performed and acts as a guidance for organization and thorough investigation rather than a formal, published report.

---

[46] Arkivum/1+1
[47] Arkivum/100
[48] Digital Safe Phase 3 PID
[49] Arkivum website
[50] Stanbridge, Nik, "New digital preservation solution from Arkivum, shaped to grow with your data"
[51] Arkivum FAQ, page 23
[52] Archivematica documentation

Furthermore, the OIDLPP project goal was to audit an approach to a service that first, does not exist yet, and second, has limited user feedback and no examples to measure. This may be remedied after the early adopters have time to give feedback. Finally, there is no metric for evaluating a project approach, nor is there an updated, universally accepted metric for evaluating a dark archive. This is understandable given the early stages that digital preservation and storage are in, but does cause some difficulty in attempting to measure Digital Safe.

Methodology: TRAC

In consulting Michael Popham, Head of Digital Collection and Preservation Services for BDLSS; Neil Jefferies, Research & Development Project Manager for BDLSS and project leader for Digital Safe; and the Center for Research Libraries' assessment tools, TRAC was chosen as the metric to compare to Digital Safe. Trustworthy Repositories Audit & Certification (TRAC) was released by OCLC in 2007, and is what the ISO 16363 standard (aka the Trusted Digital Repository Checklist) is based on.[53] It is significantly more detailed than the Data Seal of Approval or the Ten Principles, but is less complicated than ISO 16363. This allows for a compromise in complexity.

The word "repository" is TRAC's choice of word to describe the collection of materials. There is some controversy surrounding this word, particularly in the United Kingdom, because a digital repository is by definition an open collection. Additionally, a dark archive tends to hold different content than a repository, have a different business model, and requires higher security. For the purposes of this project, the word repository should be replaced with dark archive and not acknowledged in the traditional definition in this report.

TRAC Limitations

As its name implies, TRAC is built for auditing repositories, not for dark archives or approaches to digital preservation services. TRAC is also meant to be used on a single repository. In the case of the University of Oxford, each college, department, institution, and so on, would have their own archive with different permissions, workflows, and policies. The chosen technologies, Arkivum and Archivematica, also have separate documentation that may overlap with each other or with Digital Safe policies, or may not apply to TRAC at all.

TRAC also calls for accessibility and transparency. A dark archive by definition is not widely accessible. While Digital Safe may provide some information on the type of content that might be included, who has access and why, any more specific information is not appropriate. Each individual college will also likely not publish their workflows and policies added to the basic ones developed by Digital Safe.

Purpose of this Document

This document is a combination of the background and context of the Digital Safe project, the process of producing the audit, and the results of the informal TRAC audit.

Vocabulary

- Client: Client is the language chosen to indicate a member of the University of Oxford who is utilizing the Digital Safe service. It removes any assumption of who will be using the service, particularly if they fall outside of the primary users identified in Phase 1 (See section A3.1).
- DS: refers to Digital Safe's responsibilities as outlined in the informal audit

[53] Center for Research Library Metrics

- Electronic Archives Pilot Project (EAPP): This is the original title of the project, which changed during Phase 2 to Digital Safe
- NJ: refers to Neil Jefferies, project lead for Phase 3 of Digital Safe
- ST: refers to Susan Thomas, Head of Archives & Modern Manuscripts and stakeholder in Digital Safe

Rating System

Nancy McGovern of MIT Libraries developed a TRAC review tool in Drupal that allows institutions to self-review themselves. The tool was built in 2013 but is currently only hosted on Archivematica and also requires a DRUPAL installation. This installation and the lack of documentation or examples of the tool omitted its use in the methodology. However, one of the features is a rating system for TRAC compliance, which is utilized in this informal audit. This system provides a straightforward review of each section, and is included in this audit. The rating system as described on Archivematica's TRAC review tool page:

- 4 = fully compliant - the repository can demonstrate that has comprehensively addressed the requirement
- 3 = mostly compliant - the repository can demonstrate that it has mostly addressed the requirement and is on working on full compliance
- 2 = half compliant - the repository has partially addressed the requirement and has significant work remaining to fully address the requirement
- 1 = slightly compliant - the repository has something in place, but has a lot of work to do in addressing the requirement
- 0 = non-compliant or not started - the repository has not yet addressed the requirement or has not started the review of the requirement

Audit Structure

The audit follows TRAC Criteria structure. Each response includes the criteria, the response, and the example evidence provided by TRAC. The responsibilities of Digital Safe as a service, Arkivum as a storage platform, and Archivematica as a digital preservation workflow are all included. If they do not have a responsibility for the criteria it is acknowledged as such. TRAC can be repetitive, and the audit responses can be repetitive because the same answer may apply to several questions or are not yet fully developed.

Audit Recommendations

A brief review of the recommendations for Digital Safe as a result of this informal audit are as follows:

Previously Identified Goals

- Obtain funding (in progress)
- Develop and cost and service model to be integrated into IT Services
- Define contract specifications with Arkivum
- Deploy service to early adopters

Audit-Identified Goals

- Secure University web space to host information about the service, including basic policy and purpose description, help guides and contacts, training material, and an access matrix (as outlined

in section C3.3) to provide transparency on use of the service as well as access to content. Also provide links to the Bodleian Library and IT Services

- Continue updating the Wordpress blog with a link to the University web space
- Determine a permanent Steering Committee or governance committee for the next phase of the project. If this is already in place unofficially, develop a formal policy on term length, responsibilities, and a process for member replacement.
- Develop a written policy that defines the best practices for utilizing Archivematica and Arkivum, specifically in recommended digital preservation workflows and in what Arkivum product best suits the client's needs.
- Continue building on this, or a similarly constructed, informal audit. The purpose of this audit is meant to review and evaluate what level of completion and trustworthiness Digital Safe has obtained. Developing multiple versions that build on this pre-Phase 3 audit might be beneficial for tracking progress and highlighting existing or new issues.

Digital Safe Informal TRAC Audit

A. Organizational Infrastructure

Average Rating: 1.45/4

A1. Governance & organizational viability

A1.1 Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.

Audit Rating: 2

DS: The original Electronic Archive Pilot Project's mission statement is as follows: The Electronic Archive Pilot Project will establish the feasibility of a working electronic archive for the use of the whole of the Collegiate University. The archive will support the safe and secure storage of all classifications of non-public record data that individual departments, colleges and associated units are required to keep legally or would like to keep for historic reasons. The pilot project aims to develop a cost recovered service" (EAPP).

The Digital Safe service itself as yet lacks a mission statement. It will likely draw on the phrase from the Phase 3 PID to "deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis once in production," (2).

Arkivum: "Arkivum provides industry-leading big data preservation and archiving solutions to organisations in higher education, healthcare, life sciences, and digital heritage. These solutions assure the long-term value, trustworthiness and authenticity of data irrespective of whether it's terabytes or petabytes being preserved, and irrespective of whether the retention period is years, decades, or a quarter of a century. Through active data management, chain of custody and ISO 27001 compliance processes, Arkivum's unique technology provides rapid, low-latency access to archived data and provides an unrivalled 100% data integrity guarantee. Backed by indemnity insurance, this is our commitment to protect, curate and preserve data for the future and to eliminate the needless loss of information and knowledge. Arkivum works with partners to deliver integrated, scalable and flexible solutions for data discovery and sharing; publishing; file format preservation; and information portals" (Arkivum, About Us).

Archivematica: "Archivematica is a free and open-source digital preservation system that is designed to maintain standards-based, long-term access to collections of digital objects. Archivematica is packaged with the web-based content management system AtoM for access to your digital objects," (What is Archivematica).

A1.2 Repository has an appropriate, formal succession plan, contingency plans, and/or a escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.

Audit Rating: 4

DS: Digital Safe chose Arkivum in part due to its contingency plan and relies on Arkivum to maintain and implement this plan if necessary.

Arkivum: Arkivum provides several safety measures. For Arkivum/1+1, one copy is stored in a secure data center and one copy is saved on LTO data tape in Escrow. For Arkivum/100, two copies are stored in secure, geographically separate data centers and one copy is saved on LTO data tape in Escrow. The Escrow copy is the backup for any data loss or corruption. If there is any data loss in the data centers or if Arkivum Limited should cease to operate, the LTO tapes are delivered to the client. If there is complete data loss, Arkivum provides a "financial guarantee underwritten by an Information and Communication Technology Professional Liability Insurance Policy," which provides coverage for direct loss relating to data loss (FAQ, 7). Arkivum also provides multiple client sites, so if one site is compromised the data may be retrieved at another site (FAQ, 22). For more information:

- section A3.8 for more information on Arkivum/1+1 and Arkivum/100
- section C3.2 for information on Arkivum's physical storage locations
- Stages of Archiving
- Chain of Custody
- Security Model
- Arkivum's FAQ document

Archivematica: Archivematica is a digital preservation workflow tool. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. Archivematica is integrated with Arkivum via Arkivum's A-Stor, which will be built into the contract between the University and Arkivum, and is therefore protected by the security measures of Arkivum.

Evidence: Succession plan(s); escrow plan(s); explicit and specific statement documenting the intent to ensure continuity of the repository, and the steps taken and to be taken to ensure continuity; formal documents describing exit strategies and contingency plans; depositor agreements.

A2. Organizational structure & staffing

A2.1 Repository has identified and established the duties that it needs to perform and has appointed staff with adequate skills and experience to fulfill these duties.

Audit Rating: 1

DS: The project team for Phase 3 of Digital Safe is established, but the ultimate short and long-term staffing duties have not yet been established. The plan is that the service will be built into the University of Oxford's IT department. A business and service model is a priority for phase 3 of Digital Safe and will determine staffing needs and training. Ideally the Steering Committee on the Digital Safe project will also develop into a governance committee that will act as a checks and balances to ensure that local management is useful, that will review policy annually, and to ensure funding.

Digital Safe relies on the technologies to maintain their own staffing.

Arkivum: Digital Safe relies on Arkivum to ensure their own staffing needs and is not responsible for Arkivum maintaining their services. According to Arkivum, "In addition to technical change in the archive system, managing staff transitions of those who run the system, for example support staff and administrators, is required" (Data Integrity). More information on their team may be found on their About Us page.

Archivematica: Archivematica is created and staffed by Artefactual Inc. More information can be found on their Team page.

Evidence: A staffing plan; competency definitions; job description; development plans; plus evidence that the repository review and maintains these documents as requirements evolve.

A2.2 Repository has the appropriate number of staff to support all functions and services.

Audit Rating: 1

See the response in section A2.1 for details.

Evidence: Organizational charts; definitions of roles and responsibilities; comparison of staffing levels to commitments and estimates of required effort.

A2.3 Repository has an active professional development program in place that provides staff with skills and expertise development opportunities.

Audit Rating: 2

DS: Arkivum provides all virtual training. The current plan is to bring in Arkivum employees for training and a training course will be developed for staff here based on Arkivum's resources (NJ Interview). There will also be basic tutorials for Archivematica on the tools available, though this training will be a broader, Oxford-level and the individual clients will determine their own specific workflow, especially in regards to their use of Archivematica. The training will then be added to the Bodleian Library's current collection of training materials. The training materials will be reviewed and updated on a 4-5 year cycle by the governance committee.

Arkivum: Arkivum provides all virtual training with courses being built into the start-up fees. Additional help and support is also available via email and phone. Contact information is as follows: "For initial support please contact your reseller where appropriate. Should this not be possible or you need to speak to Arkivum, then please call our support staff on +44 1249 400 001 or e-mail support@arkivum.com. Support services are provided weekdays, during the hours of 8:30 a.m. to 5:30 pm UK time. Outside of these times automated alerting systems are in operation, with escalation to a designated analyst" (FAQ, 12). Additional information is found on their website.

Archivematica: Archivematica's manufacturer Artefactual provides training for Archivematica use online, onsite in workshops, and via VMs for classroom training (Training). These workshops are priced separately from the service, but there may be a discount included with the contract between the University of Oxford and Arkivum. Technical support is also an option for Archivematica clients (Maintenance Services), which is also priced separately and could be included in the contract.

Support services may be accessed by phone at +1 604 527 2046 or via email at info@artefactual.com (Contact page).

Evidence: Professional development plans and reports; training requirements and training budgets, documentation of training expenditures (amount per staff); performance goals and documentation of staff assignments and achievements, copies of certificates awarded.

A3. Procedural accountability & policy framework (documentation)

A3.1 Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.

Audit Rating: 4

DS: The key users for this service have been identified in Phase 1 as: College Archivists; University Archive; Central Administrative Records Management; Departmental Research Records Management; and Bodleian Electronic Archives and Manuscripts. These users have materials that require high-security and low-accessibility, including administrative records, student records, financial records, personal communication, medical reports, non-anonymized case studies, and other material that has personal, identifiable information. These users were identified after interviewing various colleges and departments on campus and determining a need for a universal storage system (PID, 4). Users are not limited to only these categories, however, as the service is open to all who want to use the service and are affiliated with the University of Oxford.

Arkivum: In brief, Arkivum was developed for the long term, large-scale management, protection, and curation of data primarily from institutions based in Healthcare and NHS, Digital Heritage, Higher Education, and Life Sciences. This does not limit their scope, and more information can be found in their Solutions examples (see the Higher Education example), and in their FAQ document.

Archivematica: Archivematica was developed to provide "archivists and librarians with limited technical and financial capacity the tools, methodology and confidence to begin preserving digital information today," (Archivematica) which expands into the entire digital preservation community overall.

Evidence: Mission statement; written definitions of the designated community(ies); documented policies; service-level agreements.

A3.2 Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolves.

Audit Rating: 2

DS: The official governance for this service has not yet been developed. Ideally the Steering Committee on the Digital Safe project will also develop into a governance committee that will work in tandem with the staff in the IT department to review and update any general policies, training materials, and announcements and event information. Also see the response for section A2.1 for staffing information.

Digital Safe relies on the technologies to remain updated on and implement any evolving best practices in the field.

Arkivum: Arkivum has policies and automated processes in place for running integrity checks on the data, software, and hardware. The data is retrieved annually and given an integrity test based on checksums. Arkivum has also identified software and hardware obsolescence to occur on a cycle of generally 3 to 5 years, so Arkivum's policy is that data is migrated to new media following the LTO roadmap. The LTO data tapes in Escrow are also migrated every 5 years. More detailed information on media upgrades can be found on their Data Integrity page.

Archivematica: In addition to being committed to maintaining standards-based tools for those interested in digital preservation tools, Archivematica also relies on their community to help steer the tool in the most useful direction. “We're constantly working with our community to improve the application, and all enhancements are bundled into our public releases. This means that whenever one person or institution contributes resources, the entire community benefits,” (Archivematica).

Evidence: Written documentation in the form of policies, procedures, protocols, rules, manuals, handbooks, and workflows; specification of review cycle for documentation; documentation detailing review, update, and development mechanisms. If documentation is embedded in system logic, functionality should demonstrate the implementation of policies and procedures.

A3.3 Repository maintains written policies that specify the nature of any legal permissions required to preserve digital content over time, and repository can demonstrate that these permissions have been acquired when needed.

Audit Rating: 2

DS: Legal permissions will be the entire responsibility of the client. The content will likely comprise of files created by the client, e.g. student records, financial records, etc., and legal permissions are moot. Other material may be acquisitions to the Bodleian Library that have their own documentation and standards that are separate from this service. Each client may also have their own policies on permissions and permissions workflow that are independent of each other. See section A5 for more information on policy and permissions development.

Arkivum: In regards to service agreements between Arkivum and clients, there will be a contract between individual clients and Arkivum that articulate the clients’ product choice, workflow preferences, and additional storage space options (NJ Interview). Furthermore, according to Arkivum “Our operations at all sites, including our business offices, is certified to ISO 27001 information security standards,” (FAQ, 19) and Arkivum is regularly audited externally to maintain ISO 27001 certification and welcomes client audits as well (FAQ, 10). These certifications enable Arkivum to legally store ingested content.

Archivematica: Archivematica is a digital preservation workflow tool, not a storage space. Both the workflow and the content are not accessible to users who are not specifically the client developing and using the workflow.

Evidence: Deposit agreements; records schedule; digital preservation policies; records legislation and policies; service agreements.

A3.4 Repository is committed to formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements.

Audit Rating: 2

DS: The official governance for this service has not yet been developed. Ideally the Steering Committee on the Digital Safe project will also develop into a governance committee that will review the contracts and funding options annually (NJ Interview). Information on this governance, contact information, and any updates or announcements will be provided on University web space in the IT Services space that will also be linked to the Bodleian homepage. The number of individuals in the University of Oxford who will be using the service will likely lead to a self-supporting committee that will further discuss use and policies (ST Interview).

Arkivum: Arkivum policy is under regular review as they aim to maintain data integrity by regular integrity checks and data migrations. See section A3.2 for more information, as well as Arkivum's long-term Data Integrity page. Arkivum is also ISO 27001 certified and "conforms to the controls within ISO 27002 to maintain its certifications against ISO 27001" (Security Model).

Archivematica: Archivematica seeks input from their user community. "We're constantly working with our community to improve the application, and all enhancements are bundled into our public releases. This means that whenever one person or institution contributes resources, the entire community benefits," (Archivematica). Archivematica is also committed to updated format policies as standards evolve (FPR section), which they maintain my monitoring and reacting to community discussions. See section B2.7 for more information on the FPR.

See the technologies' mission statements in section A1.1.

Evidence: A self-assessment schedule, timetables for review and certification; results of self-assessment; evidence of implementation of review outcomes.

A3.5 Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.

Audit Rating: 1

DS: This section is not fully developed because web space has not yet been devoted to Digital Safe. The service will be built into the University of Oxford's IT department, and a business and service model is a priority for Phase 3 of Digital Safe. Web space will be devoted to Digital Safe in the IT Services space that will also be linked to the Bodleian homepage. Ideally the webpage would have contact information for the governance committee or other local managing team who can assist in troubleshooting smaller issues and directing to training and Arkivum and Archivematica help pages (NJ Interview). This audit recommends that contact information and a Help and Feedback section are included in the University web space for the long-term.

Arkivum: According to Arkivum policy, "The security and audit model above has been developed in partnership with Arkivum customers who have confirmed that the model meets their regulatory requirements as part of a due-diligence/audit process that they have conducted on Arkivum. This includes due-diligence by customers in clinical and financial sectors where regulation is strict" (Chain of Custody). There is also contact information at the bottom of every page on Arkivum.

Archivematica: Archivematica was originally a project use case for OAIS to "process analysis to synthesize the specific, concrete steps that must be carried out to comply with the OAIS functional model from Ingest to Access." This project expanded beyond OAIS into its current state as an open-source digital preservation workflow tool based on user feedback (Intro page). Clients do have to navigate to the manufacturer page in order to contact Artefactual for assistance. See A3.4 for more information on their communication with their user community.

Evidence: A policy that requires a feedback mechanism; a procedure that addresses how the repository seeks, captures, and documents responses to feedback; documentation of workflow for feedback (i.e., how feedback is used and managed); quality assurance records.

A3.6 Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content.

Audit Rating: 2

DS: Digital Safe will update any announcements and training material for the service, and the local governance committee will not have any documentation that will need regularly updated due to the sensitive information and privacy of the dark archives Digital Safe is providing. Digital Safe relies on Arkivum and Archivematica to handle obsolescence, migration, data integrity, and generating any new training materials.

Arkivum: Arkivum regularly updates its materials, see section A3.2 and Arkivum's long-term Data Integrity page.

Archivematica: Archivematica has detailed Documentation of their tool on their website. See section A3.5 for information on their community.

Evidence: Policies, procedures, and results of changes that affect all levels of the repository: objects, aggregations of objects; object-level preservation metadata; repository's records retention strategy document.

A3.7 Repository commits to transparency and accountability in all actions supporting the operation and management of the repository, especially those that affect the preservation of digital content over time.

Audit Rating: 2

DS: Digital Safe is a service dedicated to allowing University of Oxford institutions the ability to store their high-security records and materials in a dark archive that utilizes the same technology University-wide. Because the purpose is to securely store material and not to offer easy accessibility, transparent access is not applicable to this service. Though it does not yet exist, this audit recommends that University web space for Digital Safe is created to briefly describe the key users identified by phase 1 and briefly explain why accessibility is limited to the clients of the University. Other users outside of the University can utilize the platform and technologies, but without any benefits from accessing them via the University.

Arkivum: Arkivum provides detailed Documentation on their website about their technology and processes on the Technical Overview page. Digital Safe relies on Arkivum to maintain their transparency and accountability and are not responsible if Arkivum does not. Access to materials will be strictly monitored. Individuals with the encryption keys, likely the Archivist or a similar position, will determine user access. These users will have Active Directory permissions that can be integrated into individual segments of the archive. All of the access is user-controlled, and these users will not be publishing their policies to anyone but their own staff. More information can be found in the FAQ on page 10.

Archivematica: Archivematica has detailed Documentation of their tool on their website. Archivematica is a digital preservation workflow tool, not a storage space. Both the workflow and the content are not accessible to users who are not specifically the client developing and using the workflow.

Evidence: Comprehensive documentation that is readily accessible to stakeholders; unhindered access to content and associated information within repository.

A3.8 Repository commits to defining, collecting, tracking, and providing, on demand, its information integrity measurements.

Audit Rating: 4

DS: Digital Safe is a service and bears no responsibility for materials being collected. Digital Safe also relies on Arkivum for all tracking, data integrity, and storage needs.

Arkivum: In brief, Arkivum provides a workflow and safety measures for integrity. For Arkivum/1+1, one copy is stored in a secure data center and one copy is saved on LTO data tape in Escrow. For Arkivum/100, two copies are stored in secure, geographically separate data centers and one copy is saved on LTO data tape in Escrow. The files will retain their original file names. Checksums are used to ensure data is correct and complete after migration and during storage. In the case of network errors, the workflow ensures that the event (e.g. a transfer) and its progress is tracked, and any result that is not deemed successful is automatically repeated or queued until the network problem is solved. Safety checks are provided to clients to ensure nothing is deleted until the ingest process is completed. More information on Arkivum's workflow and policies can be found on their Maintaining Data Integrity page.

Arkivum also followed the LTO roadmap for storage obsolescence and aims to prevent data loss by determining what LTO generation their system is and introducing new generations well in advance to any system failure. Escrow copies are also migrated every five years to prevent data loss. More information on their workflow and policies can be found on Arkivum's long-term Data Integrity.

Archivematica: Archivematica is a digital preservation workflow tool. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum, and the content is only accessible by the client importing it. Archivematica is integrated with Arkivum via Arkivum's A-Stor, which will be built into the contract between the University and Arkivum, and is therefore protected by the security measures of Arkivum.

Evidence: An implemented registry system; a definition of the repository's integrity measurements; documentation of the procedures and mechanisms for integrity measurements; an audit system for collecting, tracking, and presenting integrity measurements; procedures for responding to results of integrity measurements that indicate digital content is at risk; policy and workflow documentation.

A3.9 Repository commits to a regular schedule of self-assessment and certification and, if certified, commits to notifying certifying bodies of operational changes that will change or nullify its certification status.

Audit Rating: 2

DS: Digital Safe does not require certification as it is a service provided by the University via an outsourced service. Digital Safe relies on Arkivum to maintain its ISO 27001 certification and regular audit checks of data, methods, technology, and physical locations certification and recognizes a breach in contract if Arkivum does not maintain certification.

Digital Safe will update the University web space with any new general policies, training material, and news and announcements.

Arkivum: Arkivum is ISO 27001 certified and "conforms to the controls within ISO 27002 to maintain its certifications against ISO 27001" (Security Model). Additionally, "The data centres used by Arkivum are Tier 2 or Tier 3 and are ISO 27001 certified or have FACT accreditation. They are inspected by Arkivum on a regular basis and have also been inspected by our ISO 27001 auditor" (Security Model). Other audited sections include the production system access, building access, and logical access to data. For more information, see Arkivum's Security Model page. The Escrow copy is the backup for any data loss or corruption. If there is any data loss in the data centers or if Arkivum Limited should cease to operate, the LTO tapes are delivered to the client.

Should Archivematica cease to exist, the contract between the University and Arkivum will need to be reviewed.

Archivematica: Archivematica is a digital preservation workflow tool and not a storage service, and in the event of Archivematica ceasing to operate, there would be no chance of data loss. If Archivematica requires an update, more information can be found in their Installing from packages section.

Evidence: Completed, dated audit checklists from self-assessment or objective audit; certificates awarded for certification; presence in a certification register (when available); timetable or budget allocation for future certification.

A4. Financial sustainability

A4.1 Repository has short- and long-term business planning processes in place to sustain the repository over time.

Audit rating: 1

DS: This section has not yet been developed. A cost model is a priority for phase 3 (PID, 2) and will determine the start-up costs only. The success of phase 3 will determine and long-term cost model based on the contract between the University of Oxford and Arkivum, and the support from the University for the service.

The projected expense report for Phase 3 includes the following factors (Phase 3 PID, 12) but does omit any information on the actual budget for the project for the sake of confidentiality:

Expense type	
1. IT Services internal staff	
2. Non-IT Services staff	
3. Hardware, software, training and equipment / storage (Archivum)	
4. Advertising, consumables and room bookings	
Total project cost	
0.5% charge for PMO	
Contingency at 8.5% (according to Monte Carlo Simulation)	
Total project costs, including PMO and contingency	
Forecast on-going charges, per year	

Table 1 Projected Expense report outlining expense types for Phase 3

As based on the NSMS example, this service will be based on the cost recovery model. According to the Phase 3 PID, "The cost recovery solution will encompass the recovery of; all third-party charges, the local FTE resource needed to manage and maintain this service, and any local costs towards the infrastructure and power needed to run this service" (12).

Arkivum: Arkivum does not publish their financial reports, business plans, or blank contracts. However, Arkivum has several well-known institutions as clients, including the Museum of Modern Art, University of Westminster, and the Oxford Molecular Diagnostics Centre, among many others noted in case studies on their website that have been successful. Their Solutions tab offers several reports and case studies in various fields, such as Higher Education, that are evidence of success.

Archivematica: Archivematica does not publish their financial reports or business plans. Archivematica is an open-source tool and therefore does not charge clients. However, they provide paid services, including storage, training, technical support, all noted on their Services page. In the long-term, they have had several successful clients list on their Clients page.

Evidence: Operating plans; financial reports; budgets; financial audit reports; annual financial reports; financial forecasts; business plans; audit procedures and calendars; evidence of comparable institutions; exposure of business plan to scenarios.

A4.2 Repository has in place processes to review and adjust business plans at least annually.

Audit Rating: 1

DS: The official governance for this service has not yet been developed. Ideally the Steering Committee on the Digital Safe project will also develop into a governance committee that will review the contracts and funding options annually See section A3.4 for more information.

Arkivum: Arkivum policy is under regular review as they aim to maintain data integrity by regular integrity checks and data migrations. See section A3.2 and A3.4 for more information, as well as Arkivum's Data Integrity page.

Archivematica: Archivematica is continuously evolving as standards develop and evolve, and relies on their community to help steer the tool in the most useful direction. "We're constantly working with our community to improve the application, and all enhancements are bundled into our public releases. This means that whenever one person or institution contributes resources, the entire community benefits," (Archivematica).

Evidence: Business plans, audit planning (e.g., scope, schedule, process, and requirements) and results; financial forecasts; recent audits and evidence of impact on repository operating procedures.

A4.3 Repository's financial practices and procedures are transparent, compliant with relevant accounting standards and practices, and audited by third parties in accordance with territorial legal requirements.

Audit Rating: 1

DS: Funding has not yet been secured. Start-up funding is a priority for Phase 3 (PID, 2). Ideally the Steering Committee on the Digital Safe project will also develop into a governance committee that will review the contracts and funding options annually. See section A3.4 and A4.1 for more information.

Arkivum: Arkivum does not publish their financial reports, business plans, or blank contracts. The relationship between the client and Arkivum will be dictated by a contract. Given the number of clients the University would bring to Arkivum there may be a discount for Arkivum and Archivematica service. The benefits to this service will be outlined in the policies for University clients.

Archivematica: Archivematica does not publish their financial reports or business plans. Archivematica is an open-source tool and therefore does not charge clients. However, they provide paid services, including storage, training, technical support, all noted on their Services page.

Evidence: Demonstrated dissemination requirements for business planning and practices; citations to and/or examples of accounting and audit requirements, standards, and practice; evidence of financial audits already taking place.

A4.4 Repository has ongoing commitment to analyze and report on risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).

Audit Rating: 1

DS: Funding has not yet been secured for the short or long-term. Start-up funding is a priority for Phase 3 (PID, 2). Ideally the Steering Committee on the Digital Safe project will also develop into a governance committee that will review the contracts and funding options annually. See section A3.4 and A4.1 for more information.

Arkivum: As they are a private business, Arkivum does not publish their financial reports, business plans, or blank contracts.

Archivematica: Archivematica does not publish their financial reports or business plans.

Evidence: Risk management documents that identify perceived and potential threats and planned or implemented responses (a risk register); technology infrastructure investment planning documents; cost/benefit analyses; financial investment documents and portfolios; requirements for and examples of licenses, contracts, and asset management; evidence of revision based on risk.

A4.5 Repository commits to monitoring for and bridging gaps in funding.

Audit Rating: 0

DS: Funding has not yet been secured for the short or long-term. Start-up funding is a priority for Phase 3 (PID, 2). Ideally the Steering Committee on the Digital Safe project will also develop into a governance committee that will review the contracts and funding options annually. See section A3.4 for more information.

Arkivum: Arkivum relies on Digital Safe service to monetarily commit to the agreed upon contract.

Archivematica: Archivum relies on Digital Safe service to monetarily commit to the agreed upon contract.

Evidence: Fiscal and fiduciary policies, procedures, protocols, requirements; budgets and financial analysis documents; fiscal calendars; business plan(s); any evidence of active monitoring and preparedness.

A5. Contracts, licenses, & liabilities

Audit Rating: 0

This entire section cannot be audited as no contracts or liabilities exist yet. A priority of Phase 3 of Digital Safe will be in finalizing the contract with Arkivum and in determining any preservation rights and copyrights. Services that are part of the library may come under the Heritage Institution exception for the right to change objects and make copies, which would occur in the regular migration of data in Arkivum and in any Archivematica workflow the data is pushed through. Though much of the material may be produced by the clients at the University (e.g. financial records, etc.), some of BEAM's content may apply to the exception.

A5.1 If repository manages, preserves, and/or provides access to digital materials on behalf of another organization, it has and maintains appropriate contracts or deposit agreements.

Evidence: Deposit agreements; policies on third-party deposit arrangements; contracts; definitions of service levels; Web archiving policies; procedure for reviewing and maintaining agreements, contracts, and licenses.

A5.2 Repository contracts or deposit agreements must specify and transfer all necessary preservation rights, and those rights transferred must be documented.

Evidence: Contracts, deposit agreements; specification(s) of rights transferred for different types of digital content (if applicable); policy statement on requisite preservation rights.

A5.3 Repository has specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.

Evidence: Submission agreements/deposit agreements/deeds of gift; written standard operating procedures.

A5.4 Repository tracks and manages intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.

Evidence: A policy statement that defines and specifies the repository's requirements and process for managing intellectual property rights; depositor agreements; samples of agreements and other documents that specify and address intellectual property rights; demonstrable way to monitor intellectual property; results from monitoring.

A5.5 If repository ingests digital content with unclear ownership/rights, policies are in place to address liability and challenges to those rights.

Evidence: A definition of rights; citations for relevant laws and requirements; policy on responding to challenges; documented track record for responding to challenges in ways that do not inhibit preservation; examples of legal advice sought and received.

B. Digital Object Management

Average Rating: 3.5/4

B1. Ingest: acquisition of content

B1.1 Repository identifies properties it will preserve for digital objects.

Audit Rating: 3

DS: The metadata fields that Digital Safe recommends the clients' preserve are: Title, Description, Creator(s), ID Type, ID Value, Retention review date, Retention rationale, Resource type, Technical description(s), Covering date(s), Finding aid(s), Rights & Licensing information (ICTF powerpoint). It is ultimately the client's decision what properties are preserved.

Arkivum: Arkivum only requires a file for anything to be ingested. Arkivum has no responsibility for anything additional included in the ingest and relies on the client to upload any additional information. Arkivum can ingest any file format and will maintain a copy of the original file alongside a normalized file.

Archivematica: Archivematica is an open-source workflow tool that can be integrated with Arkivum via Arkivum's A-Stor, and which will be built into the contract between the University and Arkivum. To see more about the integration of Arkivum with Archivematica, see their Storage Services page.

Archivematica allows users to do various archival activities, including adding metadata in Dublin Core, adding rights in PREMIS, data normalization, AIP storage, DIP storage, communication with other tools (e.g. Archivist's Toolkit, ArchiveSpace, Arkivum), among many other options that can be explored in Archivematica's Documentation.

A SIP begins as a transfer. "In Archivematica, Transfer is the process of transforming any set of digital objects and/or directories into a SIP. Transformation may include appraisal, arrangement, description and identification of donor restricted, private or confidential contents. The Transfer tab prepares your content for preservation in Archivematica" (Transfer). A transfer can be created with submission documentation, existing checksums, or an existing METS structmap. The transfer will be processed through several micro-services, as described in the Transfer process. This is then ingested into Archivematica after the green light is given to the client.

The client will be able to develop their own workflow and to define their own preserved properties based on their individual policies. Digital Safe can recommend best practices, though ultimately the decision of what and how to archive information will be determined by the client.

Evidence: Mission statement; submission agreements/deposit agreements/deeds of gift; workflow and policy documents, including written definition of properties as agreed in the deposit agreement/deed of gift; written processing procedures; documentation of properties to be preserved.

B1.2 Repository clearly specifies the information that needs to be associated with digital material at the time of its deposit (i.e., SIP).

Audit Rating: 3

DS: Digital Safe can recommend best practices but cannot dictate a client's policies. Recommendation for clients will be available on the Digital Safe website once it is developed. See section B1.1 for more information.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). Arkivum has no responsibility for anything additional included in the ingest and relies on the client to upload any additional information. If the client chooses to use the integrated Archivematica tool they may also employ digital preservation practices. See section B1.1 for more information.

Archivematica: Archivematica enables the client to create or submit a transfer that will then be made a SIP. More details can be found in their documentation on the Create a SIP page and Transfer page. This includes arranging SIPs, adding metadata, adding PREMIS rights, normalizing, and transcribing SIPs using the Tesseract OCR tool (Ingest). See section B1.1 for more information.

Evidence: Transfer requirements; producer-archive agreements.

B1.3 Repository has mechanisms to authenticate the source of all materials.

Audit Rating: 4

DS: Authentication will be the entire responsibility of the client. The content will likely comprise of files created by the client, e.g. student records, financial records, etc., and legal permissions are moot. Other material may be acquisitions to the Bodleian Library that have their own documentation and standards that are separate from this service. Each client may also have their own policies on permissions and permissions workflows that are independent of each other.

Arkivum: In regards to service agreements between Arkivum and clients, there will be a contract between individual clients and Arkivum that articulate the clients' product choice, workflow preferences, and additional storage space options. Furthermore, according to Arkivum "Our operations at all sites, including our business offices, is certified to ISO 27001 information security standards," (FAQ, 19) and Arkivum is regularly audited externally to maintain ISO 27001 certification and welcomes client audits as well (FAQ, 10). These certifications enable Arkivum to legally hold ingested content.

Archivematica: Archivematica is a digital preservation workflow tool and not a storage space, and has no responsibility regarding the content authentication.

Evidence: Submission agreements/deposit agreements/deeds of gift; workflow documents; evidence of appropriate technological measures; logs from procedures and authentications.

B1.4 Repository's ingest process verifies each submitted object (i.e., SIP) for completeness and correctness as specified in B1.2.

Audit Rating: 4

DS: Digital Safe relies on the client to submit and on Arkivum to retain and retrieve complete and correct files.

Arkivum: Checksums are used to verify that each file is correct and complete. For Arkivum/100 there is also the added security of the checksums generated by Arkivum being compared to the client's checksums, though this is not included in the Arkivum/1+1 product. The multiple copies of data and active data verification via annual data integrity checks complete the workflow to ensure that the data is correct. See Arkivum's archiving process for more information (Stage 2, Archiving Process).

Clients are also notified at what stage their data is at during ingestion using a "traffic light system" where Red indicates that the client copy must not be deleted; Amber indicates that the ingested files are at the Arkivum data centers; and Green indicates that the ingested files are replicated and protected in the prescribed data center(s) and in escrow and that the client file can be deleted. See the FAQ for more information (FAQ, 17).

Archivematica: The completion of all processes in Archivematica are indicated by color and text. Green indicates that a process has been completed successfully, and red indicates that the process has not been completed successfully. A client can search for content when by its name. Archivematica's naming system will retain the original name of the transfer unless a new name has been assigned to the SIP upon creation. This name will be combined with a Universal Unique Identifier that is generated and assigned during SIP formation (AIP Structure).

Evidence: Appropriate policy documents and system log files from system performing ingest procedure; formal or informal "acquisitions register" of files received during the transfer and ingest process; workflow, documentation of standard operating procedures, detailed procedures; definition of completeness and correctness, probably incorporated in policy documents.

B1.5 Repository obtains sufficient physical control over the digital objects to preserve them.

Audit Rating: 4

DS: Digital Safe is a service and therefore has no responsibility for the control of the content uploaded to Arkivum. Digital Safe relies on Arkivum to maintain and implement their storage policies. The service can recommend best practices and assist in troubleshooting.

Arkivum: Arkivum maintains high security both digital and physically. According to their FAQ documentation. "All copies of customer data are held in secure UK storage locations. Storage facilities are manned at all hours and access is strictly restricted to a list of named, trained and vetted members of the Arkivum Operations team. Our operations at all sites, including our business offices, is certified to ISO 27001 information security standards…escrow. In addition to ISO27001 certification and industry best practice for security, our customer base includes people using our service to store personal data including voice call recordings and medical treatment records. They have audited our service and satisfied themselves that our service is secure and meets their regulatory and legislative obligations" (FAQ, 19). Data is also encrypted once it leaves the client's network and passes through a secure VPN before entering a data center. The Escrow copy is located based on the contract between the client and Arkivum (FAQ, 5). For more information, see the FAQ documentation.

The client also has access to the data during the ingest process. According to their documentation, the client can "go in and get the files back with the same name and path that they used when they originally provided the data" (Technical Overview).

Archivematica: Archivematica is a digital preservation workflow tool and not a storage service. Archivematica is integrated with Arkivum via Arkivum's A-Stor, which will be built into the contract between the University and Arkivum, and is therefore protected by the security measures of Arkivum. Once the data is through the workflow it is transferred to Arkivum storage space and is fully under their security.

Evidence: Submission agreements/deposit agreements/deeds of gift; workflow documents; system log files from the system performing ingest procedures; logs of files captured during Web harvesting.

B1.6 Repository provides producer/depositor with appropriate responses at predefined points during the ingest processes.

Audit Rating: 4

DS: Digital Safe relies on Arkivum to notify the clients of progress and ingest impletion.

Arkivum: Clients are notified at what stage their data is at during ingestion using a "traffic light system" where Red indicates that the client copy must not be deleted; Amber indicates that the ingested files are at the Arkivum data centers; and Green indicates that the ingested files are replicated and protected in the prescribed data center(s) and in escrow and that the client file can be deleted. See the FAQ for more information (FAQ, 17).

Archivematica: The completion of all processes in Archivematica are indicated by color and text. Green indicates that a process has been completed successfully, and red indicates that the process has not been completed successfully (AIP Structure).

Evidence: Submission agreements/deposit agreements/deeds of gift; workflow documentation; standard operating procedures; evidence of "reporting back."

B1.7 Repository can demonstrate when preservation responsibility is formally accepted for the contents of the submitted data objects (i.e., SIPs).

Audit Rating: 3

DS: Digital Safe relies on Arkivum to notify the clients of progress and ingest completion. As the service is not yet active, there are no examples to test this process.

Arkivum: Arkivum provides a "Green light" upon a successful ingest completion that indicates the ingested files are replicated and protected in the prescribed data center(s) and in escrow and that the client file can be deleted. See the FAQ for more information (FAQ, 17) or the archiving process page (Stage 6, Archiving Process).

Archivematica: See the response for section B1.6 for information on process completion.

Evidence: Submission agreements/deposit agreements/deeds of gift; confirmation receipt sent back to producer.

B1.8 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Ingest: content acquisition).

Audit Rating: 4

DS: Digital Safe does not bear any responsibility for logging actions of content. Digital Safe does the use of audit trails recommend to clients, though the use of audit trails is determined by the client.

Arkivum: Arkivum offers audit trails as a service. These must be turned on and specified by the client upon contract agreement. Audit trails are "accessible in machine-readable XML format through Arkivum REST API calls, either on a per-file or per-folder basis. Audit trails are also accessible in human readable PDF/A format through the Arkivum Service web interface or through an API call. Audit trails in PDF/A format can be signed if necessary to show that they were generated by Arkivum Service." More detailed information concerning audit trails can be found at their Audit Trails page.

Archivematica: Archivematica utilizes fixity checks on files before AIP storage. See B2.11 for more information about Archivematica's fixity program. The client has complete control over the workflow process and can monitor at what stage the content is in as it moves through the workflow and once a client has established their digital preservation workflow, this will be recorded in their own policies. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

Evidence: Written documentation of decisions and/or action taken; preservation metadata logged, stored, and linked to pertinent digital objects.

B2. Ingest: creation of the archival package

B2.1 Repository has an identifiable, written definition for each AIP or class of information preserved by the repository.

Audit Rating: 4

DS: Digital Safe relies on the client and Archivematica to create AIPs, and on Archivematica to store and retrieve AIPs.

Arkivum: The AIP is determined by the client as they choose what archival processes are a part of their workflow. Arkivum is designed to hold AIPs. "The persistent file/folder mechanism within the Arkivum service offers excellent support for the storage of AIPs. The file will retain its original filename, and checksums are provided for incorporation in the Preservation Description Information (PDI) for the AIP. The service follows the OAIS model for Archive Storage through use of replication, fixity monitoring and repair, disaster recovery, migration, and tiered storage to deliver a specified level of performance, availability and integrity of storage" (Integration with Other Systems).

Archivematica: If the client decides to utilize the Archivematica tool they may choose to create and package an AIP. The client must determine the processing configuration, which can be left at Archivematica's default setting, or can be created by the client. See Archivematica's Processing Configuration documentation for more details.

After configuring the process as desired, the SIP can be normalized and stored in an AIP. "After normalization is approved, the SIP runs through a number of micro-services, including processing of the submission documentation, generation of the METS file, indexing, generation of the DIP and packaging of the AIP," which is packaged according to Bagit specifications (AIP Storage). Detailed information on the structure of the AIP can be found in Archivematica's AIP Structure documentation. The client may review the AIP and proceed to storing the AIP.

AIP reingest is also an option if the client wishes to add information (e.g. metadata and data normalization) after the SIP process, which can be found in their AIP Reingest documentation.

Evidence: Documentation identifying each class of AIP and describing how each is implemented within the repository. Implementations may, for example, involve some combination of files, databases, and/or documents.

B2.2 Repository has a definition of each AIP (or class) that is adequate to fit long-term preservation needs.

Audit Rating: 4

DS: Digital Safe relies on the client and Archivematica to create AIPs, and on Archivematica to store and retrieve AIPs.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). Arkivum has no responsibility for anything additional included in the ingest and relies on the client to upload any additional information. If the client chooses to use the integrated Archivematica tool they may

also employ digital preservation practices. See section B1.1 for more information on metadata, and section A1.2 for information on their storage process and security measures.

Archivematica: If the client chooses to use the Archivematica tool and creates a SIP, the data can also be normalized for AIP packaging. There are five options for normalization. For more detailed information, see:

- section B2.1
- section B2.3
- Archivematica's AIP Storage
- Archivematica's Normalization process
- Archivematica's Preservation Planning strategies

Evidence: Documentation that relates the AIP component's contents to the related preservation needs of the repository, with enough detail for the repository's providers and consumers to be confident that the significant properties of AIPs will be preserved.

B2.3 Repository has a description of how AIPs are constructed from SIPs.

Audit Rating: 4

DS: Digital Safe relies on the client and Archivematica to create AIPs, and on Archivematica to store and retrieve AIPs.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). Arkivum has no responsibility for anything additional included in the ingest and relies on the client to upload any additional information. If the client chooses to use the integrated Archivematica tool they may also employ digital preservation practices. See section B1.1 for more information.

Archivematica: If the client chooses to use the Archivematica tool and creates a transfer and a SIP, the data can also be normalized for AIP packaging. "After normalization is approved, the SIP runs through a number of micro-services, including processing of the submission documentation, generation of the METS file, indexing, generation of the DIP and packaging of the AIP," which is packaged according to Bagit specifications The AIP and any additional METS and PREMIS files can be downloaded during this stage if needed. For more detailed information, see section B2.1, B2.2, and Archivematica's AIP Storage documentation.

Evidence: Process description documents; documentation of SIP relationship to AIP; clear documentation of how AIPs are derived from SIPs; documentation of standard/process against which normalization occurs; documentation of normalization outcome and how outcome is different from SIP.

B2.4 Repository can demonstrate that all submitted objects (i.e., SIPs) are either accepted as whole or part of an eventual archival object (i.e., AIP), or otherwise disposed of in a recorded fashion.

Audit Rating: 4

DS: Digital Safe relies on the client and Archivematica to create AIPs, and on Archivematica to store and retrieve AIPs.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). Arkivum has no responsibility for anything additional included in the ingest and relies on the client to upload any additional information. If the client chooses to use the integrated Archivematica tool they may also employ digital preservation practices. See section B1.1 for more information.

Archivematica: Once the SIP has been created, reviewed, and saved, the normalization process occurs. The client can review and accept or reject the SIP during the normalization stage, as well as review and accept or reject the normalization. "After normalization is approved, the SIP runs through a number of micro-services, including processing of the submission documentation, generation of the METS file, indexing, generation of the DIP and packaging of the AIP," which is packaged according to Bagit specifications. For more detailed information, see section B1.2, B2.1, B2.2, and Archivematica's AIP Storage documentation.

Evidence: System processing files; disposal records; donor or depositor agreements/deeds of gift; provenance tracking system; system log files.

B2.5 Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects (i.e., AIPs).

Audit Rating: 4

DS: Digital Safe relies on the client and the technologies to create and organize identifiers for all archived objects.

Arkivum: The uploaded files will retain their original file names. Checksums are used to ensure data is correct and complete after migration and during storage (Stage 2, Archiving Process). Other naming systems may be generated by Archivematica.

Archivematica: Archivematica's naming system will retain the original name of the transfer unless a new name has been assigned to the SIP upon creation. This name will be combined with a Universal Unique Identifier that is generated and assigned during SIP formation. For more detailed information on this and the structure of the AIP, see the AIP Structure page.

Evidence: Documentation describing naming convention and physical evidence of its application (e.g., logs).

B2.6 If unique identifiers are associated with SIPs before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP).

Audit Rating: 4

DS: Digital Safe relies on the client and the technologies to create and organize identifiers for all archived objects.

Arkivum: The uploaded files will retain their original file names. Checksums are used to ensure data is correct and complete after migration and during storage (Stage 2, Archiving Process). Other naming systems may be generated by Archivematica.

Archivematica: Archivematica's naming system will automatically retain the original name of the transfer unless a new name has been assigned to the SIP upon creation. This name will be combined with a Universal Unique Identifier that is generated and assigned during SIP formation. For more detailed information on this and the structure of the AIP, see the AIP Structure page. Ultimately it is the decision of the client to retain the original name or generate a new one upon SIP creation.

Evidence: Workflow documents and evidence of traceability (e.g., SIP identifier embedded in AIP, mapping table of SIP IDs to AIPs).

B2.7 Repository demonstrates that it has access to necessary tools and resources to establish authoritative Representation Information of the digital objects it contains.

Audit Rating: 4

DS: Digital Safe relies on the client and the technologies to create and organize metadata for all archived objects.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). Additionally, "the main challenge in [Arkivum's] view is with discipline specific data formats, e.g. data collected from laboratory equipment, environmental sensor data, numerical simulations etc. Here bespoke or proprietary formats are often used. Whether these formats are at risk or not depends on whether the institution that has created the data, or will use data, has the necessary skills and tools to read the data and can maintain this capability. This can vary hugely between institutions even for the same data format," (FAQ, 23). If the client chooses to utilize Archivematica, Archivematica can determine file formats and normalize.

Archivematica: Archivematica contains a Format Policy Registry (FPR) that contains the default format policies and is maintained by Artefactual Systems, Inc (which as of July, 2016 does not have a public interface yet. See B2.8 for more information). This system also allows for clients to define their format policies in a local FPR that is accessible via the FPR server maintained by Artefactual. Archivematica is also committed to updated format policies as standards evolve; "A format policy indicates the actions, tools and settings to apply to a digital object of a particular format (e.g. conversion to preservation format, conversion to access format, extraction of package formats). Format policies will change over time as local and community standards, practices and tools evolve" (FPR section). For additional information on the FPR and configuring a local FPR, see the Preservation Planning page, and the full FPR page.

Evidence: "Evidence: Subscription or access to such registries; association of unique identifiers to registries of Representation Information (including format registries); Viewable records in local registries (with persistent links to digital objects); database records that include Representation Information and a persistent link to relevant digital objects.

B2.8 Repository records/registers Representation Information (including formats) ingested.

Audit Rating: 4

DS: Digital Safe relies on Arkivum to document the Representation Information for digital objects.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). Arkivum can ingest any file format, but does not record their representation information. If the client

chooses to utilize Archivematica, Archivematica can determine file formats and normalize using their Format Policy Registry, as well as creating SIPs and AIPs for record this process.

Archivematica: Archivematica contains a Format Policy Registry (FPR) that contains the default format policies and is maintained by Artefactual Systems, Inc. The latest version, FPR 1.4, does not have a public interface yet. There is a Public Roadmap wiki page outlining the development planning for a public interface. For clients who "expect to be writing/altering commands, implementing new tools, etc.," the FPR main page provides detailed directs on configuring and editing FPR policies.

Evidence: Viewable records in local format registry (with persistent links to digital objects); local metadata registry(ies); database records that include Representation Information and a persistent link to relevant digital objects.

B2.9 Repository has documented processes for acquiring preservation metadata (i.e., PDI) for its associated Content Information and acquires preservation metadata in accordance with the documented processes. The repository must maintain viewable documentation on how the repository acquires and manages Preservation Description Information (PDI).

Audit Rating: 3

DS: Digital Safe bears no responsibility for metadata acquisition or preservation and relies on the client to provide their own metadata, Archivematica to process the meta, and on Arkivum to store the metadata.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). If the client chooses to utilize Archivematica, Archivematica can ingest metadata and also has the option to generate metadata using Dublin Core standards.

Though Arkivum is not responsible for any digital preservation documentation, once the content has been imported into Arkivum, Arkivum does offer audit trails as a service. These must be turned on and specified by the client upon contract agreement. Audit trails are "accessible in machine-readable XML format through Arkivum REST API calls, either on a per-file or per-folder basis. Audit trails are also accessible in human readable PDF/A format through the Arkivum Service web interface or through an API call. Audit trails in PDF/A format can be signed if necessary to show that they were generated by Arkivum Service." More detailed information can be found on their Audit Trails page.

Archivematica: The client controls what metadata, if any, is ingested. Once a transfer has been created and processed into a SIP in Archivematica the client may also import their own metadata (Import Metadata), or they may create their metadata in in Archivematica using Dublin Core standards (Add Metadata). This can occur before or after the normalization process. The client can also add PREMIS rights (PREMIS Rights). After normalization and during AIP storage, the AIP may also be downloaded.

The client has complete control over the workflow process and can monitor at what stage the content is in as it moves through the workflow and once a client has established their digital preservation workflow, this will be recorded in their own policies. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

Evidence: Viewable documentation on how the repository acquires and manages Preservation Description Information (PDI).

B2.10 Repository has a documented process for testing understandability of the information content and bringing the information content up to the agreed level of understandability.

Audit Rating: 3

DS: The technologies chosen for Digital Safe were researched by the project team in Phase 2, which consisted of Knowledge Engineers, University Archivists, Technical Consultants, and Senior Users from the University. The project team for Phase 3 will be tracking the early adopters use of the service, which will determine what recommendations Digital Safe provides for the clients. Digital Safe will provide training for the service and act as a local manager of the service, including providing updated training materials and regular training sessions (PID, 19). Arkivum or Archivematica may need to be consulted for complex issues. See section A2.3 for more information.

Arkivum: In regards to imported content, Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). If the client chooses to utilize Archivematica, Archivematica has detailed instructions in their Documentation.

Arkivum is dedicated to regularly reviewing evolving technology and policies and implementing them accordingly. Arkivum also relies on customer feedback and client audits, stating that, "The security and audit model above has been developed in partnership with Arkivum customers who have confirmed that the model meets their regulatory requirements as part of a due-diligence/audit process that they have conducted on Arkivum. This includes due-diligence by customers in clinical and financial sectors where regulation is strict" (Audit Trails).

Access to Arkivum material after it has been processed by Archivematica is via an on-site gateway application, which requires little technical experience. More information on access to Arkivum can be found in their FAQ document, beginning on page 14 of Version 2.2. Arkivum may offer additional training and materials in their contract with the University of Oxford but will not ultimately be responsible for the clients' understanding of Archivematica.

Archivematica: Archivematica is open-source and provides generous documentation and instructions for every available option. See their Documentation page for more details. They also have various documentation on Error Handling and Error Reporting. See section A3.2 for information on their interaction with their user community.

Evidence: Retention of individuals with the discipline expertise; periodic assembly of designated or outside community members to evaluate and identify additional required metadata.

B2.11 Repository verifies each AIP for completeness and correctness at the point it is generated.

Audit Rating: 4

DS: Digital Safe relies on Arkivum to store and protect the AIP; on Archivematica to carry out the AIP generation and storage correctly and completely; and on the client for following due instructions on AIP generation and review. Digital Safe can offer recommendations but responsibility ultimately falls on the client.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). If the client chooses to utilize Archivematica, Archivematica is responsible for carry out the AIP process completely and correctly.

Once materials have been archived in Arkivum, Arkivum offers audit trails as a service, which will log any failures and successes. These must be turned on and specified by the client upon contract agreement. Audit trails are "accessible in machine-readable XML format through Arkivum REST API calls, either on a per-file or per-folder basis. Audit trails are also accessible in human readable PDF/A format through the Arkivum Service web interface or through an API call. Audit trails in PDF/A format can be signed if necessary to show that they were generated by Arkivum Service." More detailed information can be found on their Audit Trails page.

Archivematica: A SIP is generated from the transfer created by the client (see Transfer page) and then ultimately approved by the client. See section B1.1 for more information. Once a SIP has been created in Archivematica it "runs through a number of micro-services, including processing of the submission documentation, generation of the METS file, indexing, generation of the DIP and packaging of the AIP," which is packaged according to Bagit specifications (Store AIP), after which the client may review and/or download the AIP and its contents. The client may then choose to remove or store the AIP after review. The completion of all processes in Archivematica are indicated by color and text. Green indicates that a process has been completed successfully, and red indicates that the process has not been completed successfully.

Archivematica also utilizes fixity checks before AIP storage. "Archivematica generates checksums upon transfer of objects into the system, and will verify those checksums before storing the AIP. It is also possible to include pre-existing checksums, which Archivematica will also verify. To check fixity of AIPs in storage, Artefactual has written a separate command-line app called Fixity (Archivematica FAQs).

The client has complete control over the workflow process and can monitor at what stage the content is in as it moves through the workflow and once a client has established their digital preservation workflow, this will be recorded in their own policies. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

Evidence: Description of the procedure that verifies completeness and correctness; logs of the procedure.

B2.12 Repository provides an independent mechanism for audit of the integrity of the repository collection/content.

Audit Rating: 4

DS: Digital Safe relies on Arkvium to maintain its ISO 27001 certification and regular audit checks of data, methods, technology, and physical locations, and recognizes a breach in contract if Arkivum does not maintain certification.

Arkivum: Arkivum is ISO 27001 certified and "conforms to the controls within ISO 27002 to maintain its certifications against ISO 27001" (Security Model). Additionally, "The data centres used by Arkivum are Tier 2 or Tier 3 and are ISO 27001 certified or have FACT accreditation. They are inspected by Arkivum on a regular basis and have also been inspected by our ISO 27001 auditor" (Security Model). Other

audited sections include the production system access, building access, and logical access to data. For more information, see Arkivum's security model page (Security Model).

Additionally, Arkivum does not allow the editing of files once they are uploaded. "Arkivum provides an archiving service and an important feature is that once files are written to the archive then they become immutable. This is commonly known as WORM (Write Once Ready Many) and is a feature of many archive systems to ensure the integrity and authenticity of content. This means that once a file is written into our archive then it cannot be changed, for example edited to create a new version. If multiple versions of the same file need to be kept then they will need to be stored as separate files," (FAQ, 22).

Furthermore, checksums are used to verify that each file is correct and complete. For Arkivum/100 there is also the added security of the checksums generated by Arkivum being compared to the client's checksums, though this is not included in the Arkivum/1+1 product. The multiple copies of data and active data verification via annual data integrity checks complete the workflow to ensure that the data is correct. See Arkivum's archiving process for more information (Stage 2, Archiving Process).

Finally, Arkivum offers audit trails as a service. These must be turned on and specified by the client upon contract agreement. Audit trails are "accessible in machine-readable XML format through Arkivum REST API calls, either on a per-file or per-folder basis. Audit trails are also accessible in human readable PDF/A format through the Arkivum Service web interface or through an API call. Audit trails in PDF/A format can be signed if necessary to show that they were generated by Arkivum Service." More detailed information can be found on their Audit Trails page.

Archivematica: Archivematica utilizes fixity checks before AIP storage. "Archivematica generates checksums upon transfer of objects into the system, and will verify those checksums before storing the AIP. It is also possible to include pre-existing checksums which Archivematica will also verify. To check fixity of AIPs in storage, Artefactual has written a separate command-line app called Fixity (Archivematica FAQs). Clients and other organizations may also conduct a software audit on Archivematica.

The client has complete control over the workflow process and can monitor at what stage the content is in as it moves through the workflow and once a client has established their digital preservation workflow, this will be recorded in their own policies. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

Once content leaves Archivematica after client approval and is imported into Arkivum, Archivematica no longer bears responsibility to perform fixity checks.

All: See section B2.1-B2.6 for additional information.

Evidence: Documentation provided for B2.1 through B2.6; documented agreements negotiated between the producer and the repository (see B 1.1-B1.9); logs of material received and associated action (receipt, action, etc.) dates; logs of periodic checks.

B2.13 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (AIP creation).

Audit Rating: 4

DS: Digital Safe relies on Arkvium to maintain its ISO 27001 certification and regular audit checks of data, methods, technology, and physical locations, and recognizes a breach in contract if Arkivum does not maintain certification. Digital Safe can recommend best practices to clients, but ultimately the client chooses to use audit trails on their data.

Arkivum: Arkivum offers audit trails as a service. These must be turned on and specified by the client upon contract agreement. Audit trails are "accessible in machine-readable XML format through Arkivum REST API calls, either on a per-file or per-folder basis. Audit trails are also accessible in human readable PDF/A format through the Arkivum Service web interface or through an API call. Audit trails in PDF/A format can be signed if necessary to show that they were generated by Arkivum Service." More detailed information can be found on their Audit Trails page.

Archivematica: Archivematica utilizes fixity checks on files before AIP storage. See B2.11 for more information about Archivematica's fixity program. The client has complete control over the workflow process and can monitor at what stage the content is in as it moves through the workflow and once a client has established their digital preservation workflow, this will be recorded in their own policies. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

Evidence: Written documentation of decisions and/or action taken; preservation metadata logged, stored, and linked to pertinent digital objects.

B3. Preservation planning

B3.1 Repository has documented preservation strategies.

Audit Rating: 3

DS: Digital Safe has bears no responsibility for the choice in preservation strategies of the client. Digital Safe can recommend best practices for using Arkivum and Archivematica, but ultimately it is the responsibility of the client.

Arkivum: Arkivum follows the LTO roadmap for storage obsolescence and aims to prevent data loss by determining what LTO generation their system is and introducing new generations well in advance to any system failure. Escrow copies are also migrated every five years to prevent data loss. More information on their workflow and policies can be found on Arkivum's short term Maintaining Data Integrity page and long-term Data Integrity page. Also see section A3.8 for more information.

In regards to the digital preservation strategies, Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). If the client chooses to utilize Archivematica for digital preservation strategies, Archivematica is responsible for providing documentation.

Archivematica: Archivematica provides detailed documentation on installation, configuration, and use of this tool, found on their Documentation page. It is important to note that the Preservation Planning page is advice for users on how to construct and/or handle their own preservation policies and not a reflection of Archivematica's, or Digital Safe's, policies.

Evidence: Documentation identifying each preservation issue and the strategy for dealing with that issue.

B3.2 Repository has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable.

Audit Rating: 4

DS: Digital Safe relies on Arkivum and Archivematica to maintain their certification and updated format registries accordingly.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23), and is therefore only concerned with hardware, software, and data obsolescence. Arkivum can ingest any file format, but does not record their representation information. If the client chooses to utilize Archivematica, Archivematica can determine file formats and normalize using their Format Policy Registry, as well as creating SIPs and AIPs for record this process. See B3.1 for more information. Arkivum relies on Archivematica to adhere to any format registry updates.

Archivematica: Archivematica contains a Format Policy Registry (FPR) that contains the default format policies and is maintained by Artefactual Systems, Inc (which as of July, 2016 does not have a public interface yet. See B2.8 for more information). This system also allows for clients to define their format policies in a local FPR that is accessible via the FPR server maintained by Artefactual. Archivematica is also committed to updated format policies as standards evolve; "A format policy indicates the actions, tools and settings to apply to a digital object of a particular format (e.g. conversion to preservation format, conversion to access format, extraction of package formats). Format policies will change over time as local and community standards, practices and tools evolve" (FPR section). For additional information on the FPR and configuring a local FPR, see the Preservation Planning page, and the full FPR page.

Evidence: Subscription to a format registry service; subscription to a technology watch service; percentage of at least one staff member dedicated to monitoring technological obsolescence issues.

B3.3 Repository has mechanisms to change its preservation plans as a result of its monitoring activities.

Audit Rating: 3

DS: Digital Safe bears no responsibility for Arkivum and Archivematica to update their preservation mechanisms and plans. Digital Safe will update any new material for training, best practices recommendations, and announcements as needed.

Arkivum: Arkivum developed an automatic chain of custody system that prevents storage hardware and software obsolescence. Arkivum follows the LTO roadmap for storage obsolescence and aims to prevent data loss by determining what LTO generation their system is and introducing new generations well in advance to any system failure. Escrow copies are also migrated every five years to prevent data loss, which "includes using new drives when new generations of media are introduced into the system, which ensures that the media/drive combination is never near to end of life." More information on their workflow and policies can be found on Arkivum's short term Maintaining Data Integrity page and long-term Data Integrity page. Also see section A3.8 for more information.

Archivematica: See response for section B3.2 for information on their Format Policy Registry and their recommendations for preservation planning. Archivematica is a digital preservation workflow tool and not a storage service, and in the event of Archivematica ceasing to operate, there would be no chance of

data loss. If Archivematica requires an update, more information can be found in their Installing from packages section.

Evidence: Preservation planning policies tied to formal or information technology watch(es); preservation planning or processes that are timed to shorter intervals (e.g., not more than five years); proof of frequent preservation planning/policy updates.

B3.4 Repository can provide evidence of the effectiveness of its preservation planning.

Audit Rating: 2

Digital Safe: Digital Safe is currently in the planning stages. The goals of Phase 3 aim to develop an agreement with Arkivum, design a business model with IT Services, and cover up-front costs for the service (PID, 2). The service does not yet exist in the ideal form that the Digital Safe Steering Committee is aiming to accomplish, so there is not yet the means for evidencing the effectiveness of the model. This will be measured once Phase 3 has been implemented and the beta service has been deployed to early adopters.

Arkivum: Arkivum is a storage platform that does not provide digital preservation activities. In regards to long-term storage success, Arkivum has several well-known institutions as clients, including the Museum of Modern Art, University of Westminster, and the Oxford Molecular Diagnostics Centre, among many others noted in case studies on their website that have been successful. Their Solutions tab offers several reports and case studies in various fields, such as Higher Education, that are evidence of success.

Archivematica: First, Archivematica is open-source and provides generous documentation and instructions for every available option. See their Documentation page for more details. See section A3.2 and A4.2 for information on their interaction with their user community, on which they rely to ensure that the product is effective and efficient. This interaction has led to extensive documentation, including in Error Handling and Error Reporting. They also monitor their user community to stay current on their Format Policy Registry (FPR section). For additional information on the FPR and configuring a local FPR, see the Preservation Planning page, and the full FPR page.

Evidence: Collection of appropriate preservation metadata; proof of usability of randomly selected digital objects held within the system; demonstrable track record for retaining usable digital objects over time.

B4. Archival storage & preservation/maintenance of AIPs

B4.1 Repository employs documented preservation strategies.

Audit Rating: 4

DS: Digital Safe chose the Arkivum and Archivematica solution because of their ability to preserve, normalize, and store any file format. See sections B2.1, B2.7, and B2.8 for more information. Digital Safe relies on Arkivum and Archivematica to employ their documented preservation strategies.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). The AIP is determined by the client as they choose what archival processes are a part of their workflow using

Archivematica. If the client chooses to utilize Archivematica, Archivematica will be responsible for employing their documented preservation strategies.

Arkivum is designed to hold AIPs. "The persistent file/folder mechanism within the Arkivum service offers excellent support for the storage of AIPs. The file will retain its original filename, and checksums are provided for incorporation in the Preservation Description Information (PDI) for the AIP. The service follows the OAIS model for Archive Storage through use of replication, fixity monitoring and repair, disaster recovery, migration, and tiered storage to deliver a specified level of performance, availability and integrity of storage" (Integration with Other Systems).

Archivematica: Archivematica: If the client decides to utilize the Archivematica tool they may choose to store an AIP. The client must determine the processing configuration, which can be left at Archivematica's default setting, or can be created by the client. See Archivematica's Processing Configuration documentation for more details. See section B2.1, B2.2, and B2.3 for more information.

Evidence: Documentation of strategies and their appropriateness to repository objects; evidence of application (e.g., in preservation metadata); see B3.3.

B4.2 Repository implements/responds to strategies for archival object (i.e., AIP) storage and migration.

Audit Rating: 4

DS: Digital Safe chose the Arkivum and Archivematica solution because of their ability to preserve, normalize, and store any file format. See sections B2.1, B2.7, and B2.8 for more information. Digital Safe relies on Arkivum and Archivematica to employ their documented preservation strategies. Digital Safe will update any new material for training, best practices recommendations, and announcements as needed. Archivematica is an optional tool and it is the responsibility of the client to utilize Archivematica's AIP generation and storage.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). The AIP is determined by the client as they choose what archival processes are a part of their workflow using Archivematica. If the client chooses to utilize Archivematica, Archivematica will be responsible for employing their documented preservation strategies.

Arkivum is designed to hold AIPs. "The persistent file/folder mechanism within the Arkivum service offers excellent support for the storage of AIPs. The file will retain its original filename, and checksums are provided for incorporation in the Preservation Description Information (PDI) for the AIP. The service follows the OAIS model for Archive Storage through use of replication, fixity monitoring and repair, disaster recovery, migration, and tiered storage to deliver a specified level of performance, availability and integrity of storage" (Integration with Other Systems). See section A3.2 for more information on certification to safely store AIPs.

Archivematica: If the client decides to utilize the Archivematica tool they may choose to store an AIP. The client must determine the processing configuration, which can be left at Archivematica's default setting, or can be created by the client. See Archivematica's Processing Configuration documentation for more details.

After configuring the process as desired, the SIP can be normalized and stored in an AIP. "After normalization is approved, the SIP runs through a number of micro-services, including processing of the

submission documentation, generation of the METS file, indexing, generation of the DIP and packaging of the AIP," which is packaged according to Bagit specifications (AIP Storage). Detailed information on the structure of the AIP can be found in Archivematica's AIP Structure documentation. The client may review the AIP and proceed to storing the AIP. See section B2.2 and B2.3 for more information on AIP packaging and storage.

The content is then moved to the storage facility of the clients' choice, in this case Arkivum, via A-stor.

Archivematica relies on their Format Policy Registry to develop their workflow tool, and are committed to updated format policies as standards evolve. For additional information on the FPR and configuring a local FPR, see the Preservation Planning page, and the full FPR page.

Evidence: Institutional technology and standards watch; demonstration of objects on which a preservation strategy has been performed; demonstration of appropriate preservation metadata for digital objects.

B4.3 Repository preserves the Content Information of archival objects (i.e., AIPs).

Audit Rating: 4

DS: Digital Safe relies on the client to provide the archival object, Archivematica to preserve and package the archival object, and Arkivum to store and maintain the archival object. Deletion of any archival objects or their associated content information is at the discretion of the client. The client will rely on Arkivum and Archivematica for instructions on how to delete files.

Arkivum: First, any original client files are only deleted after receiving the "Green" light form Arkivum. Clients are notified at what stage their data is at during ingestion using a "traffic light system" where Red indicates that the client copy must not be deleted; Amber indicates that the ingested files are at the Arkivum data centers; and Green indicates that the ingested files are replicated and protected in the prescribed data center(s) and in escrow and that the client file can be deleted. See the FAQ for more information (FAQ, 17).

Data may be deleted from Arkivum after one year of storage. This is managed by the client's administration following a retention review or as required by the contract. There are four levels to removing data from Arkivum. First the key encryption is deleted so the data cannot be read. Second, the file is deleted to remove references to the file, making the file difficult to retrieve. Third, the data is securely erased from the storage media by overwriting so that the data cannot be recovered. Finally, the storage media is physically destroyed. For more detailed information, see the FAQ page 16.

Any unwanted data that has, for example, a set amount of time to be kept before permanent deletion, is removed at the behest of the client. The AES key is purged, the encrypted data removed from the media, and the tapes are securely erased and taken out of service (FAQ, 10). Conversely, data that has been accidentally deleted can be recovered via the physical LTO data tape held in Escrow if the administration contact Arkivum immediately following the deletion. This will require the master encryption key and will cause a delay in the data being returned to the client (FAQ, 17).

Clients cannot delete data that they are viewing as the security system only allows administration to request deletion. The deletion must be approved by the administrator and carried out through the administrative web interface. Any attempt to delete is recorded and can be tracked to a user's Active Directory. (FAQ, 12).

Archivematica: AIPs may be deleted in Archivematica before and after they have been packaged. A user may choose to remove a SIP or AIP if they wish to start over. Once an AIP has been created and packaged, the deletion begins with a request in Archivematica. The client must enter a reason for deletion. The request is sent to the administrator and if it is approved, the data is removed. If the administrator does not approve, the AIP will remain in Archivematica. For more information, see the Deleting an AIP section.

Once the content has moved to Arkivum, Archivematica does not keep the files unless they are specifically saved there temporarily. For more information on storage, see their Storage Services page.

Evidence: Policy documents specifying treatment of AIPs and whether they may ever be deleted; ability to demonstrate the chain of AIPs for any particular digital object or group of objects ingested; workflow procedure documentation.

B4.4 Repository actively monitors integrity of archival objects (i.e., AIPs).

Audit Rating: 4

DS: Digital Safe relies on Arkivum to maintain its ISO 27001 certification and regular audit checks of data, methods, technology, and physical locations certification and recognizes a breach in contract if Arkivum does not maintain certification. Digital Safe can also recommend best practices to clients in regards to utilizing Arkivum's audit trails, but it is ultimately up to the user to turn the audit trails on.

Arkivum: Arkivum has policies and automated processes in place for running integrity checks on the data, software, and hardware. The data is retrieved annually and given an integrity test. Arkivum has also identified software and hardware obsolescence to occur on a cycle of generally 3 to 5 years, so Arkivum's policy is that data is migrated to new media following the LTO roadmap. The LTO data tapes in Escrow are also migrated every 5 years. More detailed information on media upgrades can be found on their Data Integrity page.

If the client chooses to utilize them, Arkivum offers audit trails as a service. These must be turned on and specified by the client upon contract agreement. Any integrity check would be logged in the audit trail and accessible to the client at any time. For more information on audit trails, see section B1.8 and Arkivum's Audit Trails page.

Archivematica: Archivematica also utilizes fixity checks before AIP storage. "Archivematica generates checksums upon transfer of objects into the system, and will verify those checksums before storing the AIP. It is also possible to include pre-existing checksums, which Archivematica will also verify. To check fixity of AIPs in storage, Artefactual has written a separate command-line app called Fixity (Archivematica FAQs).

The client has complete control over the workflow process and can monitor at what stage the content is in as it moves through the workflow and once a client has established their digital preservation workflow, this will be recorded in their own policies. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

The client has complete control over the workflow process and can monitor at what stage the content is in as it moves through the workflow. See the response for section B1.6 for information on process completion.

Evidence: Logs of fixity checks (e.g., checksums); documentation of how AIPs and Fixity information are kept separate.

B4.5 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Archival Storage).

Audit Rating: 4

DS: Digital Safe does not bear any responsibility for logging actions of content. Digital Safe does recommend the use of audit trails to clients, though the use of audit trails is determined by the client.

Arkivum: Arkivum offers audit trails as a service. These must be turned on and specified by the client upon contract agreement. Audit trails are "accessible in machine-readable XML format through Arkivum REST API calls, either on a per-file or per-folder basis. Audit trails are also accessible in human readable PDF/A format through the Arkivum Service web interface or through an API call. Audit trails in PDF/A format can be signed if necessary to show that they were generated by Arkivum Service." More detailed information concerning audit trails can be found at their Audit Trails page.

Archivematica: Archivematica utilizes fixity checks on files before AIP storage. See B2.11 for more information about Archivematica's fixity program. The client has complete control over the workflow process and can monitor at what stage the content is in as it moves through the workflow and once a client has established their digital preservation workflow, this will be recorded in their own policies. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

Evidence: Written documentation of decisions and/or action taken; preservation metadata logged, stored, and linked to pertinent digital objects.

B5. Information management

B5.1 Repository articulates minimum metadata requirements to enable the designated community(ies) to discover and identify material of interest.

Audit Rating: 4

See the answer for section B1.1 for detail on metadata properties and how they are handled, and section A3.1 for information on the identified user communities.

DS: Retrieving data is only accessible by the client. Digital Safe relies on the client to maintain the filenames they ingest for easy retrieval. Should the client wish to allow others to access materials, the client should describe the naming system. This may occur if, for example, a researcher is granted permission by the client to browse material not yet public.

Arkivum: Arkivum maintains the original filename and identifies the file internally for integrity checks by checksums. Clients retrieving information will search using their original file names.

Archivematica: Archivematica's naming system will retain the original name of the transfer unless a new name has been assigned to the SIP upon creation. This name will be combined with a Universal Unique

Identifier that is generated and assigned during SIP formation. For more detailed information on this and the structure of the AIP, see the AIP Structure page.

Evidence: Descriptive metadata.

B5.2 Repository captures or creates minimum descriptive metadata and ensures that it is associated with the archived object (i.e., AIP).

Audit Rating: 4

DS: Digital Safe is not responsible for metadata creation. See section B1.1 for Digital Safe's recommendation on metadata preservation. Digital Safe relies on Arkivum and Archivematica for associating any descriptive metadata with the digital object.

Arkivum: The AIP is determined by the client as they choose what archival processes are a part of their workflow. If the client has chosen to package an AIP, Arkivum is designed to hold AIPs. "The persistent file/folder mechanism within the Arkivum service offers excellent support for the storage of AIPs. The file will retain its original filename, and checksums are provided for incorporation in the Preservation Description Information (PDI) for the AIP. The service follows the OAIS model for Archive Storage through use of replication, fixity monitoring and repair, disaster recovery, migration, and tiered storage to deliver a specified level of performance, availability and integrity of storage" (Integration with Other Systems). If the client does not have any associated files, the digital object will be identified by its original filename and its checksum.

Archivematica: If the client chooses to utilize Archivematica and to create an AIP, "after normalization is approved, the SIP runs through a number of micro-services, including processing of the submission documentation, generation of the METS file, indexing, generation of the DIP and packaging of the AIP," which is packaged according to Bagit specifications (AIP Storage). The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure). More detailed information on the structure of the AIP can be found in Archivematica's AIP Structure documentation. This package of the digital object and its associated files is what is ingested into Arkivum.

Evidence: Descriptive metadata; persistent identifier/locator associated with AIP; system documentation and technical architecture; depositor agreements; metadata policy documentation, incorporating details of metadata requirements and a statement describing where responsibility for its procurement falls; process workflow documentation.

B5.3 Repository can demonstrate that referential integrity is created between all archived objects (i.e., AIPs) and associated descriptive information.

Audit Rating: 3

DS: Digital Safe is not responsible for metadata creation. See section B1.1 for Digital Safe's recommendation on metadata preservation. Digital Safe relies on Arkivum and Archivematica for associating any descriptive metadata with the digital object. Digital Safe does not yet have a best practices policy in place.

Arkivum: Arkivum is designed to hold AIPs but not to create them. "The persistent file/folder mechanism within the Arkivum service offers excellent support for the storage of AIPs. The file will retain its original

filename, and checksums are provided for incorporation in the Preservation Description Information (PDI) for the AIP. The service follows the OAIS model for Archive Storage through use of replication, fixity monitoring and repair, disaster recovery, migration, and tiered storage to deliver a specified level of performance, availability and integrity of storage" (Integration with Other Systems).

Archivematica: If the client chooses to utilize Archivematica and to create an AIP it is packaged according to Bagit specifications (AIP Storage). Detailed information on the structure of the AIP can be found in Archivematica's AIP Structure documentation. The filename of the AIP is created using the original name of the transfer, unless a new name has been assigned to the SIP upon creation, and then combined with a Universal Unique Identifier that is generated and assigned during SIP formation. The directory and the METS file carry the UUID, and the object and thumbnail (if necessary). The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure). More detailed information on the structure of the AIP can be found in Archivematica's AIP Structure documentation.

Evidence: Descriptive metadata; persistent identifier/locator associated with AIP; documented relationship between AIP and metadata; system documentation and technical architecture; process workflow documentation.

B5.4 Repository can demonstrate that referential integrity is maintained between all archived objects (i.e., AIPs) and associated descriptive information.

Audit Rating: 3

DS: Digital Safe relies on Arkivum and Archivematica for associating any descriptive metadata with the digital object. Digital Safe does not yet have a best practices policy in place.

Arkivum: See section A3.2 for the response to integrity monitoring, and section B1.8 and B4.5 for the response to audit trails. If the client chooses to turn on the audit trails, this will log any changes in integrity, but does not track the creation of referential integrity.

Archivematica: See section B5.3 for the response to creating referential integrity and how it is maintained.

Evidence: Log detailing ongoing monitoring/checking of referential integrity, especially following repair/modification of AIP; legacy descriptive metadata; persistence of identifier/locator; documented relationship between AIP and metadata; system documentation and technical architecture; process workflow documentation.

B6. Access management

B6.1 Repository documents and communicates to its designated community(ies) what access and delivery options are available.

Audit Rating: 3

DS: Though it does not yet exist, it is recommended that University web space for Digital Safe is created to briefly describe the key users identified by phase 1 and briefly explain why accessibility is limited to the clients of the University. See section A3.1 and A3.4 for detailed information on Digital Safe's

identified key users. Ideally this web space would also provide brief information on the funding of the Digital Safe service.

Access parameters will be the responsibility of the client. These parameters should be outlined by the client in their own policy, but is not the responsibility of Digital Safe. Digital Safe relies on Arkivum to maintain its security.

Arkivum: Access to materials will be strictly monitored. Individuals with the encryption keys, for example the college Archivist or a similar position, are the only individuals with access to any ingested content, and will determine additional user access. These users will have Active Directory permissions that can be integrated into individual segments of the archive. More information can be found in the Arkivum FAQ document on page 10. Should the client wish to allow others to access materials, the client will set access parameters. This may occur if, for example, a researcher is granted permission by the client to browse material not yet public.

Archivematica: Archivematica is a digital preservation workflow tool, not a storage space. Both the workflow and the content are not accessible to users who are not specifically the client developing and using the workflow.

Evidence: Public versions of access policies; delivery policies; fee policies.

B6.2 Repository has implemented a policy for recording all access actions (includes requests, orders etc.) that meet the requirements of the repository and information producers/depositors.

Audit Rating: 3

DS: Digital Safe conducted user studies in Phase 2 on various technologies, including Arkivum and the beta interface developed for Digital Safe, but not since the integration of Arkivum and Archivematica. The first year of early adopters testing the service, user feedback will be the basis of any improvements to the service and if the service launches for the entire University. Digital Safe has not developed an official policy on access actions. It is recommended that contact information and a Help and Feedback section are included in the University web space for the long-term.

See section B6.1 for more information on clients determining access parameters.

Arkivum: See section B6.1 for information on the access policy for Arkivum. See section B1.8 and B4.5 for the response to audit trails. If the client chooses to turn on the audit trails, this will log any access actions.

Archivematica: Archivematica is a digital preservation workflow tool, not a storage space. Both the workflow and the content are not accessible to users who are not specifically the client developing and using the workflow.

Archivematica utilizes fixity checks on files before AIP storage. See B2.11 for more information about Archivematica's fixity program. The client has complete control over the workflow process and can monitor at what stage the content is in as it moves through the workflow. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

Evidence: Access policies; use statements.

B6.3 Repository ensures that agreements applicable to access conditions are adhered to.

Audit Rating: 3

DS: Digital Safe will evaluate all user feedback and make changes as is reasonable and possible to create an archive space that fits their access needs. The policy for Digital Safe relies on the client to utilize audit trails, which are recommended by Digital Safe, and to set their own access parameters. Digital Safe also relies on Arkivum and Archivematica to maintain their security measures.

Arkivum: See section B1.8 for information on Audit Trails, which will track all handling of the content. Clients cannot delete data that they are viewing as the security system only allows administration to request deletion or to alter a file. The deletion must be approved by the administrator and carried out through the administrative web interface. Any attempt to delete or alter is recorded and can be tracked to a user's Active Directory (FAQ, 12). See section B4.3 for more information for tracking deletion and other unauthorized activities.

Archivematica: Archivematica utilizes fixity checks on files before AIP storage. See B2.11 for more information about Archivematica's fixity program. The client has complete control over access to the workflow and the workflow process and can monitor at what stage the content is in as it moves through the workflow and once a client has established their digital preservation workflow, this will be recorded in their own policies. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

Evidence: Access policies; logs of user access and user denials; access system mechanisms that prevent unauthorized actions (such as save, print, etc.); user compliance agreements.

B6.4 Repository has documented and implemented access policies (authorization rules, authentication requirements) consistent with deposit agreements for stored objects.

Audit Rating: 2

DS: Digital Safe does not monitor the deposited content, but will need to determine any preservation rights and copyrights and dictate them in the contract with Arkivum. These details have not yet been established and will be a priority during Phase 3. See section A5 for more information. Digital Safe relies on the client to have the appropriate permissions for any content that was not generated by them, BEAM for example, and will adhere to their permissions policies in the Arkivum contract.

Arkivum: See section B1.8 for information on Audit Trails, which will track all handling of the content. "Access to files through the filesystem exposed by the Arkivum service appliance on the customer site is controlled through file permissions and Active Directory" (Security Model) and the Active Directory is tracked by audit trails.

Archivematica: Only the client has access to manipulating and using the workflow. The AIP contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure) which will log any access actions. Archivematica is integrated with Arkivum via Arkivum's A-Stor, which will be built into the contract between the University and Arkivum, and is therefore protected by the security measures of Arkivum. To see more about the integration of Arkivum with Archivematica, see their Storage Services page.

Evidence: Access validation mechanisms within system; documentation of authentication and validation procedures.

B6.5 Repository access management system fully implements access policy.

Audit Rating: 2

DS: Digital Safe relies on the client to set their security parameters and on Arkivum to maintain those parameters.

Arkivum: See section B1.8 for information on Audit Trails, which will track all handling of the content. "Access to files through the filesystem exposed by the Arkivum service appliance on the customer site is controlled through file permissions and Active Directory" (Security Model) and the Active Directory is tracked by audit trails. Anyone provided with an authorized Active Directory by the administrator may access. Unauthorized users will be denied access into Digital Safe, and even they have access to the system, they will also need an authorized Active Directory to access the files. "Access to files through the filesystem exposed by the Arkivum service appliance on the customer site is controlled through file permissions and Active Directory" (Security Model). Any attempt to delete is also recorded and can be tracked to a user's Active Directory. (FAQ, 12).

Archivematica: Archivematica utilizes fixity checks on files before AIP storage. See B2.11 for more information about Archivematica's fixity program. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure). Archivematica is integrated with Arkivum via Arkivum's A-Stor, which will be built into the contract between the University and Arkivum, and is therefore protected by the security measures of Arkivum. To see more about the integration of Arkivum with Archivematica, see their Storage Services page.

Evidence: Logs and audit trails of access requests; information about user capabilities (authentication matrices); explicit tests of some types of access.

B6.6 Repository logs all access management failures, and staff review inappropriate "access denial" incidents.

Audit Rating: 3

DS: Digital Safe does not monitor access management or review unauthorized incidents. The client is responsible for monitoring their access management and addressing any unauthorized incidents. Digital Safe relies on the technologies to notify the clients of any access management failures.

Arkivum: See section B1.8 for information on Audit Trails, which will track all handling of the content. Clients cannot delete data that they are viewing as the security system only allows administration to request deletion. The deletion must be approved by the administrator and carried out through the administrative web interface. Any attempt to delete is recorded and can be tracked to a user's Active Directory. (FAQ, 12). Arkivum notifies the administrator.

Archivematica: Alterations may occur during the workflow process, but deletions begin with a request through Archivematica. The client must enter a reason for deletion. The request is sent to the administrator and if it is approved, the data is removed. If the administrator does not approve, the AIP will remain in Archivematica. For more information, see the Deleting an AIP section. The client is responsible for acting on these notifications and may change access permissions.

Archivematica utilizes fixity checks on files before AIP storage. See B2.11 for more information about Archivematica's fixity program. The AIP also contains a /data/logs folder with transfers, normalization, malware scan, and extraction information (AIP Structure).

Evidence: Access logs; capability of system to use automated analysis/monitoring tools and generate problem/error messages; notes of reviews undertaken or action taken as result of reviews.

B6.7 Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is completed in relation to the request.

Audit Rating: 3

DS: Digital Safe relies on the technologies to produce the correct and complete DIP. As the service is not yet active, there are no examples to test this process. It may also be possible that the client has not chosen to create a DIP and is accessing the original file or the AIP.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). Arkivum can store a DIP but does not create DIPs. Arkivum will retrieve the DIP is the same format and state that it was in upon ingestion. This can be monitored by checksums.

Archivematica: If the client chooses to create a DIP, they indicate it during the normalization process and the access copies used to create it are also created during normalization (Normalization process). DIPs may be part of the AIP, or can be uploaded separately to Arkivum via Arkivum's A-Stor. For more information on storage, see their Storage Services page. For more information on DIP storage, see their Store DIP section.

Evidence: System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; logs of orders and DIP production; test accesses to verify delivery of appropriate digital objects.

B6.8 Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is correct in relation to the request.

Audit Rating: 3

DS: Digital Safe relies on the client to set and monitor the access parameters, and on the technologies to produce the correct and complete DIP. As the service is not yet active, there are no examples to test this process. It may also be possible that the client has not chosen to create a DIP and is accessing the original file or the AIP.

Arkivum: Arkivum "provides safe and secure data archiving," not digital preservation (FAQ, 23). Arkivum can store a DIP but does not create DIPs. Arkivum will retrieve the DIP is the same format and state that it was in upon ingestion. This can be monitored by checksums.

Archivematica: If the client chooses to create a DIP, the DIP is generated directly from the AIP and will have the same UUID associated with it. DIPs may be part of the AIP, or can be uploaded separately to Arkivum via Arkivum's A-Stor. For more information on storage, see their Storage Services page. For more information on DIP storage, see their Store DIP section.

Evidence: System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; logs of orders and DIP production.

B6.9 Repository demonstrates that all access requests result in a response of acceptance or rejection.

Audit Rating: 4

DS: Digital Safe relies on the client to set and monitor the access parameters, and on the technologies to maintain their security measures. As the service is not yet active, there are no examples to test this process. It may also be possible that the client has not chosen to create a DIP and is accessing the original file or the AIP.

Arkivum: Anyone with an authorized Active Directory may access. Unauthorized users will be denied access into Digital Safe, and even they have access to the system, they will also need an authorized Active Directory to access the files. "Access to files through the filesystem exposed by the Arkivum service appliance on the customer site is controlled through file permissions and Active Directory" (Security Model). As long as the client is accessing Arkivum during their scheduled time they will have access to all of their files.

Archivematica: The completion of all processes in Archivematica are indicated by color and text. Green indicates that a process has been completed successfully, and red indicates that the process has not been completed successfully. Access is determined by the client; all actions must be authorized by the administrator, and the administrator is notified of any unauthorized activity.

Evidence: System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; logs of orders and DIP production.

B6.10 Repository enables the dissemination of authentic copies of the original or objects traceable to originals.

Audit Rating: 3

DS: Digital Safe is a service for the long-term storage high-security data that will not be disseminated in the near future or ever. Student records and research data may be released in the far future, but other data may never be publicly accessible as it may be permanently deleted. Digital Safe relies on the client to upload authentic material and to set and manage their access parameters. Digital Safe will also provide training and general best-practice policies, but relies on the technologies to retrieve the correct and complete file.

As the service is not yet active, there are no examples to test this process. It may also be possible that the client has not chosen to create a DIP and is accessing the original file or the AIP.

Arkivum: Arkivum ensures that all of the ingested files are correct and complete upon retrieval. Akrivum has policies and automated processes in place for running integrity checks on the data, software, and hardware to prevent data corruption or loss. The data is retrieved annually and given an integrity test using checksums. Arkivum has also identified software and hardware obsolescence to occur on a cycle of generally 3 to 5 years, so Arkivum's policy is that data is migrated to new media following the LTO roadmap. The LTO data tapes in Escrow are also migrated every 5 years. More detailed information on media upgrades can be found on their Data Integrity page. See section B1.4 for additional information.

Arkivum maintains high security both digital and physically, and access to this content will be minimal. See section B1.5 for detailed information on the security measures.

Archivematica: Archivematica's naming system will retain the original name of the transfer unless a new name has been assigned to the SIP upon creation. This name will be combined with a Universal Unique Identifier that is generated and assigned during SIP formation. For more detailed information on this and the structure of the AIP, see the AIP Structure page. The Archivematica directory is searched by this name.

Archivematica also interacts with their user community for providing a product that suits their needs. See A3.4 for more information on their communication with their user community, or view their user forum.

Evidence: System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; production of a sample authenticated copy; documentation of community requirements for authentication.

C. Technologies, Technical Infrastructure, & Security

Average Rating: 3.1/4

C1. System Infrastructure

C1.1 Repository functions on well-supported operating systems and other core infrastructural software.

Audit Rating: 3

DS: Digital Safe relies on Arkivum for maintaining its operating systems. Digital Safe as a service will be available via the web page and on the client's local operating system.

Arkivum: Arkivum and its applications can be used from most common operating systems (FAQ, 18). Additional technical support is available for unique operating systems. Updates are automatically provided to clients as they develop and are applied by simple mouse click (FAQ, 11). If there is a fault or failure on the client's side, "a new system can be configured and archived data will still be available (FAQ, 12).

Archivematica: Archivematica's system "packages a customized Xubuntu environment as a virtual appliance, making it possible to run on top of any consumer-grade hardware and operating system" (Single install page). It is also possible to update Archivematica after installation. See their Installing from packages section for detailed information.

Evidence: Software inventory; system documentation; support contracts; use of strongly community supported software (i.e., Apache).

C1.2 Repository ensures that it has adequate hardware and software support for backup functionality sufficient for the repository's services and for the data held, e.g., metadata associated with access controls, repository main content.

Audit Rating: 3

DS: Digital Safe relies on Arkivum to maintain and carry out its back up procedures.

Arkivum: For file backup, Arkivum has policies and automated processes in place for running integrity checks on the data, software, and hardware. The data is retrieved annually and given an integrity test based on checksums. For hardware and software support, Arkivum follows the LTO roadmap for storage obsolescence and aims to prevent data loss by determining what LTO generation their system is and introducing new generations well in advance to any system failure. Escrow copies are also migrated every five years to prevent data loss. More information on their workflow and policies can be found on Arkivum's short term Maintaining Data Integrity page and long-term Data Integrity page. Also see section A3.8 for more information.

Archivematica: Archivematica is a digital preservation workflow tool. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. If necessary, Data can be backed up and temporarily stored using MySQL. More detailed instructions can be found in their Data back-up section.

Evidence: Documentation of what is being backed up and how often; audit log/inventory of backups; validation of completed backups; disaster recovery plan—policy and documentation; "firedrills"—testing of backups; support contracts for hardware and software for backup mechanisms.

C1.3 Repository manages the number and location of copies of all digital objects.

Audit Rating: 3

DS: Digital Safe relies on Arkivum to maintain the copies of digital objects. As the service is not yet active, there are no examples to test this process.

Arkivum: For Arkivum/1+1, one copy is stored in a secure data center and one copy is saved on LTO data tape in Escrow. For Arkivum/100, two copies are stored in secure, geographically separate data centers and one copy is saved on LTO data tape in Escrow. Arkivum maintains high security both digital and physically. According to their FAQ documentation. "All copies of customer data are held in secure UK storage locations. Storage facilities are manned at all hours and access is strictly restricted to a list of named, trained and vetted members of the Arkivum Operations team. Our operations at all sites, including our business offices, is certified to ISO 27001 information security standard." In addition to ISO27001 certification and industry best practice for security, our customer base includes people using our service to store personal data including voice call recordings and medical treatment records. They have audited our service and satisfied themselves that our service is secure and meets their regulatory and legislative obligations" (FAQ, 19). Data is also encrypted once it leaves the client's network and passes through a secure VPN before entering a data center. The Escrow copy is located based on the contract between the client and Arkivum (FAQ, 5). For more information, see the FAQ documentation and section B1.5. The data is also retrieved annually and given an integrity test based on checksums (long-term Data Integrity).

Archivematica: Archivematica is a digital preservation workflow tool. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. Archivematica also utilizes fixity checks before AIP storage. "Archivematica generates checksums upon transfer of objects into the system, and will verify those checksums before storing the AIP. It is also possible to include pre-existing checksums, which Archivematica will also verify. To check fixity of AIPs in storage, Artefactual has written a separate command-line app called Fixity (Archivematica FAQs).

Evidence: random retrieval tests; system test; location register/log of digital objects compared to the expected number and location of copies of particular objects.

C1.4 Repository has mechanisms in place to ensure any/multiple copies of digital objects are synchronized.

Audit Rating: 4

DS: Digital Safe relies on Arkivum to maintain synchronized copies. As the service is not yet active, there are no examples to test this process.

Arkivum: The data is retrieved annually and given an integrity test based on checksums (long-term Data Integrity). Any corruption or loss is automatically repaired and immediately makes a copy of the correct object to replace the corrupted object.

Archivematica: Archivematica is a digital preservation workflow tool. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. If necessary, Data can be backed up and temporarily stored using MySQL. More detailed instructions can be found in their Data back-up section.

Evidence: Workflows; system analysis of how long it takes for copies to synchronize; procedures/documentation of operating procedures related to updates and copy synchronization; procedures/documentation related to whether changes lead to the creation of new copies and how those copies are propagated and/or linked to previous versions.

C1.5 Repository has effective mechanisms to detect bit corruption or loss.

Audit Rating: 3

DS: Digital Safe relies on Arkivum to manage bit loss and corruption. As the service is not yet active, there are no examples to test this process.

Arkivum: The data is retrieved annually and given an integrity test based on checksums and preventative data migrations. See section A3.2 for more information, as well as Arkivum's long-term Data Integrity page. "Each copy has its integrity actively monitored and managed and any corruption or loss is automatically repaired to make the system self-healing" (Overview).

Archivemetica: Archivematica is a digital preservation workflow tool. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. Clients are notified of any failed actions and are responsible for managing these issues. Archivematica also utilizes fixity checks before AIP storage. "Archivematica generates checksums upon transfer of objects into the system, and will verify those checksums before storing the AIP. It is also possible to include pre-existing checksums, which Archivematica will also verify. To check fixity of AIPs in storage, Artefactual has written a separate command-line app called Fixity (Archivematica FAQs).

Evidence: Documents that specify bit error detection and correction mechanisms used; risk analysis; error reports; threat analyses.

C1.6 Repository reports to its administration all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data.

Audit Rating: 3

DS: Digital Safe relies on Arkivum and Archivematica to notify and handle any incidents of data loss or corruptions. As the service is not yet active, there are no examples to test this process.

Arkivum: The data is retrieved annually and given an integrity test based on checksums and preventative data migrations. See section A3.2 for more information, as well as Arkivum's long-term Data Integrity page. "Each copy has its integrity actively monitored and managed and any corruption or loss is automatically repaired to make the system self-healing" (Overview).

Archivematica: See the response for section C1.5.

Evidence: Preservation metadata (e.g., PDI) records; comparison of error logs to reports to administration; escalation procedures related to data loss.

C1.7 Repository has defined processes for storage media and/or hardware change (e.g., refreshing, migration).

Audit Rating: 3

DS: Digital Safe relies on Arkivum to maintain its policies on hardware.

Arkivum: Arkivum follows the LTO roadmap for storage obsolescence and aims to prevent data loss by determining what LTO generation their system is and introducing new generations well in advance to any system failure, and plan to migrate every 3-5 years. Escrow copies are also migrated every five years to prevent data loss. More information on their workflow and policies can be found on Arkivum's short term Maintaining Data Integrity page and long-term Data Integrity page. Also see section A3.8 for more information.

Archivematica: See the response for section C1.5.

Evidence: Documentation of processes; policies related to hardware support, maintenance, and replacement; documentation of hardware manufacturers' expected support life cycles.

C1.8 Repository has a documented change management process that identifies changes to critical processes that potentially affect the repository's ability to comply with its mandatory responsibilities.

Audit Rating: 3

DS: Digital Safe relies on the technologies to maintain their ISO certification and to update and implement their policies. Digital Safe as a service will remain active and give due notice of any maintenance issues or other cessations in operation that will affect the clients and their data.

Arkivum: See section C1.7 and A3.8 for information on LTO roadmap use. Furthermore, according to Arkivum "Our operations at all sites, including our business offices, is certified to ISO 27001 information

security standards," (FAQ, 19) and Arkivum is regularly audited externally to maintain ISO 27001 certification and welcomes client audits as well (FAQ, 10). Every change is monitored and tested bi-annually.

Archivematica: Archivematica is a digital preservation workflow tool. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. For policy and digital preservation activities, Archivematica is committed to updated format policies as standards evolve; "A format policy indicates the actions, tools and settings to apply to a digital object of a particular format (e.g. conversion to preservation format, conversion to access format, extraction of package formats). Format policies will change over time as local and community standards, practices and tools evolve" (FPR section). For additional information on the FPR and configuring a local FPR, see the Preservation Planning page, and the full FPR page.

Evidence: Documentation of change management process; comparison of logs of actual system changes to processes versus associated analyses of their impact and criticality.

C1.9 Repository has a process for testing the effect of critical changes to the system.

Audit Rating: 3

See the response for C1.8. As the service is not yet active, there are no examples to test this process.

Evidence: Documented testing procedures; documentation of results from prior tests and proof of changes made as a result of tests.

C1.10 Repository has a process to react to the availability of new software security updates based on a risk-benefit assessment.

Audit Rating: 3

See the response for C1.8. As the service is not yet active, there are no examples to test this process.

Evidence: Risk register (list of all patches available and risk documentation analysis); evidence of update processes (e.g., server update manager daemon); documentation related to the update installations.

C2. Appropriate technologies

C2.1 Repository has hardware technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluate when hardware technology changes are needed.

Audit Rating: 3

DS: During Phase 1 of the Digital Safe identified a distinct need in the community for a long-term, high-security storage space for sensitive content. See section "Phases 1&2" on page 3 of this document for more information on Phase 1, and section A3.1 for information on the identified key users. Digital Safe clients will rely on Arkivum and Archivematica for information on hardware updates and changes.

Arkivum: Arkivum and its applications can be used from most common operating systems (FAQ, 18), with technical support for installation available. Updates are automatically provided to clients via A-Stor as they develop and are applied by simple mouse click (FAQ, 11). "hardware appliances have been designed to offer industry standard fault resilience and come with a three year, four hour, on-site warranty. During this time, failed parts will be replaced under this warranty. Should the system undergo total failure or be destroyed (such as in a fire), a new system can be configured and archived data will still be available," (FAQ, 12) once the client has contacted Arkivum about the issues.

Archivematica: Archivematica's system "packages a customized Xubuntu environment as a virtual appliance, making it possible to run on top of any consumer-grade hardware and operating system" and only requires a single installation to run (Single install page). It is also possible to update Archivematica after installation. See their Installing from packages section for detailed information.

Evidence: Technology watch; documentation of procedures; designated community profiles; user needs evaluation; hardware inventory.

C2.2 Repository has software technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluate when software technology changes are needed.

Audit Rating: 3

DS: During Phase 1 of the Digital Safe identified a distinct need in the community for a long-term, high-security storage space for sensitive content. See section "Phases 1&2" on page 3 of this document for more information on Phase 1, and section A3.1 for information on the identified key users. Digital Safe clients will rely on Arkivum and Archivematica for information on software updates and changes. See section C2.1 as the procedures still apply to software changes.

Arkivum: See section C2.1 as the procedures still apply to software changes.

Archivematica: See section C2.1 as the procedures still apply to software changes.

Evidence: Technology watch; documentation of procedures; designated community profiles; user needs evaluation; software inventory.

C3. Security

C3.1 Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.

Audit Rating: 3

DS: During Phase 1 Digital Safe's project team interviewed various staff members in the colleges, IT Services, the Oxford Colleges Librarians Group, Oxford Archivists Consortium, and other related groups, committees, and departments. These interviews gathered some of the technical, security, and infrastructure requirements that these institutions might have, and chose the technologies accordingly. Digital Safe relies on Arkivum maintain and implement its security model.

Arkivum: Arkivum's procedures, hardware, and locations are all certified to ISO 27001 standards and are audited every six months. "[Arkivum's] secure storage locations are based in highly secure facilities, with our operations at all sites certified to ISO 27001 standards. Our locations are manned at all hours and access is strictly restricted to a list of named, trained and vetted members of the Arkivum Operations team. Each site is protected by best of breed firewall technology ensuring that our locations are protected from the latest advanced evasion techniques utilised by sophisticated hackers and intelligence organisations" (FAQ, 8).

Data is secured based on "the ability to separately encrypt each file stored in our service. Only encrypted data is ever stored in [Arkivum's] service. Each file is encrypted with a unique symmetric key using AES256 encryption. The symmetric key for the file is then encrypted using a public key from a public-private key pair using RSA2048 encryption. Both AES256 and RSA2048 are industry standard encryption algorithms and widely used in high security applications, e.g. electronic commerce and for sensitive government information," (FAQ, 9). For more information, see their FAQ documentation and their Security Model page. See section A1.2 for additional information.

Archivematica: Archivematica is a digital preservation workflow tool to which only the client has access. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. Archivematica is integrated with Arkivum via Arkivum's A-Stor, which will be built into the contract between the University and Arkivum, and is therefore protected by the security measures of Arkivum.

Evidence: ISO 17799 certification; documentation describing analysis and risk assessments undertaken and their outputs; logs from environmental recorders; confirmation of successful staff vetting.

C3.2 Repository has implemented controls to adequately address each of the defined security needs.

Audit Rating: 3

DS: Digital Safe chose Arkivum in part due to its security model and contingency plan. Digital Safe relies on Arkivum maintain and implement its security model.

Arkivum: Arkivum contains a strict chain-of-custody system, audit trails, and a highly detailed security model. "The security and audit model has been developed in partnership with Arkivum customers who have confirmed that the model meets their regulatory requirements as part of a due-diligence/audit process that they have conducted on Arkivum. This includes due-diligence by customers in clinical and financial sectors where regulation is strict," (Chain of Custody). Many Arkivum clients have strict compliance policies and government documents and require an even higher level of security. They also adhere to UK Government Information Levels; "IL2 and IL3 are UK Government Information Levels. These define the sensitivity of the data stored and the security procedures and processes that need to be followed when transmitting and storing this data. Arkivum currently has ISO27001 certified processes in place, which will meet the requirements set out for various Information Levels," (FAQ, 10). For more information, see Arkivum's FAQ document, their Chain of Custody page and their Security Model page.

Archivematica: See section C3.1 for the response regarding security needs.

Evidence: ISO 17799 certification; system control list; risk, threat, or control analyses; addition of controls based on ongoing risk detection and assessment.

C3.3 Repository staff have delineated roles, responsibilities, and authorizations related to implementing changes within the system.

Audit Rating: 2

DS: Digital Safe currently has a project team and their roles described in its Phase 3 PID. These individuals will design the service and have access to its private policies, but will have no access to any ingested content. Currently the project roles consist of: Project Sponsor, Senior User, Senior Supplier, Programme Manager, and Project Manager. The Steering Group will oversee the project board. The project board consists of the Senior Supplier, Programme Manager, Project Sponsors, and Senior Users. The Project Manager reports to the Project Board, and oversees the Project Roles. Project roles consist of IT Services, NSMS, and any other third parties (Phase 3 PID, 15-16).

Access to the dark archives held via Digital Safe is strict and has four different levels.

1. Clients, affiliated with the University of Oxford, who use the Digital Safe service (see section A3.1 for information on the identified key users for Digital Safe)
    a. Have control over the master encryption key
    b. Can view data, view metadata, manipulate data, delete data, export data, and retrieve escrow copies, alter digital preservation workflows, among other activities.
    c. Can determine access for other personnel and users
2. Users, perhaps researchers or auditors
    a. Can view data that is specified by the client
    b. Cannot manipulate, delete, copy, download, export, or otherwise retrieve data or access digital preservation workflows
3. Arkivum employees
    a. Can maintain servers, software, and hardware
    b. Cannot access data or digital preservation workflows
4. University of Oxford IT Services
    a. Can assist in determining digital preservation workflows
    b. Can assist in troubleshooting local error and issues
    c. Can maintain Digital Safe website
    d. Cannot access data or digital preservation workflows

Arkivum: Arkivum employees do not have access to any client data. Additionally, "there is no direct customer access to data in the Arkivum data centres, e.g. through a web interface or cloud API. This ensures all ingest and access is properly managed through Arkivum appliances," (Authentication and Access). For more information, see Arkivum's FAQ document, their Chain of Custody page and their Security Model page.

Archivematica: Archivematica employees do not have access to any client data. See section C3.1 for the response regarding additional security permissions.

Evidence: ISO 17799 certification; organizational chart; system authorization documentation.

C3.4 Repository has suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an off-site copy of the recovery plan(s).

Audit Rating: 4

DS: Digital Safe chose Arkivum in part due to its contingency plan and relies on Arkivum to maintain and implement this plan if necessary.

Arkivum: Arkivum provides a workflow and safety measures for integrity. For Arkivum/1+1, one copy is stored in a secure data center and one copy is saved on LTO data tape in Escrow. For Arkivum/100, two copies are stored in secure, geographically separate data centers and one copy is saved on LTO data tape in Escrow. The Escrow copy is the backup for any data loss or corruption. If the digital copies are corrupted or the clients need to remove content, the LTO tapes are delivered to the client. If there is complete data loss, Arkivum provides a "financial guarantee underwritten by an Information and Communication Technology Professional Liability Insurance Policy," which provides coverage for direct loss relating to data loss (FAQ, 7). Arkivum also provides multiple client sites, so if one site is compromised the data may be retrieved at another site (FAQ, 22). For more information:

- section A3.8 for more information on Arkivum/1+1 and Arkivum/100
- section C3.2 for information on Arkivum's physical storage locations
- Stages of Archiving
- Chain of Custody
- Security Model
- Arkivum's FAQ document

Archivematica: See section C3.1 for the response regarding security needs.

Evidence: ISO 17799 certification; disaster and recovery plans; information about and proof of at least one off-site copy of preserved information; service continuity plan; documentation linking roles with activities; local geological, geographical, or meteorological data or threat assessments.

Appendices

Appendix A: Commentary on Governance and Infrastructure

This section of policy building and planning requires the most attention during Phase 3. More information on what does exist is available in the Phase 3 PID and in the Informal TRAC Audit, and a brief description of the major improvement areas follows.

General Policy

As the project is still in development, the governance for the service is not yet in place. As Digital Safe is a service and not a repository itself there is no mission statement, and the project goals have altered through each phase. The primary Phase 3 goal is "to deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis,"[54] and any mission statement will likely draw from this idea. There is no additional policy formally produced that describes the current state and goals of the project outside of the Project Initiation Document. However, once Phase 3 commences the blog will be regularly updated and University web space will be developed for such policies. These will be further explored in the contract between the University and Arkivum.

Both of the technologies provide extensive documentation that is available on their websites.

Structure and Staffing

Currently there is a project group and Steering Committee. The hope is that the Steering Committee will continue, or develop into a similar governance committee[55], for the purpose of reviewing the contracts, funding, and any updates from IT Services and the technologies as needed. IT Services play a vital role in Digital Safe, as the ultimate goal is to develop and business and service model so that Digital Safe will be integrated with IT Services. This model will better determine staffing needs and training. IT Services will update the University web space for Digital Safe and assist in training, troubleshooting, and communicating with the Steering Committee. The Steering Committee will review and update the service as needed. Because the technologies are outsourced there is little technical training needed to maintain the interface, and the University can rely on Arkivum and Archivematica for maintaining their product.

Arkivum provides training virtually, on-site, and in workshops. The plan is to bring Arkivum in for training and to use and develop their training materials to do local training in the University. These materials will be broad, as it is up to the client to determine how much they want out of the service, and will be added to the Bodleian Library's current collection of training materials.[56] These materials will then be reviewed an updated on a 4-5 year cycle by the governance committee.[57] Both Arkivum and Archivematica have support services available during the week and on weekends for emergencies.

Documentation and Policy

It is recommended by this audit of Digital Safe to provide University web space to this project articulating the key users, basic policy, and contact information. The nature of a dark archive is not transparency and

[54] Digital Safe Phase 3 PID, page 2
[55] NJ interview
[56] Training materials
[57] NJ Interview

much information cannot be made public, however as a service provided to the University, general information and documentation is necessary.

Phase 2 determined that the key users for this service have been identified as: College Archivists; University Archive; Central Administrative Records Management; Departmental Research Records Management; and Bodleian Electronic Archives and Manuscripts.[58] These users have materials that require high-security and low-accessibility, including administrative records, student records, financial records, personal communication, medical reports, non-anonymized case studies, and other material that has personal, identifiable information. Users are not limited to only these categories, however, as the service is open to all who want to use the service and are affiliated with the University of Oxford.

As described above, policies have not yet been developed. The Steering Committee is currently directing the project, but much of the policy will be directed by the contract agreements between the University and Arkivum before they can be documented. Policies for the technologies, however, are well-detailed and located on their respective websites.[59] [60] It is also recommended that the Steering Committee review the policies, contracts, and funding security annually after this project is launched.

Legal permissions will largely be the responsibility of the client as much of the content will be produced by the client. Other materials that have been acquisitioned by the library may require additional policy measurements, but this is also primarily up to the library and BEAM to maintain. The contract between the University and Arkivum may need to determine if there is a need to build in a deposit agreement concerning the permissions for migration copies. It should be noted that Arkivum employees do not have access to any material as it is only accessible with an encryption key held by the client.

Outside of Digital Safe, the technologies have extensive documentation and are responsible for maintaining their certification and upholding their own policies. Arkivum has automated annual data integrity checks, five-year hardware and software migration, and developed an LTO road map to prevent technology and data obsolescence. More information is on their website. Arkivum also maintains ISO 27001 certification, is audited every six months, and welcomes client audits.

Financial Sustainability

Digital Safe does not have a full cost model developed for the long-term. Phase 3 has outlined the short-term business plan that includes University staff training, Arkivum training costs, start-up costs, 1 year of Arkivum service and storage space, and 1 year of maintenance fees. All are outlined in the Phase 3 PID. Developing this model is a priority for Phase 3 and will ultimately determine what funding is provided after the first year of service to maintain the website and any license and storage fees.

Once the service is launched and the clients choose the service they prefer, they will be responsible for paying. Phase 3 will help determine the payment method agreed upon between the University and Arkivum. If Arkivum handles the billing, they do direct invoicing for each client. If the University handles the billing it will be managed as a library service, similarly to the process for handling services like catalog use, access to electronic journals, and IT Services.

The Steering Committee will review and secure funding once it is determined if Phase 3 has been successful.

[58] Digital Safe Phase 3 PID, page 4
[59] Arkivum website
[60] Archivematica documentation

Contracts, Licenses, & Liabilities

Digital Safe and Arkivum have been contact since 2014 and are still in discussion over contract specifications. Once funding has been secured for Phase 3, this will be developed on a beta level for the first year of early adopters, and then re-evaluated after the first year is completed.

As discussed in section A5, deposit agreements and copyright issues will need to be evaluated to ensure legal inclusion of any acquisitioned material that might be included, such as from BEAM. A5 was not completed in this audit because there is no existing documentation yet. See section A5 for the review of this section.

Appendix B: Commentary on Digital Object Management

Object management is the most complete section of Digital Safe's development. Details on the practices and procedures described in Arkivum and Archivematica can be found in Section B of the Informal TRAC Audit and on the technologies' websites in their documentation.

Content Acquisition

All materials that are ingested into Arkivum will be provided by the client. Digital Safe will not be providing any content, only suggesting the type of content that might be ingested. Digital Safe also has recommendations on what metadata properties might be preserved, file format choices, and digital preservation workflow activities, but the content is strictly the responsibility of the client. Arkivum only requires a file to ingest, and depending on the complexity of the digital preservation workflow the client chooses, it can also ingest additional associated files.

Digital Preservation Workflows

Archivematica offers a customizable workflow tool to clients wanting to digitally preserve in addition to secure storage. Should the client require an OAIS model, Archivematica can create SIPs, normalize data, package AIPs, and normalize data, among other activities. A wide spectrum of processes is available, but ultimately up to the client to choose. Digital Safe may also be able to offer recommendations on best practices.

Storage

Arkivum offers multi-location, high-security storage within their own data centers and at an additional Escrow location. During the ingest process, the client can follow the process of ingest and is given a green light once the data is fully ingested and secure so that the client may delete their own copies. Arkivum also offers audit trails as a service. These must be turned on and specified by the client upon contract agreement. Audit trails will log every event affecting the content, such as an integrity check or migration, and the employee, time, reason, and any changes made. More information can be found on Arkivum's Audit Trails page[61].

Appendix C: Commentary on Technology, Technologies Infrastructure, and Security

This section has been largely handled by the technologies and is fairly compliant with TRAC guidelines. Once all three entities (Digital Safe, Arkivum, and Archivematica) have been integrated it will be easier to test the function of the entire system.

[61] Arkivum Audit Trails

Operating Systems

Arkivum and Archivematica were chosen because they built for institutions without the technical abilities or funding to develop their own infrastructure for digital preservation and storage, particularly with needs for high-security and low-access. Their systems can be operated from standard operating systems and do not require high levels of technical ability, fitting the needs of potential users identified in Phase 1.

Security

Although a mock interface has been created, the system combining Digital Safe, Arkivum, and Archivematica has not yet been completed. Once the service is deployed to early adopters, the security will be tested and become more concrete.

The technologies, infrastructure, and security for Digital Safe are largely provided by and well-documented by Arkivum. The service that the University will receive will vary based on client choices and any specific agreements made between the University and Arkivum, and by the client and Arkivum.

References

Archivematica. Program documentation. *Archivematica 1.5*. Vers. 1.5. Artefactual Systems, Inc., 2015. Web. <https://www.archivematica.org/en/docs/archivematica-1.5/>.

This documentation covers their digital preservation service.

Archivematica. Program documentation. *Archivematica Storage Service documentation*. Vers. 0.8. Artefactual Systems, Inc., 2015. Web. < https://www.archivematica.org/en/docs/storage-service-0.8/>.

This documentation covers their storage service documentation, which contains information on how to store within Arkivum.

Archivematica. Program documentation. *Format Policy Registry (FPR).* Vers. 1.1.0. Artefactual Systems, Inc., 2015. Web. < https://www.archivematica.org/en/docs/fpr/>.

This documentation covers their current Format Policy Registry, which is routinely updated.

"Arkivum." *Arkivum.* Arkivum Limited, 2016. Web. <http://arkivum.com/>.

Arkivum's documentation is spread across several pages that are all linked from this page. Individual pages are specifically included in the text, and can all be accessed here.

Arkivum. Program documentation. *Frequently Asked Questions – Arkivum*. Vers. 2.2. Arkivum Limited, 27 Feb. 2014. Web. <http://arkivum.com/wp-content/uploads/2014/04/Arkivum-Frequently-Asked-Questions.pdf>.

Arkivum also produced documentation in PDF format that is available in one comprehensive document.

"Digital Preservation Metrics." *Center for Research Libraries. Enriching Research. Expanding*

Possibilities. Since 1949. Center for Research Libraries, n.d. Web. 25 July 2016. <https://www.crl.edu/archiving-preservation/digital-archives/metrics>.

The Center for Research Libraries contains information on the metrics used to measure the trustworthiness of digital repositories. This information helped dictate what metric was used for this informal audit.

Jefferies, Neil. "Background on Digital Safe." Personal interview. Weston Library. 19 July 2016.

Neil Jefferies was informally interviewed by the creator of this document in order to answer specific questions on Arkivum's contract with the University of Oxford, the status of Phase 3, and other information that was not publicly available or known by other previous team members. This was not recorded and the notes are not publicly available. Contact Neil Jefferies for more information.

Jefferies, Neil, Brian Hicks, and Sam Rendell. *Digital Safe: Project Initiation Documentation*. Rep. Vol. 0.7. N.p. 26 Feb. 2016. Print.

The Phase 3 Project Initiation Documentation was the primary documentation for understanding Digital Safe as a service separate from the chosen technologies. It is not publicly available. Contact Neil Jefferies for more information.

Jefferies, Neil, and David Tomkins. *Digital Safe: Archiving Digital Records for the Long Term.* Bodleian Digital Library, 10 July 2014. Web. <https://digitalsafe.wordpress.com/2014/08/18/digital-safe-presented-at-the-ictf-conference-10-july-2014/>.

Neil Jefferies and David Tomkins presented Phase 1 and 2 of Digital Safe at the ICTF Conference in 2014. Information from this presentation were included in this informal audit, and images were utilized and cited in presentation related to this document.

Lawson, Sarah. *Electronic Archive Pilot Project Phase 2*. Rep. Vol. 1.0. N.p. 9 Feb. 2013. Print.

The Phase 2 Project Initiation Documentation was the primary documentation for understanding the history and trajectory of Digital Safe as it has developed. It is not publicly available.

McGovern, Nancy Y. *TRAC Review in Drupal*. MIT Libraries, 2013. Web. <http://ils.unc.edu/digccurr/curategear2013-talks/mcgovern-curategear2013.pdf>.

Nancy McGovern developed a tool for self-review using TRAC. Unfortunately, the original website is down or no longer exists, but it is hosted on Archivematica under Preservation Planning, and one of the original presentations for this tool is still available. This tool was not used, but the rating system for compliance was implemented in the informal audit.

Stanbridge, Nik. "New Digital Preservation Solution from Arkivum, Shaped to Grow with Your Data." *Arkivum*. Arkivum Limited, 21 Jan. 2016. Web. <http://arkivum.com/blog/perpetua-digital-preservation/>.

This article references the launch of the Arkivum and Archivematica integration, which is what Digital Safe will be using as their storage platform and digital preservation workflow.

Thomas, Susan. “Digital Safe and BEAM.” Personal interview. Weston Library. 19 July 2016.

Susan Thomas was informally interviewed by the creator of this document in order to answer specific questions on the current status of BEAM, to retrieve some background information on Digital Safe, and to review BEAM’s current needs from the Digital Safe service. This was not recorded and the notes are not publicly available.

Tomkins, David. "An Electronic Archive Pilot Project at the University of Oxford." Web blog post. *Digital Safe*. Wordpress, 2013. Web. <https://digitalsafe.wordpress.com>.

David Tomkins was the project manager for Phase 2 of Digital Safe and developed a Wordpress blog to publicly track the progress of Digital Safe. This provided several meeting notes and presentations with information on Phases 1 and 2 of Digital Safe. Though Phase 3 is not yet included on the blog, Neil Jefferies confirmed that once Phase 3 begins the blog will resume providing updates.

Trusted Repository Archiving Checklist (TRAC). Documentation. Vers. 1.0. The Center for Research Libraries, 2007. Web. <https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf>.

The official criteria and checklist for TRAC is the foundation of this audit, and the criteria are copied directly from this document.

Various EAPP documents provided by David Tomkins.

As the project manager for Phase 2, David Tomkins provided his documents from Phase 2 for establishing a context for that phase. These documents are not publicly available.

Appendix C: DSA Assessment of Digital Safe – Full Version

This is the full version of the results from comparing Digital Safe to DSA Guidelines. These responses include a more complete description of how Digital Safe strategies fit in with DSA Guidelines and act as a reference to any points made in the Discussion section.

0. Context

Repository type: Institutional repository, (Other: Dark archive)

Brief Description of the Repository's Designated Community:

The key users for this service have been identified in the Electronic Archives Pilot Project (renamed Digital Safe) Phase 1 as: College Archivists; University Archive; Central Administrative Records Management; Departmental Research Records Management; and Bodleian Electronic Archives and Manuscripts. This community is part of the larger University of Oxford community. These users have materials that require high-security and low-accessibility, including administrative records, student records, financial records, personal communication, medical reports, non-anonymized case studies, and other material that has personal, identifiable information. These users were identified after interviewing various colleges and departments on campus and determining a need for a universal storage system (Jefferies, Hicks, & Rendell, 2016) and are internally documented with letters of support from various colleges. Users are not limited to only these categories, however, as the service is open to all who want to use the service and are affiliated with the University of Oxford.

Level of Curation Performed

- ***E. Content distributed as deposited***
- ***F. B. Basic curation – e.g., brief checking, addition of basic metadata or documentation***
- ***G. Enhanced curation – e.g., conversion to new formats, enhancement of documentation***
- ***H. Data-level curation – as in C above, but with additional editing of deposited data for accuracy***

Digital Safe as a service is being designed to provide a wide range of digital preservation activities that suit the needs of the Designated Communities. The lowest possible level of curation ingested by Digital Safe will fall under "A. Content distributed as deposited." As discussed in Requirement 7. Data integrity and authenticity and in Requirement 8. Appraisal, the storage platform Arkivum can ingest any file and will disseminate the file in the exact same condition it was in upon ingestion. Additional digital preservation workflows are customized by the client, who will also determine the relevance and authenticity of any content they opt to store in Digital Safe.

Outsource Partners

Arkivum is a proposed contractual partner that will provide the means of digital storage in the form of ingest pipes for users of the Digital Safe service. According to their website, "Our operations at all sites, including our business offices, is certified to ISO 27001 information security standards" (Arkivum Ltd., 2014). Arkivum is also audited every six months and welcomes client audits (Arkivum Ltd., 2014). In regards to service agreements between Arkivum and clients, there will be a basic contract between the University of Oxford and Arkivum, and individual client preferences will build upon that contract and articulate the clients' product choice, workflow preferences, and additional storage space options (Jefferies, N., Personal communication, 2016, July 19).

Archivematica is a proposed contractual partner that is built into Arkivum and will provide the digital preservation workflows for users of the Digital Safe service. Ideally the University of Oxford will be able to purchase one license for Archivematica via their contract with Arkivum, rather than each individual client purchasing a license (Jefferies, N., Personal communication, 2016, July 19), though these discussion with Arkivum are ongoing.

1. Mission/Scope

Statement of Compliance: 2 The scope has been determined by the designated community but the mission statement is still in development as the project evolves. The chosen technologies have fully developed and implemented mission statements.

Self-assessment statement:

The original Electronic Archive Pilot Project's mission statement is as follows: "The Electronic Archive Pilot Project will establish the feasibility of a working electronic archive for the use of the whole of the Collegiate University. The archive will support the safe and secure storage of all classifications of non-public record data that individual departments, colleges and associated units are required to keep legally or would like to keep for historic reasons. The pilot project aims to develop a cost recovered service" (Wilson, J., 2012). The Digital Safe service itself is developing a mission statement. It will likely draw on the phrase from the Phase 3 Project Initiation Document to "deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis once in production." This document is not publicly available. Both Arkivum and Archivematica have additional mission statements found in Arkivum's "About Us"[62] section and Archivematica's "What is Archivematica"[63] page.

2. Licenses

Statement of Compliance: 1 This section is designed in theory and implementation has begun, and is the focus of the next project phase.

Self-assessment statement:

This entire section relies on a fully developed contract with Arkivum. A priority of Phase 3 of Digital Safe is finalizing the contract with Arkivum and in determining any preservation rights and copyrights. In regards to transferring control of data from a client to Arkivum, any services that are part of the library may come under the Heritage Institution exception for the right to change objects and make copies, which would occur in the regular migration of data in Arkivum and in any Archivematica workflow the data is pushed through. Though much of the material may be produced by the clients at the University (e.g. financial records, etc.), some of BEAM's content may apply to the exception (Jefferies, N., Personal communication, 2016, July 19).

3. Continuity of access

Statement of Compliance: 2 This is a theoretical concept that will be developed should Digital Safe regain funding. As an Outsource Partner, Arkivum will be responsible for maintaining continuity of access as per their mission statement.

[62] Arkivum "About US"" http://arkivum.com/about-us/

[63] "What is Archivematica:" https://www.archivematica.org/en/docs/archivematica-1.5/user-manual/overview/intro/#intro

Self-assessment statement:

As any future contract with Arkivum will dictate, Digital Safe will rely on the technologies to remain updated on and implement any evolving best practices in the field. Arkivum has policies and automated processes in place for running integrity checks on the data, software, and hardware. The data is retrieved annually and given an integrity test based on checksums. Arkivum has also identified software and hardware obsolescence to occur on a cycle of generally 3 to 5 years, so Arkivum's policy is that data is migrated to new media following the LTO roadmap. The LTO data tapes in Escrow are also migrated every 5 years. More detailed information on media upgrades can be found on their Data Integrity page. In addition to being committed to maintaining standards-based tools for those interested in digital preservation tools, Archivematica also relies on their community to help steer the tool in the most useful direction.

Digital Safe is a service and therefore their only responsibility is to maintain the service and contracts with technologies, and has no influence on the amount of time the data is held. Ultimately those using the Digital Safe service are responsible for maintaining their encryption key, restricting or releasing access to materials, and providing their own funding to ensure their space and digital preservation workflows are maintained.

There is no formal documentation on policies in place for changes in circumstances from Digital Safe. Ideally the Steering Committee will develop these policies in Phase 3 and eventually transfer responsibility to technical support, and this will also likely be dictated in the responsibilities of Digital Safe in their contract with Arkivum.

4. Confidentiality/Ethics

Statement of Compliance: 2 As per the goal of Digital Safe to "deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis once in production," Digital Safe will store confidential information and will need to consider their documentation carefully.

Self-assessment statement:

Digital Safe will need to be a dark archive because its purpose is to store confidential information. First, Digital Safe has four different levels of access described in their Phase 3 Project Initiation Document that also corresponds to features offered by Arkivum.

1. Clients, affiliated with the University of Oxford, who use the Digital Safe service (see section 0. Context for information on specific designated communities):
    a. Have control over the master encryption key
    b. Can view data, view metadata, manipulate data, delete data, export data, and retrieve escrow copies, alter digital preservation workflows, among other activities.
    c. Can determine access for other personnel and users
2. Users, perhaps researchers or auditors
    a. Can view data that is specified by the client
    b. Cannot manipulate, delete, copy, download, export, or otherwise retrieve data or access digital preservation workflows
3. Arkivum employees
    a. Can maintain servers, software, and hardware
    b. Cannot access data or digital preservation workflows
4. University of Oxford IT Services

a. Can assist in determining digital preservation workflows
b. Can assist in troubleshooting local error and issues
c. Can maintain Digital Safe website
d. Cannot access data or digital preservation workflows

Arkivum employees do not have access to any client data. Additionally, "there is no direct customer access to data in the Arkivum data centres, e.g. through a web interface or cloud API. This ensures all ingest and access is properly managed through Arkivum appliances" (Arkivum Ltd., 2014) Arkivum is build on ingest pipes that are matched with an encryption key that is unique to each client. Without the encryption key there is no access to any of the data in its original form or a copy. The encryption key is controlled solely by the client. Arkivum does not have public documentation on measures in place if disclosure of data occurs. Arkivum does, however, have a detailed contingency plan described in Requirement 16 (Arkivum Ltd.). Many Arkivum clients have strict compliance policies and government documents and require an even higher level of security. Arkivum adheres to UK Government Information Levels; "IL2 and IL3 are UK Government Information Levels. These define the sensitivity of the data stored and the security procedures and processes that need to be followed when transmitting and storing this data" (Arkivum Ltd., 2014).

Archivematica is a digital preservation workflow tool to which only the client has access. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. Archivematica is integrated with Arkivum via Arkivum's A-Stor, which will be built into the contract between the University and Arkivum, and is therefore protected by the security measures of Arkivum.

5. Organizational infrastructure

Statement of Compliance: 2 As Digital Safe project team members investigate contracts with Arkivum they are also in the process of developing a stable organization infrastructure.

Self-assessment statement:

Digital Safe is currently in the process of acquiring funding for Phase 3. There is a cost model and projected expense report developed. The projected expense report includes: IT Services internal staff, Non-IT Services staff, Hardware, software, training and equipment/storage (Arkivum), and included a .5% charge for Prime Minister's Office charge, an 85% charge for contingency per the Monte Carlo Simulation, and a forecast on on-going charges per year (Jefferies, et al., 2016).

The project team for Phase 3 of Digital Safe is established, but the ultimate short and long-term staffing duties will rely on the contract with Arkivum and collaboration with the IT department. The plan is that the service will be built into the University of Oxford's IT department. Further, a business and service model is a priority for Phase 3 of Digital Safe and will determine staffing needs and training. Ideally the Steering Committee on the Digital Safe project will also develop into a governance committee that will act as a checks and balances to ensure that local management is useful, that will review policy annually, and to ensure funding (Jefferies, N., Personal communication, 2016, July 19).

Digital Safe will on the technologies to maintain their own staffing. According to Arkivum, "In addition to technical change in the archive system, managing staff transitions of those who run the system, for

example support staff and administrators, is required"[64] Archivematica is created and staffed by Artefactual Inc.

The official governance for this service has not yet been developed. Ideally the Steering Committee on the Digital Safe project will also develop into a governance committee that will work in tandem with the staff in the IT department to review and update any general policies, training materials, and announcements and event information. Because the client will ultimately choose how much space to purchase and what data is ingested, it is their responsibility to respond to any announcements or updates in training or standards.

Digital Safe will update any announcements and training material for the service, and the local governance committee will have little public documentation aside from the aspects described above that will need regularly updated due to the sensitive information and privacy of the dark archives Digital Safe is providing. Digital Safe relies on Arkivum and Archivematica to handle obsolescence, migration, data integrity, and generating any new training materials (Jefferies, N., Personal communication, 2016, July 19).

Arkivum does not publish their financial reports, business plans, or blank contracts. However, Arkivum has several well-known institutions as clients, including the Museum of Modern Art, University of Westminster, and the Oxford Molecular Diagnostics Centre, among many others noted in case studies on their website that have been successful. Their Solutions tab offers several reports and case studies in various fields, such as Higher Education, that are evidence of success. Archivematica: Archivematica does not publish their financial reports or business plans. Archivematica is an open-source tool and therefore does not charge clients. However, they provide paid services, including storage, training, technical support, all noted on their Services[65] page. In the long-term, they have had several successful clients list on their Clients[66] page.

6. Expert guidance

Statement of Compliance: 3 The decision to outsource the technologies for Digital Safe indicates that the project team wants the expertise of established digital preservation organizations, with which contract development is ongoing.

Self-assessment statement:

Digital Safe is the result of feedback collected during Phase 1 and will be built based on designated community needs. In the future, the Digital Safe service will be built into the University of Oxford's IT department. Ideally the webpage would have contact information for the governance committee or other local managing team who can assist in troubleshooting smaller issues and directing to training and Arkivum and Archivematica help pages (NJ Interview). As previously noted, once the Steering Committee transfers Digital Safe control over to IT, IT will then be responsible for keeping training materials updated and notifying clients. Digital Safe relies on Arkivum and Archivematica to maintain

Arkivum also seeks feedback from their users. According to their policy, "The security and audit model above has been developed in partnership with Arkivum customers who have confirmed that the model meets their regulatory requirements as part of a due-diligence/audit process that they have conducted on Arkivum. This includes due-diligence by customers in clinical and financial sectors where regulation is

[64] Arkivum "Data Integrity:" http://arkivum.com/data_integrity/
[65] Archivematica Service:s https://www.artefactual.com/services/
[66] Artefactual Systems' Clients: https://www.artefactual.com/clients/

strict."[67] Expertise sources for Arkivum are not in public documentation, but their transparent and detailed descriptions of their preservation workflows and storage methods, additional case studies[68] on their website, and list of current clients offered under the Industries tab on their website offer community support and proof of successful methods.

Archivematica was originally a project use case for OAIS to "process analysis to synthesize the specific, concrete steps that must be carried out to comply with the OAIS functional model from Ingest to Access." This project expanded beyond OAIS into its current state as an open-source digital preservation workflow tool based on user feedback (Artefactual Inc.). Clients do have to navigate to the manufacturer page to contact Artefactual for assistance.

7. Data integrity and authenticity

Statement of Compliance: 3 Arkivum and Archivematica both maintain detailed documentation, which will be absorbed into Digital Safe documentation by the contract between Arkivum and Digital Safe.

Self-assessment statement:

Digital Safe is a service. Authentication of the original data. will be the entire responsibility of the client. The content will likely comprise of files created by the client, e.g. student records, financial records, etc., and legal permissions are moot. Other material may be acquisitions to the Bodleian Library that have their own documentation and standards that are separate from this service. Each client may also have their own policies on permissions and permissions workflows that are independent of each other.

In regards to service agreements between Arkivum and clients, there will be a contract between individual clients and Arkivum that articulate the clients' product choice, workflow preferences, and additional storage space options. Furthermore, according to Arkivum "Our operations at all sites, including our business offices, is certified to ISO 27001 information security standards" (Arkivum Ltd., 2014), and Arkivum is regularly audited externally to maintain ISO 27001 certification. These certifications enable Arkivum to legally hold ingested content. Arkivum offers audit trails as a service. These must be turned on and specified by the client upon contract agreement. Audit trails are "accessible in machine-readable XML format through Arkivum REST API calls, either on a per-file or per-folder basis. Audit trails are also accessible in human readable PDF/A format through the Arkivum Service web interface or through an API call. Audit trails in PDF/A format can be signed if necessary to show that they were generated by Arkivum Service" (Artefactual Systems Inc.).

Archivematica is a digital preservation workflow tool and not a storage space, and has no responsibility regarding the content authentication. Archivematica also utilizes fixity checks before AIP storage. "Archivematica generates checksums upon transfer of objects into the system, and will verify those checksums before storing the AIP. It is also possible to include pre-existing checksums, which Archivematica will also verify. To check fixity of AIPs in storage, Artefactual has written a separate command-line app called Fixity (Artefactual Systems Inc.).

8. Appraisal

[67] Arkivum Chain of Custody: http://arkivum.com/chain-custody-audit-trails/
[68] Arkivum Case Studies: http://arkivum.com/resources/#

Statement of Compliance: 3 Digital Safe as a service is planned to offer both basic and extensive digital preservation activities in addition to storage, though the specific workflows are the responsibility of the client.

Self-assessment statement:

The Designated Communities that was identified after interviewing various colleges and departments on campus determined a need for a universal storage system (Jefferies, et al., 2016). The service is open to all who want to use the service and are affiliated with the University of Oxford. Therefore, the Designated Communities determine what information is included and will have their own documentation dictating the appropriate data. Digital Safe, Arkivum, and Archivematica have no influence over what data is considered appropriate by the Designated Communities.

Arkivum can ingest any file format, but does not record their representation information. If the client chooses to utilize Archivematica, Archivematica can determine file formats and normalize using their Format Policy Registry, as well as creating SIPs and AIPs for record this process.

9. Documented storage procedures

Statement of Compliance: 3 Arkivum's long term-preservation strategies are publicly documented and will be absorbed into Digital Safe documentation upon the completion of a contract.

Self-assessment statement:

Digital Safe will rely on the technologies for storage and documented storage procedures as is outlined in their future contract.

Arkivum's documentation is publicly available on their website and specific client preferences are dictated in the final contract. Each file is encrypted with a unique symmetric key using AES256 encryption. The symmetric key for the file is then encrypted using a public key from a public-private key pair using RSA2048 encryption. Both AES256 and RSA2048 are industry standard encryption algorithms and widely used in high security applications, e.g. electronic commerce and for sensitive government information" (Arkivum Ltd., 2014). The encryption key is controlled solely by the client. Many Arkivum clients have strict compliance policies and government documents and require an even higher level of security. Arkivum adheres to UK Government Information Levels by maintaining ISO 27001 certification and can maintain IL2 and IL3 UK Government Information Levels. "These define the sensitivity of the data stored and the security procedures and processes that need to be followed when transmitting and storing this data" (Arkivum Ltd., 2014).

Arkivum has policies and automated processes in place for running integrity checks on the data, software, and hardware. The data is retrieved annually and given an integrity test based on checksums. Arkivum has also identified software and hardware obsolescence to occur on a cycle of generally 3 to 5 years, so Arkivum's policy is that data is migrated to new media following the LTO roadmap. The LTO data tapes in Escrow are also migrated every 5 years.[69]

Arkivum maintains at least two copies of the data, one in a secure data center and one on LTO data tape held in Escrow. The Escrow copy is the backup for any data loss or corruption. If the digital copies are corrupted or the clients need to remove content, the LTO tapes are delivered to the client. If there is complete data loss, Arkivum provides a "financial guarantee underwritten by an Information and

[69] Arkivum Data Integrity: http://arkivum.com/data_integrity/

Communication Technology Professional Liability Insurance Policy," which provides coverage for direct loss relating to data loss (Arkivum Ltd., 2014). Arkivum also provides multiple client sites, so if one site is compromised the data may be retrieved at another site (Arkivum Ltd., 2014).

Archivematica is the digital preservation workflow that occurs before ingest into Arkivum storage and therefore does not have data back-up. Clients are notified of any failed actions and are responsible for managing these issues. Archivematica also utilizes fixity checks before AIP storage. "Archivematica generates checksums upon transfer of objects into the system, and will verify those checksums before storing the AIP. It is also possible to include pre-existing checksums, which Archivematica will also verify (Artefactual Systems Inc.).

10. Preservation plan

Statement of Compliance: 2 Arkivum's long term-preservation strategies are publicly documented and Archivematica's options are publicly documented, and individuals contracts between clients and Arkivum will dictate specific preservation plans.

Self-assessment statement:

The preservation plan for each user of the service Digital Safe will be unique to their context. They will specify their preferences for storage and transfer custody to Arkivum in the contract with Arkivum, and they will design their own digital preservation workflow in Archivematica based on recommendations from both Digital Safe and Archivematica. They will determine their own preservation level and length of time the data is to be held, and communicate with Arkivum directly. Arkivum follows a strict chain of custody that allows for minimal contact with client data. According to their chain of custody, "The starting point for data authenticity is the chain of custody established with the customer to ensure data has been correctly copied into the service. Once ingested, files become read only and cannot be updated or overwritten. Deletion of files is through a strictly controlled process that requires a request to be made to Arkivum. The default is that once a file is in the service then it remains in the service and does not change when it is within the service."[70] Any access to ingested data is restricted to individuals with the encryption key.

Digital preservation activities are solely the responsibility of the client, who will design them using Archivematica. Archivematica allows users to do various archival activities, including adding metadata in Dublin Core, adding rights in PREMIS, data normalization, AIP storage, DIP storage, communication with other tools (e.g. Archivist's Toolkit, ArchiveSpace, Arkivum), among other options. A SIP begins as a transfer. "In Archivematica, Transfer is the process of transforming any set of digital objects and/or directories into a SIP. Transformation may include appraisal, arrangement, description and identification of donor restricted, private or confidential contents."[71] A transfer can be created with submission documentation, existing checksums, or an existing METS structmap. The transfer will be processed through several micro-services, as described in the Transfer process. This is then ingested into Archivematica after the green light is given to the client. The completion of all processes in Archivematica are indicated by color and text. Green indicates that a process has been completed successfully, and red indicates that the process has not been completed successfully. A client can search for content when by its name. Archivematica's naming system will retain the original name of the transfer

[70] Arkivum Chain of Custody: http://arkivum.com/chain-custody-audit-trails/

[71] Archivematica Transfer: https://www.archivematica.org/en/docs/archivematica-1.5/user-manual/transfer/transfer/#transfer-checksums

unless a new name has been assigned to the SIP upon creation. This name will be combined with a Universal Unique Identifier that is generated and assigned during SIP formation (AIP Structure).

11. Data quality

Statement of Compliance: 2 Arkivum and Archivematica maintain best practices, and Digital Safe intends to provide recommendations, but it will be the responsibility of the client to determine what level of quality their data maintains.

Self-assessment statement:

The metadata fields that Digital Safe recommends the clients' preserve are: Title, Description, Creator(s), ID Type, ID Value, Retention review date, Retention rationale, Resource type, Technical description(s), Covering date(s), Finding aid(s), Rights & Licensing information (Jefferies & Tomkins, 2014). It is ultimately the client's decision what properties are preserved.

Arkivum only requires a file for anything to be ingested. Arkivum has no responsibility for anything additional included in the ingest and relies on the client to upload any additional information. Arkivum can ingest any file format and will maintain a copy of the original file alongside a normalized file.

12. Workflows

Statement of Compliance: 1 While extensive documentation for Arkivum is well-established, the integration into Digital Safe documentation is still in progress and are a major focus for Phase 3.

Self-assessment statement:

The ongoing nature of Digital Safe means that documentation of processes is also developing. The current documentation consists primarily of: meetings minutes from the Steering Committee and the Oxford Colleges Librarians Group; Project Initiation Documents; letters of support from various Oxford Colleges; Programme and Project Highlight Reports; conference presentation materials; Project Request Forms; Request for Change forms; End Project Reports; and various newsletters and copies of email communications. Project member Neil Jefferies has began developing a contract with Arkivum, though this is not yet available (Jeferies, N. Personal communication, 2016, July 19). Contracts with Arkivum will also help Digital Safe to establish documentation on deposits, security, and best practices for digital preservation workflow.

Digital Safe has also established a Designated Community and a corresponding access matrix, seen in detail in Requirement 4. Confidentiality/Ethics, which will guide documentation evolution.

Arkivum and Archivematica both maintain extensive documentation that is publicly available. See the response to Requirement 9. Documented storage procedures for more specific information.

13. Data discovery and identification

Statement of Compliance: 2 The Designated Communities will only have access to their own data, for which they will have provided the identifier that will be maintained by Arkivum.

Self-assessment statement:

Arkivum is designed to hold AIPs. "The persistent file/folder mechanism within the Arkivum service offers excellent support for the storage of AIPs. The file will retain its original filename, and checksums are provided for incorporation in the Preservation Description Information (PDI) for the AIP. The service

follows the OAIS model for Archive Storage through use of replication, fixity monitoring and repair, disaster recovery, migration, and tiered storage to deliver a specified level of performance, availability and integrity of storage."[72] If the client decides to utilize the Archivematica tool they may choose to create and package an AIP. The client must determine the processing configuration, which can be left at Archivematica's default setting, or can be created by the client.

After configuring the process as desired, the SIP can be normalized and stored in an AIP. "After normalization is approved, the SIP runs through a number of micro-services, including processing of the submission documentation, generation of the METS file, indexing, generation of the DIP and packaging of the AIP," which is packaged according to Bagit specifications (Artefactual Systems Inc.). AIP reingest is also an option if the client wishes to add information (e.g. metadata and data normalization) after the SIP process.

14. Data reuse

Statement of Compliance: 1 While documentation for Arkivum is well-established, the projected Digital Safe Designated Communities complicate the process of establishing licenses with vendors.

Self-assessment statement:

Digital Safe is designed to hold confidential data with high access restrictions. The Digital Safe Steering Committee has opted to utilize Arkivum as the storage technology and will rely on Arkivum to implement their responsibilities in data reuse that is outlined in their contract.

Ideally, Digital Safe will offer best practices based on recommendations from their Outsource Partners and other experts in the BDLSS based on Designated Community needs. Based on feedback from the Designated Communities, the metadata fields that Digital Safe recommends the clients' preserve are: Title, Description, Creator(s), ID Type, ID Value, Retention review date, Retention rationale, Resource type, Technical description(s), Covering date(s), Finding aid(s), Rights & Licensing information (Jefferies & Tomkins, 2014). It is ultimately the client's decision what properties are preserved.

Arkivum's full mission statement outlines its service: "Arkivum provides industry-leading big data preservation and archiving solutions to organisations in higher education, healthcare, life sciences, and digital heritage. These solutions assure the long-term value, trustworthiness and authenticity of data irrespective of whether it's terabytes or petabytes being preserved, and irrespective of whether the retention period is years, decades, or a quarter of a century. Through active data management, chain of custody and ISO 27001 compliance processes, Arkivum's unique technology provides rapid, low-latency access to archived data and provides an unrivalled 100% data integrity guarantee. Backed by indemnity insurance, this is our commitment to protect, curate and preserve data for the future and to eliminate the needless loss of information and knowledge. Arkivum works with partners to deliver integrated, scalable and flexible solutions for data discovery and sharing; publishing; file format preservation; and information portals" (Arkivum Ltd., 2015). Specifically, Arkivum only requires a file for anything to be ingested. Arkivum has no responsibility for anything additional included in the ingest and relies on the client to upload any additional information. Arkivum can ingest any file format and will maintain a copy of the original file alongside a normalized file.

[72] Arkivum Integration with other systems: http://arkivum.com/integration-with-other-systems/

Digital preservation activities are solely the responsibility of the client, who will design them using Archivematica. Archivematica allows users to do various archival activities, including adding metadata in Dublin Core, which is also ingestible by Arkivum.

15. Technical infrastructure

Statement of Compliance: 2 Outsourcing the technologies allows Digital Safe to customize their infrastructure based on pre-existing infrastructure, rather than building their own, which is being developed between the project team and Arkivum.

Self-assessment statement:

During Phase 2 of the Digital Safe project, the team investigated service and infrastructure models that included BEAM and ORA-Data systems at the University of Oxford, and also investigated DataBank as an Outsource Partner. BEAM infrastructure, developed in 2005, is held on a stand-alone server that has recently not been able to keep up with the increase in acquisitions and the level of organization and security that the BEAM would prefer (Thomas, S. Personal communication, 2016, July 19). ORA-Data was found incompatible for the type and amount of security measure that would need to be implemented. Building an entirely new infrastructure was also investigated but would not have been time or cost-efficient. DataBank was also ruled out for not including all the aspects necessary for the project in one platform, requiring additional outsourcing or increased time and money for the BDLSS. Additionally, given that the Designated Communities span across a range of departments, colleges, and expertise, the interface design will need to be user-friendly and technical support is necessary. This investigation lead the project team to opt for outsourcing storage to Arkivum, outsourcing optional digital preservation activities to Archivematica as a built-in tool in Arkivum, and create a local interface to be maintained long-term by the University IT Services. The process of considering multiple options and choosing Arkivum is evidence that Digital Safe is maintaining its mission statement to deliver a secure, long-term records archiving service for the University and Colleges operating on a cost recovery basis once in production" even in the development stages. This does not fulfill this infrastructure, but it establishes a record of effort to build stable technical infrastructure.

Arkivum adheres to UK Government Information Levels by maintaining ISO 27001 "Information security standards" certification and can maintain IL2 and IL3 UK Government Information Levels. Should Digital Safe be built on a contract with Arkivum, it would allow Digital Safe to absorb their certification and partially fulfill this Requirement. Arkivum and its applications is also constructed to be used from most common operating systems (Arkivum Ltd., 2014). Additional technical support is available for unique operating systems. Updates are automatically provided to clients as they develop and are applied by simple mouse click (Arkivum Ltd., 2014).

16. Security

Statement of Compliance: 3 The decision to outsource to Arkivum is heavily influenced by the security levels maintained by Arkivum, which will be absorbed into Digital Safe documentation upon the completion of a contract.

Self-assessment statement:

If Digital Safe were to establish a contract with Arkivum, security would largely be the responsibility of Arkivum. According to project member Neil Jefferies, a primary reason Arkivum is the choice for building Digital Safe is their contingency plan (Personal communication, 2016 July 19). First, he data is retrieved annually and given an integrity test based on checksums and preventative data migrations where

“each copy has its integrity actively monitored and managed and any corruption or loss is automatically repaired to make the system self-healing” (Arkivum Ltd.). Arkivum follows the LTO roadmap for storage obsolescence and aims to prevent data loss by determining what LTO generation their system is and introducing new generations well in advance to any system failure, and plan to migrate every 3-5 years. Escrow copies are also migrated every five years to prevent data loss.

Arkivum’s procedures, hardware, and locations are all certified to ISO 27001 standards and are audited every six months. “[Arkivum’s] secure storage locations are based in highly secure facilities, with our operations at all sites certified to ISO 27001 standards. Our locations are manned at all hours and access is strictly restricted to a list of named, trained and vetted members of the Arkivum Operations team. Each site is protected by best of breed firewall technology ensuring that our locations are protected from the latest advanced evasion techniques utilised by sophisticated hackers and intelligence organisations” (Arkivum Ltd., 2014). Data is secured based on “the ability to separately encrypt each file stored in our service. Only encrypted data is ever stored in [Arkivum’s] service. Each file is encrypted with a unique symmetric key using AES256 encryption. The symmetric key for the file is then encrypted using a public key from a public-private key pair using RSA2048 encryption. Both AES256 and RSA2048 are industry standard encryption algorithms and widely used in high security applications, e.g. electronic commerce and for sensitive government information,” (Arkivum Ltd., 2014).

Arkivum contains a strict chain-of-custody system, audit trails, and a highly detailed security model. “The security and audit model has been developed in partnership with Arkivum customers who have confirmed that the model meets their regulatory requirements as part of a due-diligence/audit process that they have conducted on Arkivum. This includes due-diligence by customers in clinical and financial sectors where regulation is strict” (Arkivum Ltd., 2014).

Arkivum employees do not have access to any client data. Additionally, “there is no direct customer access to data in the Arkivum data centres, e.g. through a web interface or cloud API. This ensures all ingest and access is properly managed through Arkivum appliances” (Arkivum Ltd., 2014) which is reliant on client control of the encryption keys.

Finally, Arkivum provides a workflow and safety measures for integrity. For Arkivum/1+1, one copy is stored in a secure data center and one copy is saved on LTO data tape in Escrow. For Arkivum/100, two copies are stored in secure, geographically separate data centers and one copy is saved on LTO data tape in Escrow. The Escrow copy is the backup for any data loss or corruption. If the digital copies are corrupted or the clients need to remove content, the LTO tapes are delivered to the client. If there is complete data loss, Arkivum provides a “financial guarantee underwritten by an Information and Communication Technology Professional Liability Insurance Policy,” which provides coverage for direct loss relating to data loss (Arkivum Ltd., 2014). Arkivum also provides multiple client sites, so if one site is compromised the data may be retrieved at another site (Arkivum Ltd., 2014).

Archivematica is a digital preservation workflow tool to which only the client has access. For Digital Safe, content is not stored in Archivematica for longer than it takes the workflow process to complete before it is stored in Arkivum. Archivematica is integrated with Arkivum via Arkivum’s A-Stor, which will be built into the contract between the University and Arkivum, and is therefore protected by the security measures of Arkivum.