# DELIVERABLE

**Project Acronym:** DCH-RP

**Grant Agreement number:** 312274

**Project Title:** Digital Cultural Heritage Roadmap for Preservation - Open Science Infrastructure for DCH in 2020

---

## Deliverable D4.1 Trust Building Report

**Revision: 1.0 FINAL**

---


**Authors:**

Raivo Ruusalepp (EVKM)
Börje Justrell (Riksarkivet)
Licia Florio (TERENA)

**Contributors:**

Eva Montenegro Piñeiro (EVKM)

**Reviewers:**

Tim Devenport (EDItEUR)
Claire Loucopoulos (DEDALE)
Norbert Meyer (PSNC)


| Project co-funded by the European Commission within the ICT Policy Support Programme | | |
|---|---|---|
| Dissemination Level | | |
| P | Public | P |
| C | Confidential, only for members of the consortium and the Commission Services | |

# Revision History

| Revision | Date | Author | Organisation | Description |
|---|---|---|---|---|
| 0.1 | 11/6/2013 | Raivo Ruusalepp | EVKM | Initial structure |
| 0.2 | 23/9/2013 | Raivo Ruusalepp | EVKM | Updated structure |
| 0.3 | 25/10/2013 | Raivo Ruusalepp | EVKM | Draft outline |
| | | Borje Justrell | RA | Updated outline |
| 0.5 | 28/01/14 | Raivo Ruusalepp | EVKM | First draft of introduction |
| 0.6 | 26/02/14 | Raivo Ruusalepp | EVKM | First draft of trust models |
| 0.7 | 04/03/14 | Raivo Ruusalepp, Eva Montenegro | EVKM | First draft of risk profile |
| 0.8 | 31/03/14 | Borje Justrell | RA | New Executive summary and a new section 1, integrating sections 7 – 10 and references |
| 0.9 | 07/04/14 | Raivo Ruusalepp | EVKM | Final Executive summary and correction of spelling errors |
| 1.0 | 08/04/14 | Claudio Prandoni Antonella Fresa | PROMOTER | Final check |

**Statement of originality:**

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

# TABLE OF CONTENTS

# LIST OF FIGURES

# EXECUTIVE SUMMARY

The DCH-RP project is developing a roadmap for preservation services where cultural heritage institutions rely on co-operation with e-Infrastructures. This necessitates the analysis of trust as one of the key concepts that this collaboration rests on, and that cultural heritage institutions need to be transparent about towards their stakeholders.

This deliverable outlines the design of a new trust model suitable for the case of cultural heritage institution—e-Infrastructures collaboration, including recommendations for the user authentication and access control system(s). It also documents the results of a survey on access and authentication services from e-Infrastructures to support trustable services of memory institutions.

This deliverable is divided into two parts. The first part is about digital cultural heritage institutions' requirements for trust in e-Infrastructure services. It presents an analysis of the concept of trust in digital preservation organisations (repositories) and discusses how the traditional trust model is transforming in distributed preservation architectures. The distributed architecture of today's digital archiving solutions adds extra complexity to the traditional trust model that memory institutions have relied on. The analysis shows that neither the existing trusted digital repository assessment methods nor the emerging cloud trust and digital trust methods can be applied to this new distributed preservation service architecture. The former do not cater sufficiently for distributed architectures and outsourcing of core services hitherto confined to the organisation that owns the repository. The latter focus too narrowly on security and performance issues leaving aside the legal, organisational and policy level aspects that repositories' trust is built on. As a first step to fill the gap in establishing trustworthiness of the joint repository—e-Infrastructure preservation service, this report proposes a risk assessment tool (in Appendix 1) that is focussing on the organisational and legal aspects of service provision.

The second part of the deliverable focusses on a more technical approach to access rather than governance and organisation level discussed in the first part. It documents and discusses the results of the DCH-RP survey on access and authentication services from e-Infrastructures to support trustable services of memory institutions. This part also includes recommendations for the user of authentication and access control systems that would be most suited for the digital cultural heritage.

Federated access is a key element of the DCH-RP roadmap and can bring many benefits for the users as well as for the resource providers. However, deploying federated access requires considerable technical expertise and manpower to set up the technical infrastructure, which is not always available in the arts and humanities sector. This support should ideally come from the federation operators, but their funding model and the availability of resources on their side does not always allow for that.

This deliverable provides a set of recommendations for the DCH-RP roadmap and DCH sector:

| Recommendations | Actors |
|---|---|
| *Engage the e-Infrastructure community in discussion of organisational, policy, legal and security risks that are specific to digital preservation to develop an understanding of the issues in both domains and guidelines for managing the identified risks* | DCH-RP partners |
| *Monitor the emergence of trusted digital repository audit and certification services and once launched, organise an interest group of cultural heritage institutions that are collaborating with e-Infrastructures to pilot the audit methods on distributed preservation architectures* | Digital preservation community |
| *Applications should use simple graphic interfaces, rather than* | Application developers and |

| | |
|---|---|
| *command line, to encourage wider usage* | resource providers |
| *Where federated access is provided, best practice guidelines should be followed to improve user satisfaction* | Resource providers |
| *Consider adding a cost/benefit analysis in the roadmap, which also includes considerations around reusing/sharing applications (federated access) versus managing services at institutional level* | DCH-RP partners |
| *Engage more actively with national R&E federation operators and ensure that funding is allocated to the federations for support activities* | DCH-RP partners |

# 1. INTRODUCTION

## 1.1 OBJECTIVES OF THE DELIVERABLE

The recent report by the European High Level Expert Group on Scientific Data (Riding the Wave 2010, p. 22) listed as one of the key challenges of e-Infrastructures the question of trust:

> "How can we make informed judgements about whether certain data are authentic and can be trusted? How can we judge which repositories we can trust? How can appropriate access and use of resources be granted or controlled?"

Traditionally cultural heritage institutions have enjoyed a high level of trust in society – people feel that they trust information they receive from archives, libraries, museums, because these institutions have been in the business of preserving information for a very long time. In the digital environment and networked architectures this is changing – digital content alters the picture as new technologies, competencies, risks, service models and stakeholders enter the preservation scene. The traditional trusted role of cultural heritage institutions to act as (long-term) mediators of information between its creators and users needs to be re-established when memory institutions involve third parties in their core business – preservation.

The DCH-RP project is developing a roadmap for preservation services where cultural heritage institutions rely on co-operation with e-Infrastructures. This necessitates the analysis of trust as one of the key concepts that this collaboration rests on, and that cultural heritage institutions need to be transparent about towards their stakeholders.

This deliverable outlines the design of a new trust model suitable for the case of cultural heritage institution—e-Infrastructure collaboration, including recommendations for the user authentication and access control system(s) that would be most suited for the digital cultural heritage. It also documents the results of a survey of access and authentication services from e-Infrastructures to support trustable services of memory institutions.

## 1.2 STRUCTURE OF THE DOCUMENT

This deliverable is divided into two parts. The first part – A (sections 2 - 6) is about digital cultural heritage institutions' requirements for trust in e-Infrastructure services. It presents an analysis of the concept of trust in digital preservation organisations (repositories) and discusses how the traditional trust model is transforming in distributed preservation architectures. It concludes with a proposal for a risk profile of distributed digital preservation services as a next step in this area.

The second part of the deliverable – B (sections 7 - 11) focuses on a more technical approach to access rather than governance and organisation level discussed in the first part. It documents and discusses the results of the DCH-RP survey on access and authentication services from e-Infrastructures to support trustable services of memory institutions. This part also includes recommendations for the use of authentication and access control systems that would be most suited for digital cultural heritage sector.

Appendix 1 includes a risk analysis tool that memory institutions can use to assess the policy, legal and organisational level risks in using distributed digital preservation services.

# A: DIGITAL CULTURAL HERITAGE INSTITUTIONS' REQUIREMENTS FOR TRUST IN E-INFRASTRUCTURE SERVICES

# 2. TRUST AND DIGITAL PRESERVATION – STATE OF THE ART

Digital curation is a complex field that requires competence in preservation, technology, metadata, risk management and so on (DCC, 2010), as well as availability of infrastructure and tools for carrying out both passive and active digital preservation. Not all digital repositories can be expected to deliver all digital preservation services to the same level of quality. Hence, questions over why should one trust a given repository to preserve digital content appear justified. When the repository involves third party service providers to preserve its clients' data, the trustworthiness and reliability of the third parties also becomes a demonstrable need.

## 2.1 THE CONCEPT OF A TRUSTED DIGITAL ARCHIVE

Claims of trustworthiness of digital archives are easy to make but are difficult to justify or objectively prove. A trusted digital repository is one whose mission is to provide reliable, long -term access to managed digital resources to its designated community, now and in the future (RLG/OCLC 2002, p 5). Trust in a digital repository is related not only to trusting the preservation methods applied by the repository, but also to broad organisational issues like funding base, policy framework, staff training, existence of transferable skills, and so on. A trustworthy digital repository will understand threats to and risks within its systems and organisation.

In 1996, the Commission on Preservation and Access (CPA) and the Research Libraries Group (RLG) joint Task Force on Archiving of Digital Information called the existence of a sufficient number of trusted organizations capable of storing, migrating, and providing access to digital collections "a critical component of the digital archiving infrastructure". The Task Force report proposed that a "process for certification of digital archives is needed to create an overall climate of trust about the prospects of preserving digital information" (CPA/RLG 1996).

An understanding of what are digital archive components and how is the preservation function embedded into the overall archive workflow is presented in the OAIS reference model. OAIS is a pivotal standard in the digital preservation domain, ISO 14721 *Space data and information transfer systems – Open archival information system – Reference model*. It is a functional framework that presents main components and basic data flows within a digital preservation system. It defines six functional entities that synthesise the most essential activities within a digital archive: ingest, preservation planning, archival storage, data management, administration and access. As a reference model, the OAIS standard does not imply a specific design or formal method of implementation (cf. Lavoie, 2004). Instead, it is left to the users to develop their own implementation by analysing existing business processes and matching them to OAIS functions.

Among the first to explore the characteristics of a trusted digital repository was the RLG and Online Computer Library Centre (OCLC) Working Group on Digital Archive Attributes. It released its report *Trusted Digital Repositories: Attributes and Responsibilities* in 2002 (RLG/OCLC 2002). RLG and OCLC sought to define the characteristics of "sustainable digital archives that could serve large-scale, heterogeneous digital collections held by national libraries, university libraries, special collections, archives, and museums". One of the qualities of the trusted digital repository (TDR) was set as: "compliance with the Reference Model for an Open Archival Information System (OAIS)". The OAIS Reference Model supplies a common framework, including terminology and concepts, for describing architectures and operations of digital archives.

Through this conjecture the concept of a 'trustworthy digital repository' became linked with a standard workflow model that a digital archive has to follow. Although the OAIS reference model does not prescribe any specific technologies or technology architectures, the 'trusted digital repository' also came to be understood as a centralised, single organisation-based preservation service model where the institution that provides the preservation service is also the owner of the digital repository system that houses digital objects. The practice of applying the TDR criteria over the next decade has demonstrated that the word 'trusted' in this concept should more appropriately have been 'quality' because essentially the TDR is about ensuring quality at the operational level of repository work. Being trusted to deliver quality service requires a step further – making the compliance with quality criteria transparent and verifiable by external stakeholders. Thanks to a strong striving towards increased reputation among repositories, the digital preservation community has accepted the term 'trusted' as a replacement for 'quality' and has gone on to develop audit methods that instead of establishing compliance with quality standards are claiming to establish the trustworthiness of a repository.

## 2.2 THE TRUSTED DIGITAL REPOSITORY AUDIT METHODS

To begin answering questions on trustworthiness of digital preservation repositories a number of approaches have been proposed that rely on different methods of audit.

The 2002 TDR report also provided a comprehensive look at the organisational context for a digital preservation program and made a direct call for the development of a digital audit and certification program. The following year (2003) RLG and NARA established the joint Digital Repository Certification Task Force with membership from the U.S., U.K., France, and the Netherlands representing multiple domains including archives, libraries, research laboratories, and data centres from government, academic, non-profit, e-science, and professional organizations (Ambacher 2007, p. 3). The task force worked on developing an audit checklist that was released as a draft for public comment in August 2005 and aimed to develop criteria to "identify digital repositories capable of reliably storing, migrating, and providing access to digital collections" (RLG/NARA 2005). Certification of repositories was foreseen to instil confidence in data creators, resource allocators, and users that a certified repository meets recognised standards and can fulfil its preservation and access provision mission.

The final version of the audit checklist was published in March 2007 as the *Trustworthy Repository Audit and Certification (TRAC) Criteria and Checklist* (OCLC/RLG 2007). The checklist presents around 90 organisational, technological and digital object management criteria for digital repositories. Many are based on the principles, terminology and functional characteristics outlined in the OAIS reference model (ISO 14271).

In 2004 the German Network of Expertise in Long-term Storage of Digital Resources (nestor) established a working group on the certification of trustworthy archives.[1] Building on the draft version of the TRAC checklist, the nestor group focused on identifying features and values that are relevant for evaluating both existing as well as planned digital object repositories. The *nestor criteria for auditing digital preservation repositories* were released in 2006 (nestor 2006) and updated in 2008 (nestor 2008). The nestor checklist covers the technical, organisational and financial characteristics of a digital repository with examples and perspectives that are of particular relevance to the legal and economic contexts and operational situation in Germany (Dobratz et al. 2007). On the conclusion of the nestor project, work on the trustworthiness criteria was transferred to the German national standards body and a new version of the criteria was

[1] http://www.langzeitarchivierung.de/Subsites/nestor/EN/Workinggroups/arbeitsgruppen_node.html

published as a national standard DIN 31644:2012 *Information and documentation - Criteria for trustworthy digital archives*.[2]

In early 2007 the DigitalPreservationEurope project (DPE) and the UK Digital Curation Centre (DCC) published their joint work as the *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA) (Hofman et al. 2007). This tool presents a methodology for repository self-assessment and characterises digital curation as a risk-management activity; the job of the digital curator is to rationalise the uncertainties and threats that inhibit efforts to maintain digital object authenticity and understandability, transforming them into manageable risks. The DRAMBORA methodology helps to determine whether the repository has made every effort to avoid and contain the risks that can impede its ability to receive, curate and provide access to authentic, and contextually, syntactically and semantically understandable digital information. Awareness of threats to its systems and operations, and the ability to deal with the expressed risks are the basis for the claim of being a trustworthy digital repository. An online assessment tool was released in 2008 to guide and document the repository assessment.[3]

The Data Archiving and Networked Services (DANS) in the Netherlands published 16 guidelines to help a data archiving institution striving to become a trusted digital repository in 2008. The guidelines are called the *Data Seal of Approval* (DANS 2009).[4] The assessment is a two-stage process where the repository carries out its own self-assessment, publishes the results and then applies for a peer-review by a member of the international DSA assessment group. The reviewer recommends to the board whether the guidelines have been complied with and whether the DSA logo can be awarded to the data repository (Harmsen 2008, p. 1). The Data Seal of Approval does not include a site visit and relies on the availability of public documentation and the public nature of all self-assessment statements that result in a Seal being awarded as a means of ensuring trust in the process of peer-review.

The TRAC checklist and nestor criteria became a basis for developing a new set of criteria on which formal audit and certification of digital repositories can be based. This work resulted in 2012 as an ISO standard in support of the OAIS reference mode – ISO 16363:2012 *Audit and certification of trustworthy digital repositories*. The scope of the checklist is explicitly the entire range of digital repositories; its criteria are empirically derived and consistent measures of effectiveness have been ascertained (Ruusalepp et al. 2012, p. 124). A team of experts conducted a series of pilot audits in 2011 as part of the APARSEN project, to test the methodology of the ISO 16363 standard (APARSEN 2012).

The same working group is working on an adjunct standard *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories* (ISO/DIS 16919). Once completed, this standard will provide normative rules against which an organization providing audit and certification of digital repositories may be judged, and it describes the auditing process.

The checklists or metrics that TRAC, ISO 16363 and DIN 31644 provide as a basis for audits are presented as quality criteria that a trustworthy repository should meet. They provide digital repositories of all sizes with directions for demonstrating their adherence to quality and consistency, their respect for data integrity, and a commitment to the long-term preservation of and access to the information entrusted to their care. The metrics are mostly derived from practice and based on the OAIS Reference Model. As such they identify the quality of work as complying with the OAIS model's principles. The criteria are divided into groups that reflect levels and types of activity involved in running a digital repository, for example (ISO 16363):

[2] http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855&artid=147058907
[3] http://www.repositoryaudit.eu/
[4] http://www.datasealofapproval.org/

- Organizational infrastructure, that addresses issues such as governance and organizational structure, staffing, procedural accountability, the policy framework, financial sustainability, and contracts, licenses, and liabilities.
- Digital object management, that assesses the acquisition of content, creation of the Archival Information Package (AIP), preservation planning, the actual preservation of the AIPs, and the management of information (i.e., metadata) and access.
- Technical infrastructure and security risk management.

The future certification process based on these standards is guided by a Memorandum of Understanding, signed as a European Framework for Audit and Certification of Digital Repositories.[5] It divides certification of trusted digital repositories into three levels:

- Basic Certification should be granted to repositories that obtain DSA certification through a process of self-audit and the public release of a peer-reviewed statement from another organization which has previously received the DSA;
- Extended Certification is granted to Basic Certification repositories that also perform a structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644; and
- Formal Certification is granted to repositories that in addition to Basic Certification obtain full external audit and certification based on ISO 16363 or the equivalent DIN 31644.

Of these three levels, currently only DSA certification is operational and so far 24 organisations have received the Seal. Some federated data preservation and access services are beginning to make DSA a precondition for membership in a network of repositories (e.g. CESSDA,[6] CLARIN,[7] EUDAT[8]). The Center for Research Libraries (CRL) in the U.S. has been carrying out repository audits based on the TRAC checklist and has issued four certificates of a trusted digital repository based on this.[9] The German nestor network has started a *nestor Seal for Trustworthy Digital Archives* scheme that is issued, for a fee, after successful extended certification.[10] Formal certification based on ISO 16363 can commence once the associated standard ISO/DIS16919 is finalised and published.[11] DRAMBORA assessments are being carried out continuously and close to 300 repositories have created their risk profiles using the on-line toolkit.

In **summary**, despite the criteria describing and checklists for assessing trustworthiness of digital repositories having been around for over a decade, the practice of applying them has been limited to self-assessment and only a handful of repositories have been formally audited and certified.

Self-assessment against any of the five criteria described can improve the quality of repository work. However, if the results of this assessment are not made public, its positive impact on trust towards the repository is indirect and only emerges over a long period of time. If the results of self-assessment are made public, this may increase trust towards (and eventually reputation of) the repository among some stakeholders. However, there is no objective benchmark available for conferment of 'trust' or for measuring how much the trust will increase, because the assessment criteria deal instead with quality of operations within the digital repository. If the self-assessment results are made public and interpreted by peer-reviewers or an external auditing committee, it is possible for a repository to receive a confirmation of this fact in the form of a certificate (e.g. Data Seal of Approval, nestor Seal, TRAC certificate of a TDR).

[5] http://www.trusteddigitalrepository.eu/Site/Trusted%20Digital%20Repository.html
[6] http://cessda.org/project/doc/D10.4_Data_Formats.pdf, p. 16
[7] http://www.clarin.eu/node/3767
[8] http://www.eudat.eu/system/files/EUDAT-DEL-WP4-D4%203-Trust%20Establishment%20Report.pdf, p. 9
[9] http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories
[10] http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html
[11] http://www.iso16363.org/preparing-for-an-audit/

The essence of such certificates at present is that the quality of work the repository is delivering has been externally reviewed and confirmed as being at the level that the repository has itself ascertained. Maturity of digital archive quality standards has not yet reached the same level as in quality management, risk management, records management or information security management, where the so called management system standards (MSS) exist.[12] The MSS provide a formally agreed mechanism for maintaining quality in-between audits that are undertaken at regular intervals. The quality and trustworthiness of digital repository services to external stakeholders is not explicitly part of the existing five assessment methods because the assessment does not involve external parties, their expectations or satisfaction with the services delivered. The trust models in digital archives require, therefore, further analysis and expansion to include architectures that involve third party service providers.

## 2.3 MODELLING TRUST IN A CULTURAL HERITAGE INSTITUTION

Trustworthiness in a digital preservation repository needs to be demonstrated towards both internal and external stakeholders. First of all, the management, staff, fund-makers and partners of a repository must all be satisfied that their efforts are capable of meeting formal mandates and expectations. Similarly, information creators, depositors and users are interested in obtaining similar assurances of the competencies of the organisation providing maintenance, preservation and dissemination services.

The original *Trusted Digital Repositories: Attributes and Responsibilities* (2002) report discussed trust-building on three levels:

1. How repositories earn the trust of their designated communities.
   Thus far, libraries, archives, and museums have shown they can create and provide access to digital materials. Users now rely on institutions to provide ongoing development of systems that support long-term access to the materials. Over time, institutions will keep the users' trust so long as they sustain reliable access to information.

2. How repositories trust third-party providers.
   Service providers gain the trust of cultural institutions through a combination of proven reliability, fulfillment of contractual responsibilities, and demonstrated sensitivity to community issues. Without demonstrated experience, the service provider cannot prove reliability. To resolve the tension between a repository's appropriately high standards and the attempts to meet the challenge, a combination of repository attributes and other criteria must be identified to foster interaction and begin to lay the foundation for trust between cultural institutions and third-party providers. One option may be to identify certain attributes in a third-party service provider that the institution requires of itself.

3. How users trust the documents provided to them by a repository
   A user must be able to trust digital documents provided by digital repositories. Authentication of digital information includes the ability to detect change to a digital document.

This initial trust network of a digital repository has been amended through subsequent work on criteria of trustworthiness that has in addition focused on the operational level of carrying out digital preservation activities in a repository (cf. digital object management in ISO 16363 and DIN 31644). The stakeholders included in a typical trust network of a digital repository today can be grouped into four types (cf. Figure 1 below):

1. Data creators – for example, publisher, record-creating agency, donor – are typically interested in trusting the preservation and access services of the archive;

[12] Cf. ISO 9000, ISO 31000, ISO 30300, ISO 27000 families of standards that establish a management system for continuous improvement of quality

2. Data users – readers, researchers, agencies, etc. – are interested in trusting both the services and the data they get from the archive;
3. Repository owners or funders – for example, government, shareholder, agency that the digital archive is part of, etc. – are usually interested in trusting the services of the archive as a whole;
4. Data rights owners – the subjects of data held by the repository, like an author of a book in digital libraries, citizens as co-creators of public records – are interested in trusting the organisation, the digital curation services and the data.

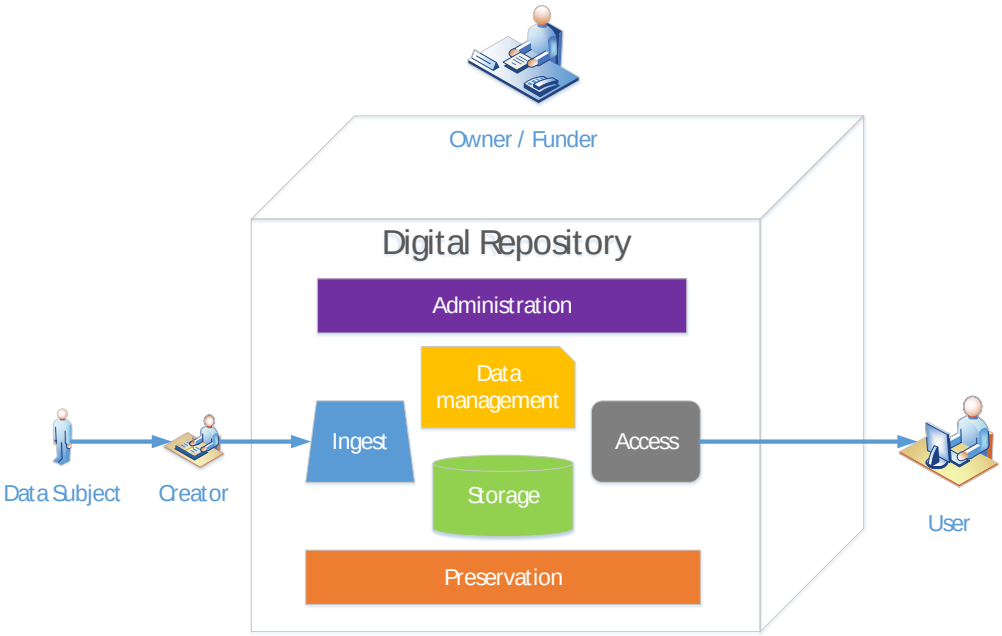


*Figure 1: Typical trust network of a digital repository.*

Depending on the mission and mandate of the archive, the trust network can be larger or have specific relationships. Quality measures that inculcate trust may depend on the stakeholder, and may differ from one stakeholder to another. For data users a trusted, usable, contextualised object is directly experienced and used while the repository funders will not be directly dependent on the quality of individual data objects, but will be concerned when data users are dissatisfied with the quality or reliability of preserved data. The data creators need to trust that the data managers/curators will take care of proper data lifecycle management over a long period of time and that appropriate credit is given to the creators of the stored data (EUDAT 2013, p. 7).

The trust models that currently exist for archive services are predominantly based on the idea of a centralised digital repository owned and managed by the archive itself (see Figure 1). The changes that a distributed digital archiving and preservation service brings to the trust network need to be further elaborated and analysed in order to support outsourcing of such services to e-Infrastructures.

# 3. DISTRIBUTED DIGITAL PRESERVATION SERVICE MODELS

The DCH-RP deliverable D3.1 *Study on a Roadmap for Preservation* analysed digital preservation service models (Ch. 3.1 and 3.2) and concluded that although the basic archiving workflow is provided by the OAIS reference model, it does not articulate clearly how it can cater for distributed archiving architectures. Cloud, grid and e-Infrastructure service architectures vary significantly and do not allow for a uniform mapping of preservation services to a single architectural model. Conceptualising and modelling the joint service architecture is only in developing phases.

> "The main challenge remains how to piece the two types of services together – whether to use the OAIS as the underlying model and map the grid/cloud services to it as "add-ons", or use the service and architecture models provided by the e-Infrastructures and embed preservation services into them."

The issue in short is the immature state of conceptualisation of digital preservation services for distributed architectures. The predominant current service model is where a single institution owns a digital repository and is itself responsible for all of its operations. A recent analysis from the National Repository of Ireland described this service model as a single site repository that (DRI 2013):

> "hosts databases and the associated functions of archiving, including data preparation and preservation, within one location. Many national repositories are single-site repositories. Many will locate off-site back-up in multiple locations, but the main technological infrastructure is located in one site."

The DRI report also notes that "since 2009 there has been a demonstrated shift towards the establishment of multi-site repositories, in which the technical infrastructure is federated across a number of repository sites". The multi-site repositories host data within a federated structure that allows sharing of metadata and data across institutions.

At present there are not formal (reference) models that describe distributed digital preservation services because the practice of using distributed service architectures is only emerging. An early description of the distributed digital preservation (DDP) model was described in the Educopia Institute and MetaArchive report *A Guide to Distributed Digital Preservation* (Skinner, Schultz 2010). The report describes the principles and advantages of a federated repository architecture based on the MetaArchive Cooperative experience with a Private LOCKSS Network (PLN). The report sets requirements for the number of copies of each archived object and their storage conditions that focus on best practice disaster preparedness (pp. 12-13):

- Content should be replicated at least three times.
- Sites preserving the same content should not be within a 75-125-mile radius of one another.
- Preservation sites should be distributed beyond the typical pathways of natural disasters, such as hurricanes, typhoons, and tornadoes.
- Preservation sites should be distributed across different power grids.
- Preservation sites should be under the control of different systems administrators.
- Content preserved in disparate sites should be on live media and should be checked on a regular basis for bit-rot and other issues.

Researchers at the University of Tsukuba, Ibaraki, Japan explored ways to align the OAIS reference model with a layered model of cloud computing, in which services are abstracted and shared between layers. Their conclusion is that despite some apparent incompatibilities, notably the often synchronous nature of preservation workflows, there is value in adopting a layered model with aspects of trusted bit level storage/API, information description and function distributed across PaaS and SaaS layers of the

cloud architecture. They define a layered model for a cloud archiving system that allows sharing of functionality and information objects by making them available as services to higher layers of cloud services (Askhoj et al. 2011, p. 180):

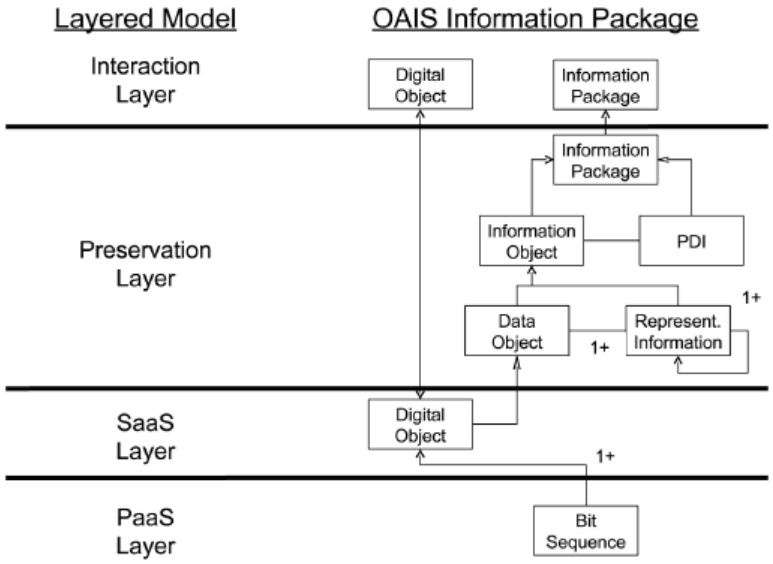


*Figure 2: OAIS entities mapped to cloud service levels.*

The SHAMAN project implemented a middleware solution that linked an iRODS-based archiving solution with a Grid service (Wittek, Darányi 2012). Their proposed service framework includes middleware for digital preservation that is agnostic to whether the environment is a Grid or a cloud. The middleware hides the complexity of the switch between a Grid or a cloud, irrespective of whether the need for change arises from storage requirements or computational demand, enabling a smooth transition between the different types of infrastructures.

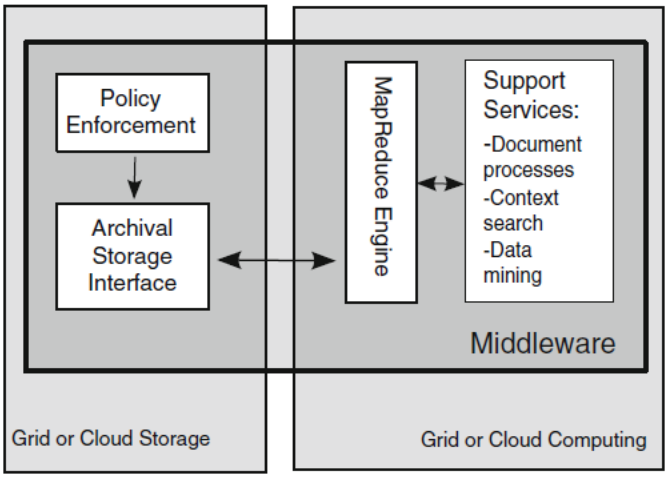


*Figure 3: Proposed SHAMAN grid service framework.*

The TIMBUS project[13] has continued to study the data grids track and has defined a grid federation framework to support further replication on the storage layer of preservation (Atunes, Pina 2011).

[13] http://timbusproject.net/

Some general requirements on the cloud for offering digital preservation services were listed by the authors of the Long-Term Digital Retention and Preservation Reference Model in 2011. Their *Cloud Digital Archive and Digital Preservation Service Requirements*[14] defines Cloud Digital Preservation Service as a "service providing digital preservation of information and data". A digital preservation service includes a comprehensive management and curation function that controls its supporting infrastructure, information, data, and storage services in accordance with the requirements of the information objects it manages to accomplish the goals of digital preservation.

A JISC workshop Curation in the Cloud produced a report *Digital Curation and the Cloud* (Aitken et al. 2012) that includes telling examples of using different cloud services in the digital archive workflow. The main distinctions in the report are drawn between commercial (public) cloud, community cloud and private cloud, and hybrid combinations of in-house services combined with one of the cloud types. In the analysis of best fit between requirements of digital curation and services offered by the cloud, the report concludes that (pp.10-11):

> "The cloud is considered a cheap storage opportunity (more so than SAN) and an appropriate choice for large, seldom-accessed resources. Fears over the cost implications of frequent cloud accesses limit its viability for other content. For tasks that are infrequent and/or difficult to anticipate and plan for, the cloud is a good fit, offering resource elasticity and a metered charging model. Interactive, application-style processes are generally less able to capitalise on such opportunities, and are more suited to traditional service models, or the highest level SaaS cloud model."

Of these initiatives, only the DDP model continues to be developed – it is regularly presented and discussed at key community events (Schultz, Gore 2010; Trehub, Halbert 2012; Skinner et al. 2013; Zierau, Schultz 2013).

The DCH-RP deliverable 3.1 *Study on a Roadmap for Preservation* (DCH-RP 2013) identified some examples of integrating cloud, grid[15] and e-Infrastructures into the preservation services model. All of these demonstrate that cloud and grid adoption need not be considered as an all-or-nothing process. Different cloud services can be effectively combined, and similarly there are numerous examples of cloud services integrating effectively with local provisions. The following generalised service models summarize the existing permutations of offering digital preservation services as federated digital archives or as distributed services relying on cloud or grid providers:

1) A **cooperative** file sharing model where each participating archive is a node in a network that hosts some other node's data. The best-known example of such a network is LOCKSS (Lots of Copies Keeps Stuff Safe).[16] Figure 1 below is an abstraction of the federated model where DP stands for Digital Preservation. In the case of LOCKSS networks, DP is limited to bit-level preservation.

[14] http://www.ltdprm.org/reference-model/preservation-in-the-cloud/cloud-archive-requirements

[15] Grid and cloud architectures are similar in using distributed resources. Grid architectures are based on shared resources while cloud computing is based on leasing resources. There are also divergences – the grids are mostly based at universities and academic institutions while cloud services mostly come from the commercial sector. Two popular types of grids are data grids and computation grids; the idea of shared storage was naturally appealing to the digital preservation community, given the scale of preservation tasks. With the introduction of cloud services, the concept of what such shared resources could offer evolved further and now includes offering of software, infrastructure and platforms as services (SaaS, IaaS, and PaaS). All these are relevant to preservation.

[16] http://www.lockss.org/

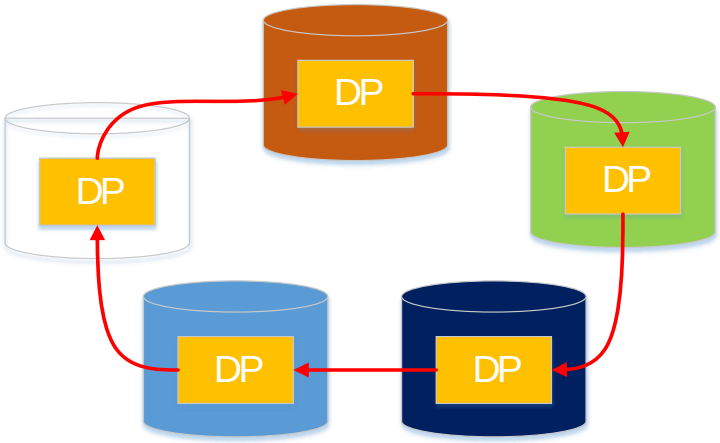


*Figure 4: A simplified federated repository service model.*

The LOCKSS service model is based on a secure, closed-access network of servers set up between the network members. Each institution in the network runs a server that is linked securely to the network but maintained by different systems administrators. A new ingested object is replicated to other nodes in the network for preservation. The servers also check in with each other to make sure that all copies of the objects are identical. If a mismatch is detected, the servers come to a consensus regarding which copies are correct and which do not match, and then the network repairs the "bad" files.

Several cooperative efforts that are based on the LOCKSS or Private LOCKSS Network (PLN) have emerged, mostly in United States, that share some functions and services of a digital archive:

- LOCKSS[17] - the LOCKSS Program is an open-source, library-led digital preservation system built on the principle that "lots of copies keep stuff safe".
- Data Preservation Alliance for the Social Sciences (Data-PASS)[18] – is a partnership of five major U.S. institutions with a strong focus on archiving social science research.
- MetaArchive Cooperative[19] - is a cooperative membership organization where each member runs a server for the MetaArchive network and prepares its own content for ingest.
- UK LOCKSS Alliance[20] - is a cooperative movement of UK academic libraries that are committed to identify, negotiate, and build local archives of material that librarians and academic scholars deem significant.
- LuKII (LOCKSS und KOPAL Infrastruktur und Interoperabilität)[21] - is a Private LOCKSS Network to conceptualize and implement interoperability among several preservation systems. This program has demonstrated bi-directional content moves between the German National Library's KOPAL system and a Germany-wide Private LOCKSS Network.
- DPN[22] – Digital Preservation Network in the US serves as a preservation backbone that replicates ingested content among several nodes with robust (bit) auditing and repair functions.

[17] http://www.lockss.org/
[18] http://thedata.harvard.edu/dvn/dv/datapass/
[19] http://metaarchive.org/
[20] http://www.lockssalliance.ac.uk/
[21] http://www.lukii.hu-berlin.de/
[22] http://www.dpn.org/

2) A **centralised** archive that acts as a service provider for a number of institutions participating in a network. The central archive uses external cloud or grid service for its storage layer or as an extra off-site storage.

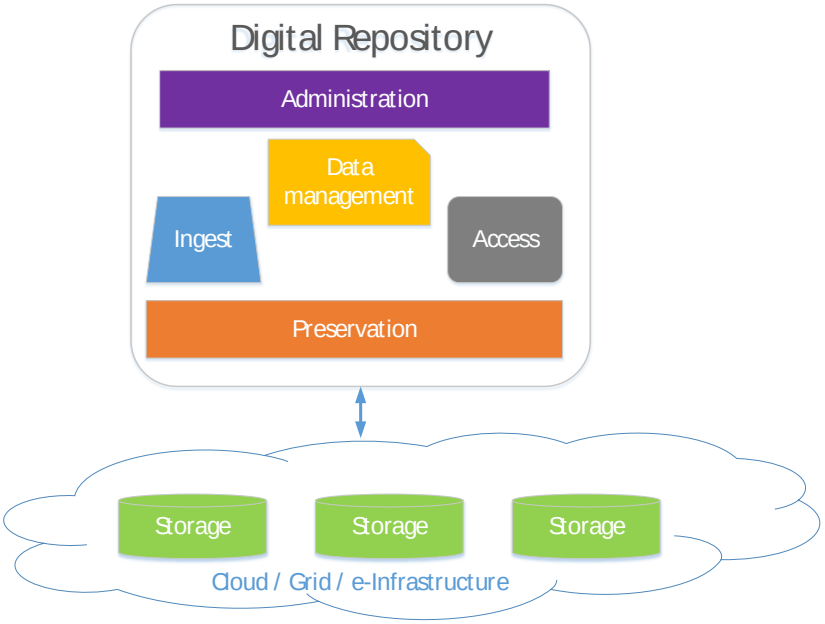


*Figure 5: A centralised archive service that outsources storage to cloud or grid service.*

Examples of such centralised service provision models are numerous since the replication of storage adds to the security of preservation services. Examples of such service provision include:

- Chronopolis[23] - is a digital preservation data grid framework developed by the San Diego Supercomputer Center (SDSC). It provides cross-domain collection sharing for long-term preservation and is based on an iRODS (Integrated Rule-Oriented Data System)[24] data grid.
- Texas Digital Library[25] and APTrust[26] that rely on DuraCloud service[27] for their storage component.
- DRI[28] - Digital Repository of Ireland that collaborates with the Trinity College Dublin Centre for High Performance Computing[29] to develop a storage solution for the repository based on a Ceph storage system.[30]

3) A **network of repositories** that share a cloud or grid-based storage that is replicated between multiple sites to achieve more secure replication of stored data.

[23] http://chronopolis.sdsc.edu/
[24] http://www.irods.org/
[25] http://www.tdl.org/
[26] https://wiki.duraspace.org/display/aptrust/Architecture
[27] https://wiki.duraspace.org/display/DURACLOUD/DuraCloud
[28] http://www.dri.ie
[29] http://www.tchpc.tcd.ie/node/981
[30] http://ceph.com/

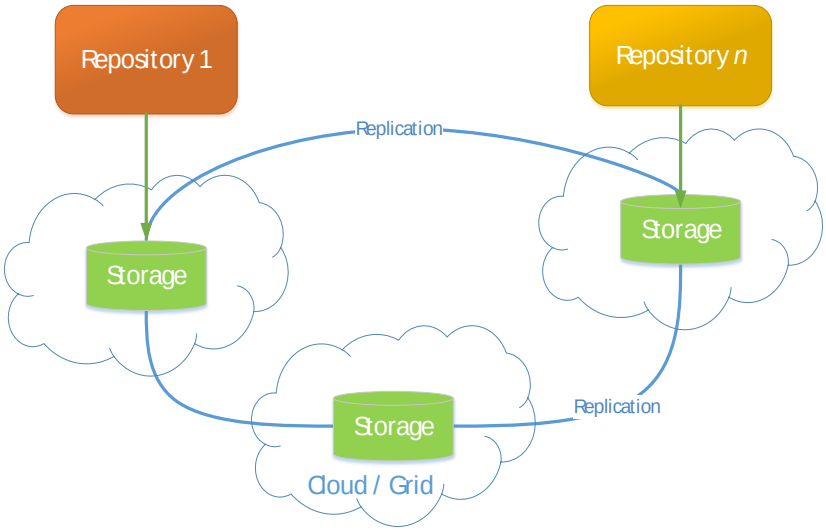


*Figure 6: A network of repositories sharing a storage network.*

A shared storage service layer based on grid or cloud infrastructure is analogous to the previous service model (see 2) above) but there are extra services agreed between several storage providers to ensure data redundancy and fixity.

- This service model has recently been deployed by the EUDAT[31] project that offers data repositories a data replication service as part of its B2SAFE service.[32]

4) **Repository outsources** one or several functions other than storage to a cloud or grid service provider. The outsourced functions could be computationally intensive, like quality assurance and conversion at the ingest stage of archive workflow; metadata management combined with user access and authentication (see the second part of this deliverable for discussion of user authentication issues); user access that is storage and computationally intensive (e.g. for audio-visual content); or active digital preservation processing (e.g. file format conversion processes on large collections).

[31] http://www.eudat.eu/services
[32] http://eudat.bsc.es/b2safe

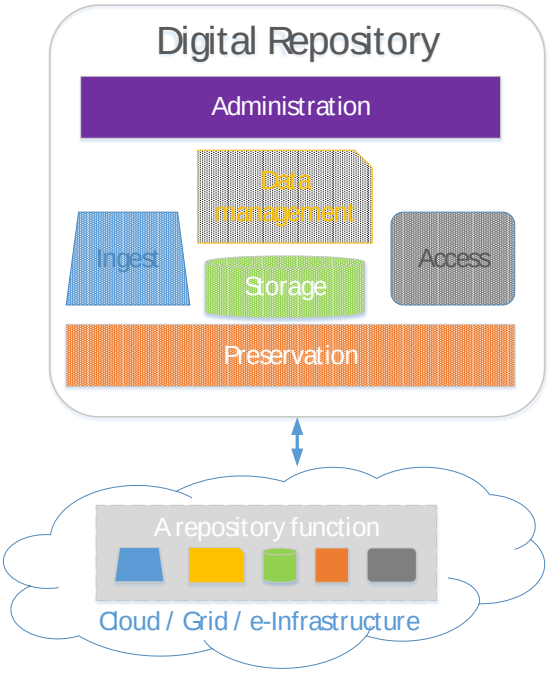


*Figure 7: A repository function outsourced to cloud or grid service provider.*

An example of such a solution is the e-Culture Science Gateway that was developed as part of the Indicate project,[33] one of the predecessors of the DCH-RP project and has now been updated to become the DCH-RP e-Culture Science Gateway (eCSG).[34] The Indicate project developed eCSG to host the catalogue and user access system to Italian libraries. As part of the DCH-RP project, further pilots are being implemented on the platform (see also D5.2 and D5.3).[35]

5) **A cloud or grid service provider offers** all repository services and effectively becomes a digital preservation repository. Although no examples of full-scale digital repositories being supported on cloud or grid platforms are available, these solutions are being discussed (cf. D3.1 ch. 4.1.1) and it is likely that an institution somewhere may have implemented a private cloud technology to host its repository system. The DuraCloud service[36] in the US is at present the closest known example to this scenario.

[33] http://www.indicate-project.eu/
[34] http://ecsg.dch-rp.eu/
[35] http://ecsg.dch-rp.eu/proofs-of-concept
[36] http://duracloud.org/

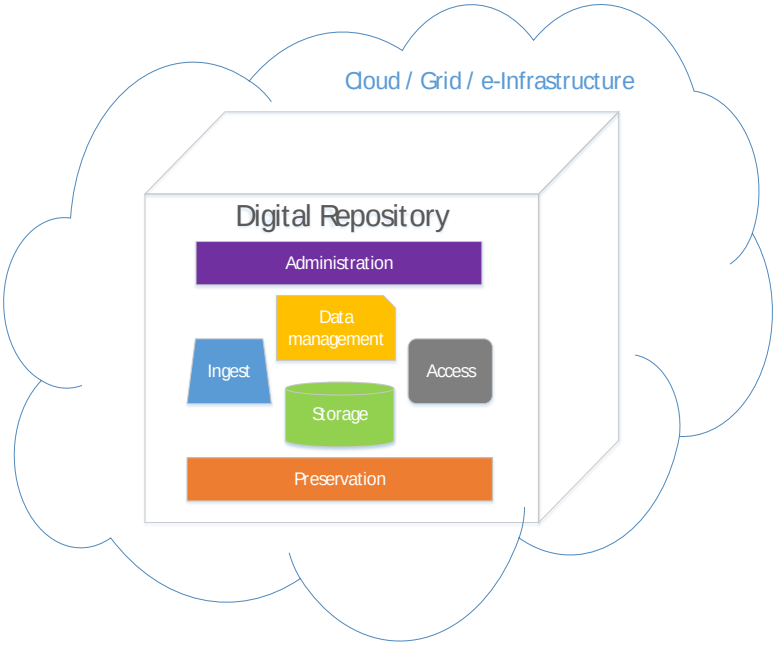


*Figure 8: Repository fully hosted by a cloud or grid service provider.*

The next chapter will model the trust relationships for each of these service models.

# 4. MODELING TRUST IN DISTRIBUTED PRESERVATION SERVICES

Similar to the lack of a reference model for distributed digital preservation services (cf. also DCH-RP deliverable D3.1), no trust model for a distributed preservation repository system yet exists. The need for a trust model for distributed digital preservation solutions has been discussed through a number of research papers.

Berman et al. (2007) describe the Chronopolis cooperative as a virtual organisation (federation) that exhibits trust both from dispositional (the natural tendency of an individual to trust other people) and situational (dispositional trust combined with structural and situational factors) perspectives. They conclude that "in formalizing the trust relationships between preservation providers, partners, and users, many issues are left unresolved".

Day (2008) discusses how trust comes to the fore in many areas of digital preservation where collaboration is necessary; this includes participation in strategic alliances and research initiatives, and in the provision of shared services like registries. He suggests that self-assessment tools like DRAMBORA could be used to help develop shared organisational cultures that are focused on solving long-term preservation challenges in an incremental and managed way.

Walters and McDonald (2008) use the example of the US Federal Reserve Bank regional governance (trust federation) model as an exemplar for centralized authority while providing for distributed independent organizational governance.

Schultz and Gore (2010) stress that "distributed digital preservation solutions must communicate trust to their Designated Communities as they continue to mature". Applying the TRAC checklist to the MetaArchive Cooperative distributed digital preservation solution (PLN) revealed that the current metrics for gauging trust in digital preservation could be readily applied to distributed solutions but because these metrics often presume a more centralized approach to preservation, there is a pressing need to "apply them carefully and with great thought".

The EU High Level Experts Group on Scientific Data advise in their final report that "if science is to advance, […] questions of trust must be answered by the infrastructure, itself, because data-intensive science operates at a distance and in a distributed way, often among people who have never met, never spoken, and, sometimes, never communicated directly in any form whatsoever. They must share results, opinions and data, but in truth, they have no real way of knowing for sure if, on the other end of the line, they will find man or machine, collaborator or competitor, reliable partner or con-artist, careful archivist or data slob. How will we judge the reliability and authenticity of data that moves from a personal archive into a common scientific repository?" (Riding the Wave 2010, p. 17).

The recent EUDAT report on *Trust Establishment* (EUDAT 2013) describes the attributes of data objects that contribute towards trustable data and discusses organisational components that engender trust in a networked service:

- Agreements, legal framework, governing structure
- Years of collaboration
- Relations within communities
- Configurable solutions.

The trust network for EUDAT's replication service between two data centres and their partners is depicted as follows (EUDAT 2013, p. 19):

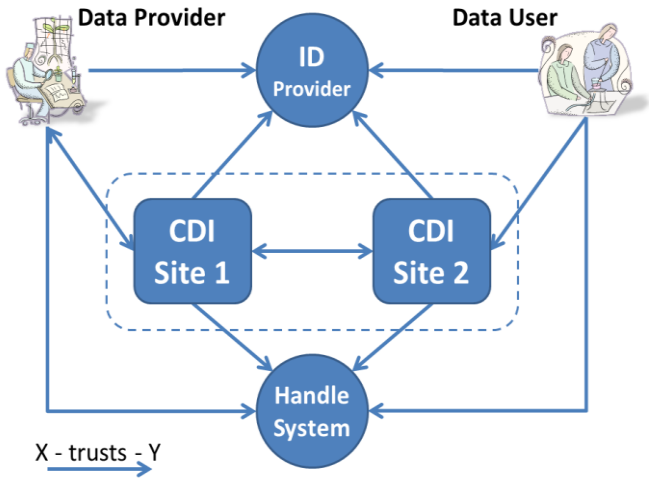


*Figure 9: EUDAT replication service trust network.*

The described trust network relies strongly on a shared user authentication and identification solution (ID Provider) and a service providing persistent identifiers to the objects in the digital repository (Handle System).

The need for a transitive trust model for distributed digital preservation solutions is, thus, accepted in the research literature, but as yet none of them have been implemented or could be relied on as working best practice.

## 4.1 TRUST NETWORK OF DIGITAL REPOSITORY-E-INFRASTRUCTURE PARTNERSHIP

When the trust network of a single-institution digital repository that was presented in Figure 1 (see Ch. 2.3 above) is augmented with e-Infrastructure as a service provider / subcontractor to the repository, the trust network becomes more complex (see Figure 10).

The distributed digital preservation trust model includes both inter-organisational trust relationships, where both the trustor and the trustee are organisations, as well as individual-organisation type trust relations where an individual (trustor) is interested in trusting an organisation or organisations (trustees). The individual-organisation trust relationships are by definition asymmetric because neither Depositors nor Users will, under normal circumstances, have sufficient competence to evaluate the quality of preservation operations and services provided by the Repository. The inter-organisational trust relationships are usually governed by contractual agreements (SLA, OLA) and involve the required degree of transparency and accountability from both sides to inculcate trust.

Trust decisions are always situated and tied to a specific context. The trust relationship between two agents is summarised in the following formula (Castelfranchi and Falcone 2010, p. 36):

$$\text{TRUST}(X\ Y\ C\ \tau\ g_x)$$

The formula reads: X trusts Y in context C to perform task τ and (thus) realising the goal g (result that X was expecting). For the above trust network, an example of the trust relationship would thus be: the Creator trusts the Digital repository, with the use of services from an e-Infrastructure, to preserve the deposited data for the agreed period of time without jeopardising the quality of data. Or for the inter-organisational scenario: the Repository trusts the e-Infrastructure to provide secure storage service under the conditions of the agreed Service Level Agreement (SLA).

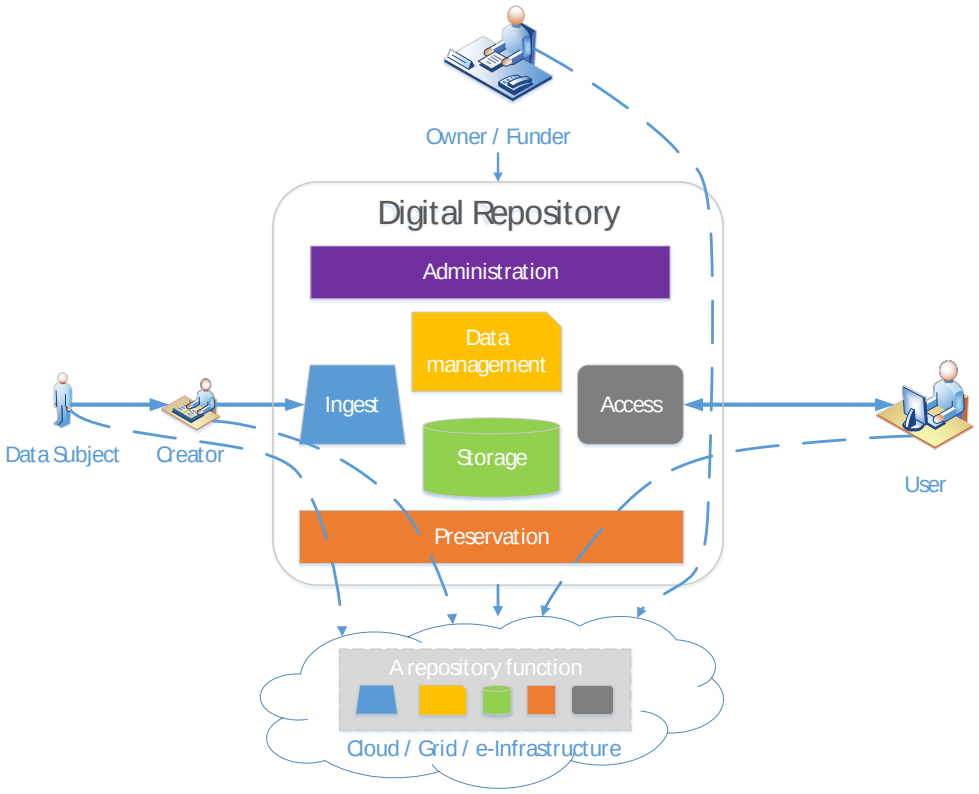


*Figure 10: A general trust network of a digital repository and e-Infrastructure joint service.*

The main change in the trust model from the Creator, Owner/Funder and User perspective that adding e-Infrastructure to the preservation service provided by the repository is causing is in the context (C) of the above formula. These stakeholders still assume they can trust the repository to deliver its services as usual, or improve the service quality, when an e-Infrastructure becomes a third party in the service provision context. The exact nature of the person-organisation trust relationship in this network has not yet been studied in detail, but is dependent at least on:

- The nature of material that is being preserved – for example, an in-copyright e-book will have different preservation and access requirements from open public records or a digitised image of a museum artefact.
- The purpose of preserving the object – organisations tends to impose less stringent requirements for short term retention of objects for, for example compliance with regulatory requirements than when depositing collections for long-term digital preservation.
- The intended users of the preserved object – both depositors and users tend to prefer subject or discipline or data type specific repositories to carry out preservation and disseminate data to a specific user group. For example a collection of video recordings is more likely to be deposited to a repository specialising in preservation of audiovisual material; a composer's private archive is likely to end up being preserved by a museum dedicated to theatre and music studies, because its core user group is comprised of researchers in these subjects and hence this repository would provide the best possible exposure to the composer's archive.

### 4.1.1 The nature of trust relationships

Mapping out all possible combinations of trust conditions for different data types, reasons for preserving data, intended user groups and the five distributed preservation scenarios (see Chapter 3 above) and for each of the four trust stakeholders, would be impractical. Instead, this chapter describes briefly the nature of each trust relationship in the network model (see Figure 10).

The **Depositor -> Repository** trust type can be described as 'dispositional trust'. It is the trustor's belief that it will have a certain goal B in the future and, whenever it will have such a goal and certain conditions obtain, the trustee will perform A and thereby will ensure B. This generalised expectancy – dispositional trust – is operational when a decision to trust or not to trust is made in the absence of direct evidence as to whether another is or is not trustworthy (Rotter 1967, p. 651). In the Depositor -> Repository situation, the trust is almost invariably of dispositional type – when the Depositor consigns material to a Repository he trusts the Repository to carry out active digital preservation actions on the deposited content in the future. Hence, the Depositor's trust expectation is directed towards the future activities of the Repository, yet the Depositor have very little (and sometimes no) evidence at his disposal that these activities will result in a successful outcome. The digital preservation community has not yet developed a practice for collecting an evidence base for successful and unsuccessful digital preservation actions. The Depositor, therefore, is left with observing current behaviour and studying documentary evidence on Repository activities, and making only predictions on the future reliability of the Repository services based on its policies. Reliance primarily on dispositional trust introduces uncertainty and risks into the trust-decision outcomes.

The **Depositor -> e-Infrastructure** trust relationship is dependent on the type of support service the e-Infrastructure is providing to the Repository. In most cases, cooperation with an e-Infrastructure service provider will improve the Repository's service quality and therefore increase the Depositor's trust in Repository's service. Thus:

Depositor -> (Repository + e-Infrastructure) > Depositor -> Repository

However, this statement is conditional and will depend on the type of services e-Infrastructure is providing to the repository – storage level services increase the chances of data replication and thus improve the chances of data being retained in case of technical failures in the storage systems. e-Infrastructures can have demonstrable track records of storing large quantities of valuable data for other clients and the Repository can benefit from this inferred trust context. The opposite may also be true if the Depositor is not confident in the ability of the e-Infrastructure to provide required level of security or control over individual objects in a shared storage space, e.g. a confidential document or a publication protected by copyright that will lead to litigation when leaked or mis-used.

There can also be **Repository -> Depositor** trust issues related to the authenticity and reliability of the deposited material that will influence the quality of service the Repository is providing to Users. Providing bad quality data to users will reduce the trust Users have in the Repository and will reflect badly on the reputation of the Repository. However, the Repository -> Depositor trust relationship has no bearing on the involvement of the e-Infrastructure provider in the service model.

The **User -> Repository** trust is mostly concerned with the quality of data provided to the user. But as recent studies (Prieto 2009; Yakel et al. 2013; Donaldson 2013) have shown, increasingly, the quality of digital preservation operations of the repository is also being considered:

> "Data reusers appear to be noticing repository functions, particularly data processing, metadata and selection, and have expectations about how these should be handled." (Yakel et al. 2013)

The quality of repository operations can be evaluated using the self-assessment and audit methods as described in chapter 2.2 above, but none of these methods currently extends to third party services contracted by the repository.

The Repository's reliance on support from e-Infrastructures should normally not have an impact on the User's trust towards the Repository, as long as the Repository can provide sufficient proof that the digital objects provided for re-use are authentic and reliable. The **User -> e-Infrastructure** trust relationship can be influenced by the User's perception whether storage, processing or other services (e.g. access, authentication) provided by the e-Infrastructure are sufficient to guarantee the delivery of tamper-proof objects to the User.

The **Owner/Funder -> Repository** trust relationship can be individual-organisation type or organisation-organisation type when, for example, the digital repository is a unit within a larger organisation or belongs to a federated structure as described in Scenario 1 in chapter 3. The main trust drivers for owners and funders are customers' (Users, Depositors) satisfaction with the services received from the Repository and the efficiency of operations, including return on investment, of the Repository. Thus the **Owner/Funder -> e-Infrastructure** trust is also conditioned by increased efficiency and cost-efficiency that can be achieved by subcontracting e-Infrastructure providers; as well as increased trust towards the Repository from its other stakeholders (and the concomitant reputation increase).

The **Repository -> e-Infrastructure** trust is inter-organisational and established through governance and fail-safe mechanisms like in subcontracting situations in general through the use of service and operation level agreements, contracts and agreed terms.

## 4.2 ASCERTAINING TRUSTWORTHINESS

As discussed in Chapter 2, repository assessment and audit methods – despite being called 'trusted digital repository' audit methods – are primarily concerned with the quality of operations in repositories and do not include objective benchmarks for establishing trustworthiness or trustedness of repositories. Furthermore, these methods do currently not cater for distributed repository architectures nor for distributed digital preservation services. The organisational viability criteria are designed to assess the ability of an organisation to maintain its own repository and preservation functions. Management of outsourced services through contracts is foreseen but at present requires significant interpretation of the set criteria to cover the five distributed service scenarios described in Chapter 3. Extracting sets of quality criteria from these assessment methods for individual services (e.g. Ingest, Data management) that a repository may want to outsource to an e-Infrastructure provider is possible, but will have to include some organisational and technical criteria as well. To what extent such sub-sets of criteria can engender trust from stakeholders has not been studied.

Mechanisms for awarding 'trusted digital repository' status based on the assessment methods are beginning to emerge through formal auditing and certification. They could be applied to the five distributed digital preservation service models described in Chapter 3 as follows.

> Scenario 1. Federated repositories need to and can trust each other using the 'circular trust model' where each member of the network is applying the same (quality) criteria and exercises the same level of transparency of assessment results towards each other. The set of criteria applied equally across the network of repositories will depend on the (trust) expectations of their stakeholder communities – it can be the Data Seal of Approval that all members of CLARIN and CESSDA networks are expected to achieve; or the TRAC certificate of a TDR that the LOCKSS service is currently aspiring towards.

Scenarios 2 and 3. Repositories extending their storage function to e-Infrastructures can be assessed using the Technical infrastructure and security risk management sections of ISO 16362, DIN 31644 or TRAC. However, the trust requirements could equally well be catered for by the information security controls as set out in ISO 27001.

Scenario 4. Individual criteria could be extracted from the Digital object management sections of ISO 16362, DIN 31644 or TRAC to cover the individual preservation functions that the repository is outsourcing to e-Infrastructure providers, but without the organisational governance aspects included, these criteria will have little impact on the trust from repository's stakeholders.

Scenario 5. When the repository is hosted in an e-Infrastructure, the full set of TDR audit criteria could be applied and TDR status could be sought through formal certification.

In the absence of other methods of establishing trustworthiness and the ill-suited TDR assessment methods that expect the third party service provider (e-Infrastructure) to fully meet the exact same requirements as the digital repository itself, other possibilities must be considered. Most e-Infrastructure service providers have no ambition to become certified as trusted digital repositories or even to act as repositories solely for the DCH sector. The next chapter will explore risk as an alternative method for creating conditions for trust in the distributed digital preservation architecture.

# 5. RISK ASSESSMENT AS A FORM OF ESTABLISHING TRUST

Situations of trust have come to be described as "a subclass of situations involving risk". They are situations in which the risk one takes depends on the performance of another actor (Coleman 1990, p. 91). According to this formulation, trust is warranted when the expected gain from placing oneself at risk to another is positive. Indeed, the decision to accept such a risk is taken to imply trust (Williamson 2006, p. 55). The DRAMBORA risk assessment method for digital repositories[37] is relying on this very concept that awareness of threats to repository systems and operations, and the ability to deal with the expressed risks are the basis for the claim of being a trustworthy digital repository. DRAMBORA describes a formalised process that assists repositories in establishing a comprehensive self-awareness of their objectives, activities and assets before identifying, assessing and managing the risks implicit within their organisation. The assessment report is essentially a risk register, presented in ten categories that helps communicating the problem areas to the repository staff and management, but also supports the trust decision-making for partners and external stakeholders who need to be able to estimate the risks they are taking when entrusting the repository.

First examples of developing domain-specific repository risk profiles have started to appear (Ross et al. 2008; OCLC 2010) and have the potential of evolving into an ontology of repository attributes (McHugh 2012). As individual classes of repository are increasingly identified and described, their common services and characteristics can be understood and ultimately linked with objective measures of success.

The same approach has been taken to develop risk profiles for third-party cloud services and specifically for outsourcing digital archives services. Early groundwork in this area was done by collaborative working groups like the European Union Agency for Network and Information Security (ENISA)[38] on risk management (ENISA 2009); the UK and Ireland Archives and Records Association (ARA) study on storing information in the cloud (ARA 2010a and 2010b); Cloud Sweden's recommendations on outsourcing preservation services to cloud providers (Cloud Sweden 2011); the US National Institute of Standards and Technology (NISO 2011). The results of these teams have led to certification frameworks

[37] http://www.repositoryaudit.eu/
[38] http://www.enisa.europa.eu/activities/risk-management

like the Cloud Security Alliance Security, the Trust & Assurance Registry (STAR),[39] and systematic studies of risks around outsourcing digital preservation services to the cloud (Aitken et al. 2012).

These works demonstrate that risk has proved itself as a useful and universally understood concept when communicating trustworthiness factors and that the current state of art provides better tools for creating a risk profile for distributed digital preservation services than a practical yet formal trust model. This has been summarised well in a recent NDSA report (NDSA 2013, p. 20):

> "The reliability, design, and behaviour of both centralized and distributed preservation networks is just beginning to be understood. It is critical to develop robust trust frameworks that address the risks, because institutions need to be able to measure and evaluate and monitor the reliability and trustworthiness of trustworthy repositories, collaborating organizations and third ‑party services (such as cloud computing). Measuring and evaluating the trustworthiness of such organizations and services is a substantial challenge for policy research."

In the absence of a universally accepted trust model for distributed digital preservation architectures, the search for alternatives has led to risk assessment as a method of establishing and communicating trustworthiness of a preservation service. The Digital Repository Assessment Method Based on Risk Assessment (DRAMBORA) has been in active use since 2007 and has proved that risk registries are an effective means of engaging stakeholders and managers of repositories in discussion of trust and sustainability of services. Indeed, risk is viewed by many of these stakeholders as the "other side of the coin" of trust.

The key concerns with outsourcing preservation services to third parties like cloud or e-Infrastructures have roots in different jurisdictions that govern cultural heritage institutions and e-Infrastructure providers, as well as with the general nature of distributed computing. The main areas of risk are related to:

- legal and governance – incompatibility of regulatory frameworks, legal liabilities;
- security – loss of data or service;
- data transfer – bottlenecks due to bandwidth restrictions, entrenchment due to vendor "lock-in".

Addressing these groups of risks is vital for the digital repository for both maintaining its level of service as well as the level of trust it enjoys with its stakeholders.

Appendix 1 includes a risk analysis tool that repositories can use to assess the policy, legal and organisational level risks when negotiating a service contract with an e-Infrastructure or, indeed a cloud service provider.

Security risks that relate primarily to fixity of information, information loss and security, multi-tenancy and shared technology issues in distributed infrastructures, but also to insecure or incomplete data deletion, are well documented in literature and standards (e.g. ISO 27001). Complete risk registers developed for outsourcing preservation service can be consulted (see for example Cloud Sweden 2011 and Cloud Security Alliance) and are not copied into this report.

Data transfer issues and exit strategies can be mitigated in service level agreements between the digital repository and the service provider.

Risks specific to preservation activities can be identified with the help of the DRAMBORA toolkit.[40] Since in a majority of cases the repository will not be outsourcing core digital preservation decision-making to an e-Infrastructure, these risks are not part of the trust-forming issues. Nevertheless, transparency and communicated accountability for digital preservation activities would contribute towards increased trustworthiness of the repository.

[39] https://cloudsecurityalliance.org/star/
[40] http://www.repositoryaudit.eu/

The intended use of the risk analysis tool in Appendix 1 is described below as a use case scenario that cultural heritage institutions (CHI) can modify according to their own specific needs of services that they are outsourcing to e-Infrastructures.

1) CHI conducts a risk analysis of its own operations or the particular service that it is looking to outsource, using the DRAMBORA toolkit, the risk analysis tool presented in this report or a risk profile tailored specifically for the CHI or its service.
2) CHI drafts requirements for the service(s) it plans to outsource and highlights the specific vulnerabilities / risk areas that it considers vital components for its services to continue to be trusted.
3) The e-Infrastructure and CHI jointly analyse risks related to the listed service requirements, agree on risk mitigation measures and how these can be made public (without disclosing technical or business details that may jeopardise the competitive advantage of either or both parties).
4) The resulting risk register is published, reviewed and updated at regular intervals.
5) CHI can additionally conduct a self-assessment using one of the repository assessment methods (DSA, TRAC, DIN 31644, ISO 16363) and include the risk register of outsourced services in the assessment results. Once auditing and formal certification service of digital repositories becomes available, the CHI may consider applying for certification based on the results of the self-assessment.

# 6. TRUST BUILDING – CONCLUSIONS

Digital repositories can be large or small, handle a wide range of materials from cultural heritage, research, government, or business institutions; they have different organisational contexts and operating situations, technical architectures and institutional responsibilities. Defining a common "yardstick" for measuring whether all the different digital repositories could be trusted by an array of different stakeholder groups is a challenge that has at present been resolved through quality criteria for the operations level. The underlying concept of repository audit tools is that a repository is trusted if it can demonstrate its capacity to fulfil its specified functions, and if those specified functions satisfy an agreed set of criteria, most of which are based on a standard model for a repository (i.e. ISO 14721).

Distributed computing platforms – grid, cloud and e-Infrastructure in general – involve a variety of service levels (e.g. IaaS, PaaS, SaaS) and many permutations of combining actual services into sets for different customer groups. Trustworthiness of these services has mostly been expressed through reliability and security that are subjected to risk assessment.

Combining these two worlds to jointly provide preservation-related services to repository customers extends both the preservation service model that, so far, has been centred around a digital repository maintained by a single organisation, and the trust model of the service that also has been developed around a single organisation being in control of the preservation function.

Neither the existing trusted digital repository assessment methods nor the emerging cloud trust and digital trust methods can be applied to this new service model. The former do not cater sufficiently for distributed architectures and outsourcing of core services hitherto confined to the organisation that owns the repository. The latter focus too narrowly on security and performance issues leaving aside the legal, organisational and policy level aspects that repositories' trust is built on.

As a first step to fill the gap in establishing trustworthiness of the joint repository-e-Infrastructure preservation service, this report proposed a short yet practical risk assessment tool focussing on the organisational and legal aspects of service provision. The risk tool can be combined with other, security, performance and preservation risk tools to create a full risk register for the distributed preservation service.

## 6.1 RECOMMENDATIONS FOR THE DCH-RP ROADMAP

Based on the analysis of both theoretical aspects and practical implementations of trust assessment, this report makes the following recommendations for the DCH-RP roadmap:

1) Adapt the use case scenario described in Chapter 5 to be tested and evaluated as a proof of concept in WP5.
2) Continue monitoring the work on distributed digital preservation (DDP) models and the trust models that will emerge as part of this work.
3) Engage the e-Infrastructure community in discussion of organisational, policy, legal and security risks that are specific to digital preservation to develop an understanding of the issues in both domains and guidelines for managing the identified risks.
4) Monitor the emergence of trusted digital repository audit and certification services and once launched, organise an interest group of cultural heritage institutions that are collaborating with e-Infrastructures to pilot the audit methods on distributed preservation architectures.
5) Share cultural heritage institutions' preservation risks and trust concerns with (research) data community that has an only partially overlapping set of their own trust/risk issues, with the aim of arriving at a fuller risk register of preservation risks that the e-Infrastructures can learn to manage when providing digital preservation services.

# B: THE USE OF AUTHENTICATION AND AUTHORISATION INFRASTRUCTURE IN DISTRIBUTED DIGITAL PRESERVATION

## 7. INTRODUCTION TO AUTHENTICATION AND AUTHORISATION

With the increased interest in digital data, it is important that the arts and humanities domains be able to make their content digitally available. The transition to the digital world brings several challenges, one of the biggest being the availability of an e-Infrastructure to share, manage and store the data for this sector.

The goal of the Digital Cultural Heritage (DCH) project is to define a roadmap to eventually implement a federated infrastructure dedicated to supporting "open science" in the arts and the humanities. As part of the roadmap, the DCH-RP project is also exploring how best to harmonise data preservation policies in the DCH sector and to promote collaboration among DCH institutions, e-Infrastructure providers, research groups and private organisations.

Any e-Infrastructure for the DCH community has to offer a range of capabilities that span different fields such as:

- **management of users** (Authentication and Authorisation Infrastructure, AAI) in the form of a service to authenticate and authorise users located in various countries in Europe and beyond;
- **data management** (High Performance facilities) in the form of services for data preservation and curation, guaranteeing the authenticity of data, rights management, etc.
- **data storage** in the form of (storage systems) services to store and share the digital data.

This part of the deliverable offers recommendations on authentication and authorisation best practices that are suitable for the digital cultural heritage and indications on which of the existing e-Infrastructures could satisfy the requirements of this community.

## 8. DATA LIFE CYCLE MANAGEMENT

The process used to digitise existing content has implications for the data life cycle management that is the set of procedures and policies to store and manage the data, particularly in a distributed environment.

The diagram below depicts a possible model for the preservation of digital cultural data.

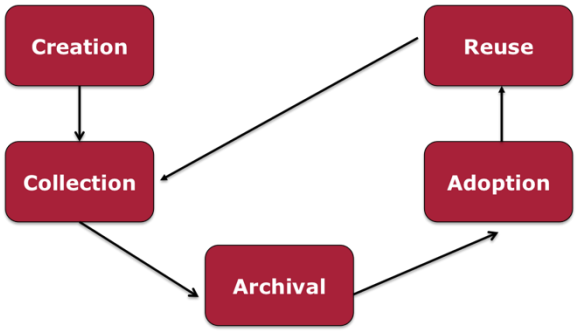


*Figure 11: Data life cycle management*

Although the detailed life-cycle management for digital data is beyond the scope of this document, the way in which data are managed and accessed has implications on the authentication and authorisation requirements, as also highlighted in the AAA Study (AAA 2012):

- As the aim is to make digital data available for future reuse, it is important that data remains authentic, reliable and usable; procedures for ensuring the **provenance**, protecting privacy, confidentiality and intellectual property should be implemented.
- The types of **access** rights associated with data may be different among different participants. These requirements should be specified by the providers that handle these resources providers.
- Users and institutions need to **trust** the infrastructure; policies should enable DCH users located in different countries in Europe and beyond, to access the available services.

Three main elements: provenance (copyrights, authenticity, etc.), access rights and trust are relevant for the AAI, particularly for defining user authentication procedures (social credentials may be accepted, user need to be authenticated face-to-face and so on) and authorisation models (how much does the resource need to know about the users?).

# 9. AAI REQUIREMENTS OF THE DCH COMMUNITY

Before delving into detailed AAI requirements, it is worth looking at the types of service that are available for the DCH community and DCH users.

## 9.1 TYPES OF DCH SERVICE

As emerged from the DCH-RP deliverable D3.1 (DCH-RP 2013), most services used for digital preservation to date are in-house, developed to meet technical requirements of specific institutions, where the digital preservation processes and facilities are operated on their own premises and within user-controlled environment.

This approach makes it difficult to share services among institutions and to reuse potentially useful tools. As highlighted in deliverable D3.1, although there are numerous services and tools that can be used to support automation and preservation tasks, their service description, the types of support offered, technical features and levels of documentation of these tools and services are very heterogeneous. The integration of these services into an existing environment and providing access for users that are outside the organisations hosting the services are therefore rather difficult.

Moving forward with the creation of a DCH infrastructure and gaining benefits from such an infrastructure implies a change in the current way of working for the DCH community. The main changes envisaged should facilitate sharing of key services among different institutions (therefore reducing the cost to operate them), use of storage facilities that are in the cloud or grid (therefore facilitating data sharing and access across institutions) and easy-to-use user interfaces for non-technical users (to facilitate the deployment of services). This requires:

- Establishing trust relationships among the DCH partners via technical frameworks and legal agreements. The usage of external resources may require changing the technical and organisational aspects of the digital preservation process within the DCH institutions themselves.
- Achieving an understanding that to share a service across the e-Infrastructure, someone else will manage the users' working environment, the users themselves and/or the data.
- Access to tools offered via e-Infrastructure (e.g. document analysis tools that are available through an e-Infrastructure, storage facilities) should be possible outside organisational boundaries.

- For the e-Infrastructure to provide added value, access control for users should be managed in a federated fashion, rather than on a per application basis. This will ensure that services and tools are used by a large number of users, will reduce duplication of tools across the partners and will ensure that users' personal data and credentials are stored in one single location (that is the institution that manages the user).

## 9.2 DCH-RP AAI SURVEY

In 2013, TERENA and INFN (with the support of GARR) carried out a survey among the DCH-RP partners. The response to the survey was lower than expected with a total of 25 answers. Although the statistical significance may not be so high, due to the low number of responses, they did highlight the following points:

1. A clear requirement for cross-institutional access: examples provided were access to digital cultural material for research purposes, access to catalogues, access to content for internal projects and so on;
2. More than half of the people who answered the survey are aware of federated access and that R&E federations are operational in many countries;
3. Lack of federated credentials among the respondents, also among those who were aware of federated access;
4. User authentication is required to upload and download the data;
5. The requirements for authorisation were less clear. The services that answered said that in most cases users just need to be authenticated to access the services. Some services maintain access control list mechanisms, but some others say that access should be open;
6. There is a willingness of institutions to be part of federations (for those that are aware of federated access) but there is also a lack of know-how and manpower and in some cases a lack of trust in the infrastructure.

The main requirements for AAI are summarised below:

| Requirements | Technical Implications |
|---|---|
| Support access management policies to ensure that specific content is protected even when Open Access is offered | This requires there to be support for different authentication models for different users. |
| A mechanism for tracking who uploads, changes data | This requires an account mechanism to be in place. It requires authentication of the users and authorisation, to ensure that only specialised users can change and upload data. Download of data could be open (with or without authentication, depending on the data) |
| Easy access for both local and international users | Federated access addresses this use-case |
| Can service trust the way in which users are authenticated by different institutions, possibly not even located in the same country? | Federated access has built in mechanisms to ensure that trust is established among different parties |

| | |
|---|---|
| Can digital cultural institutions trust a service provider that handles the digital cultural data on their behalf, for instance in the cloud? | The service provider should clearly define the terms and conditions of the service |

*Table 1: Summary of AAI requirements*

# 10. FEDERATED ACCESS

## 10.1 KEY CONCEPTS

Methods of accessing services have evolved dramatically in the last decade. One of most important changes relates to the way in which users access applications and how applications manage users.

In the past user access was managed centrally by each application, which meant that users had to register and get application-specific credentials. This model did not work efficiently with the proliferation of applications and with the need from institutions to offer services beyond their organisational borders.

The current best practice is that authentication and authorisation are decoupled from the application:

- Authentication of the users is done by their user **Identity Providers** (i.e. the user's organisation), while
- Authorisation is done by the services (**Service Providers or Relying Parties**) based on the information (identity information) received by the Identity Providers and on the characteristics of the services.

Access to resources that follows this model is known as **Federated Access.** Identity Federations are the infrastructures deployed to enable federated access: these encompass a number of institutions that agree to inter-operate and offer services under a set of well-defined rules.

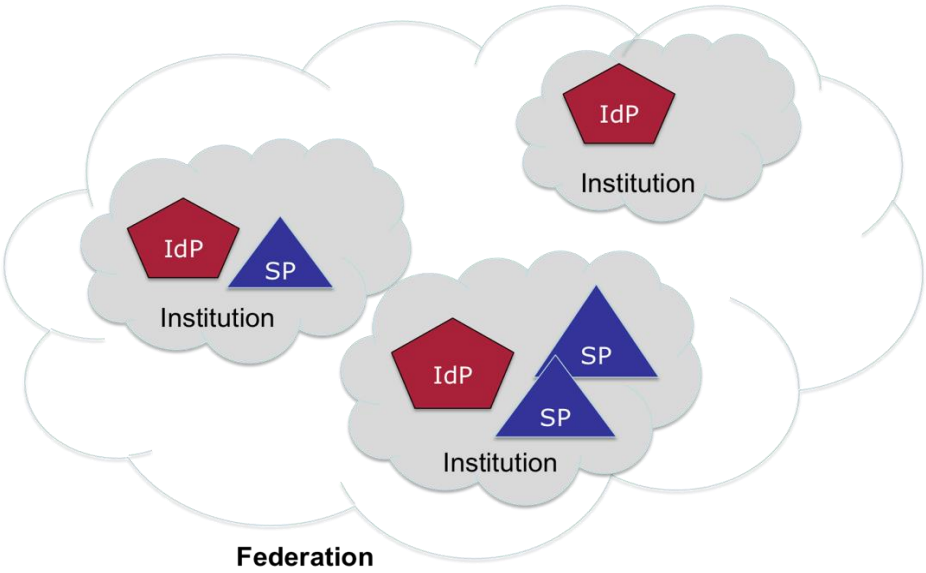


*Figure 12: Identity Federation Model*

The Security Assertion Markup Language, SAML2.0, is an open standard used to build identity federation systems. The SAML protocol supports the secure exchange of authentication and authorisation data between identity providers and service providers or parties relying on these services. Products used to build identity federations include Shibboleth, SimpleSAMLPhP and the Active Directory Federation Service (ADFS).

Federated Access has brought several advantages both for users, who can benefit from a better user experience (less credentials to remember, log in once and access multiple applications, lower risk of forgetting their credentials) and for the service operators, who in practice outsource the user management life-cycle and can focus on authorisation. Federated access also increases security, by using a trusted connection between the identity provider (IdP) and the service provider; this trust connection is built by using standard protocols, legal framework and policies that are shared by the participating entities.

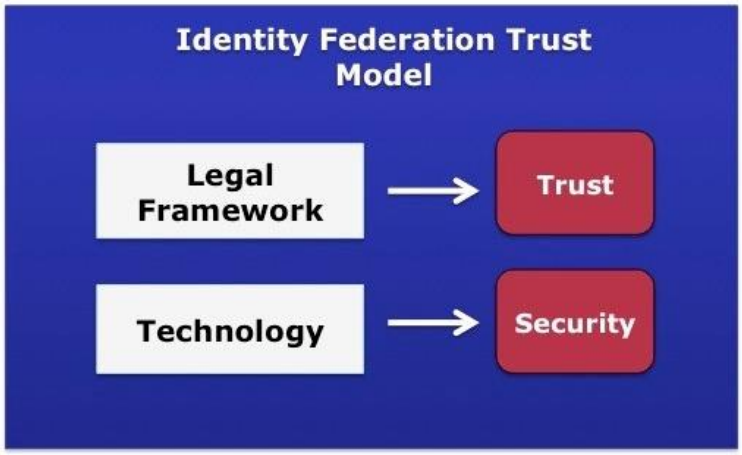


*Figure 13: Trust Model in Federated Access*

Typically R&E federations are operated nationally by the National Research and Education Networks (NRENs) for their community in the country or region concerned. In a federation resources are offered to the participating partners of that federation.

The map below shows where such federations are already available. The coverage of federations in terms of users that have federated credentials, the ways in which services are supported and to a certain extent their policies vary from federation to federation.

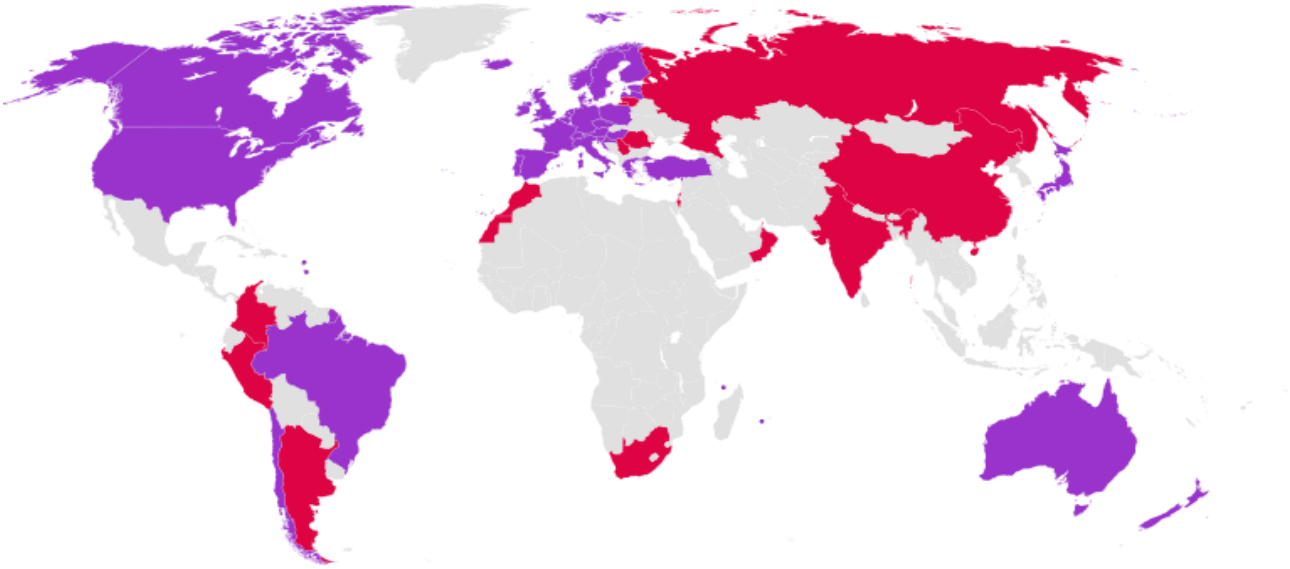

*Figure 14: R&E Identity Federations in the world. For countries shown in purple the federations are already live, whilst for those shown in red federated arrangements are in pilot*

However services can also be offered across borders (which is much more in line with the idea of international collaboration): in this case we talk about **inter-federations**. Inter-federation requires participating federations to agree on a set of policies, to follow Data Protection laws to exchange users' data across countries and a technical infrastructure that enables services and users to 'talk to each other'. Such an infrastructure exists for the R&E community and is called **eduGAIN**; eduGAIN offers support for both the technical exchange of data as well as the exchange of personal information across countries in Europe.

## 10.2 EDUGAIN

eduGAIN[41] is an infrastructure developed in the context of the GÉANT[42] project to enable trustworthy exchange of information for authentication and authorisation purposes among the GÉANT partners and other cooperating parties.

eduGAIN has been designed to address inter-federation and to provide Web Single Sign-On (WebSSO), which enables users to log into multiple services, provided by different federations, using a single, one-step login process. This approach requires an infrastructure that supports the exchange of information between different entities (often located in different countries) and a legal framework[43] (such as a contractual agreement) in line with the Data Protection Directive to ensure that the users' personal data are securely handled.

eduGAIN builds on existing national federations; therefore in order to participate in eduGAIN an existing infrastructure is needed.

There is a strong demand to extend the SSO facility to other applications and service areas. A typical example is a researcher who needs access to Grid-based services and scientific instruments that do not use web browser clients and protocols. Development is on-going in the national federations and in eduGAIN to support these use-cases. The map below shows the R&E federations that are participating in eduGAIN.

[41] http://www.edugain.org/
[42] http://www.geant.net/
[43] http://www.edugain.org/policy

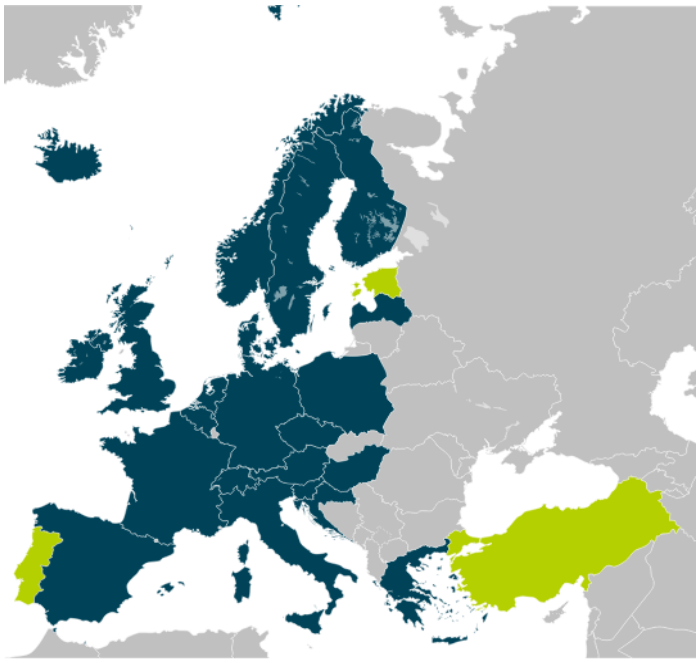

*Figure 15: eduGAIN Status in Europe as of March 2014.*
*The countries in green are in the process of joining eduGAIN.*

## 10.3 E-CULTURE SCIENCE GATEWAY

The e-Culture Science Gateway (eCSG)[44] is a web-portal, provided by INFN, to provide simple access to grid resources for non-technical users.

One of the main obstacles for non-IT-expert users to use e-Infrastructures, such as Grids, is the fact that they are based on complex security mechanisms such as Public Key Infrastructures (PKI) and accessed through low level (command-line based, I.e. non-graphical) user interfaces.

The eCSG solves these problems and makes services available to the largest possible number of users by:

- Presenting users with an easy-to-use web portal;
- Relying on the authentication performed by the users' IdP for users that belong to a federation;
- Allowing users that do not have federated credentials to authenticate use social network.

The eCSG is registered as one of the services of the R&E Italian federation IDEM and through IDEM it is available via eduGAIN; this enables users from other federations participating in eduGAIN to authenticate with eCSG. Being authenticated, however, does not automatically mean that users are authorised to use a resource.

The authorisation of users is done in accordance with the policies of the resources that are available via the eCSG. Users that wish to access a resource available via the portal need to register via the eCSG; if their request is approved, their name is stored in a registry together with the roles for which they then have privileges.

[44] http://ecsg.dch-rp.eu/

# 11. RECOMMENDATIONS FOR AN AAI FOR THE DCH COMMUNITY

To move towards a vision of an infrastructure that offers tools as well as harmonised policies to manage, store and preserve cultural heritage, trustworthiness, ease of use and ease of access for distributed users are key conditions.

A successful future infrastructure will need to be able to manage a variety of access policies where there are legitimate restrictions on data access to protect human privacy, and cultural and cultural heritage, but at the same time it should allow for open access of some of the content.

For the DCH-RP project federated access is a key element, both in terms of using federated storage to handle preservation of cultural heritage data distributed all over Europe and in terms of user management. Federated access is in fact particularly desirable in a situation where services are offered across institutions to users that do not belong to the same institution that offers the service.

## 11.1 HOW CAN FEDERATED ACCESS HELP?

Federated access provides the technical and policy framework to allow for services to be shared in a trustworthy fashion across borders. How authentication is carried out by the institutions and how rights management is carried out by the service provider is left up to the respective parties.

When deciding whether to offer federated access, services should assess their potential user-base: whether they expect many local users or many users coming from different institutions. Federated access cater for the latter use-case and brings the following benefits:

- **Users** will be able to log in once (single sign-in) using their institutional credentials and access multiple services (sign on), Single Sign-On, whilst having the assurance that their personal data will not be disclosed to third parties.
- **Digital cultural curators and cultural institutions participating** will be free of the burden of user name and password administration, and will have access to more tools for managing data. On a large scale of users this means reduced administration and service provisioning costs; and it avoids duplications of identity stores.
- Collaboration among different parties becomes easier.

## 11.2 HOW CAN INSTITUTIONS JOIN A FEDERATION?

You can expect to encounter the following entities when joining a federation:

1. **Identity Providers** (IdPs) – typically organisations that hold information about users and manage user credentials, used to access to resources
2. **Service Providers** (SPs) – publishers, storage services, data management services, blogs, wikis – in fact anyone who wants to provide a 'sign-in' to resources without the hassle of managing user information.
3. **A policy or agreement** – that IdPs and SPs sign up to agree how to interact with each other. These are typically implemented at a national level.
4. **Registration** – a place to sign up and give to a federation information about your IdP or SP - also called your 'entity'.
5. **Metadata** – the collected information about entities, brought together in one place and typically digitally signed by a federation and published to its members.
6. **Discovery service** – a tool used by Service Providers to allow users to select their own Identity Provider.

Institutions in a federated context can act both as IdPs and SPs, or they can only act as either IdPs or SPs.

The first step to join a federation is to talk to the federation operator in a specific country. The list of existing federations is available online at: https://refeds.org/resources/resources_list.html

## 11.3 CONSIDERATIONS ABOUT ENABLING FEDERATED ACCESS FOR DCH

Federated access requires considerable technical expertise to set up the technical infrastructure, whether this is about creating an IdP or an SP. In the context of digital cultural heritage this is potentially a barrier as emerged from the survey. Sadly at the moment there is no software that can be easily installed.

Some federations offer greater support to their users, for instance by installing/operating the IdPs and by offering technical support for non-commercial services. However the cost-recovery model and the availability of manpower make it hard to follow this model for all federations.

As highlighted in the AAA study "To date most NRENs in Europe offer federated access for their users. However, the level of deployment, the participation of institutions and the amount of services available via different federations is in same cases below the desired level" (AAA 2012). The study also recommended Federations to lower the entry level of existing infrastructures for new users and providers and support communities to benefit from existing AAIs.

GARR for instance has developed an elegant way to support small institutions, by offering an "IdP in the Cloud". A virtual machine is offered to each organisation on which the IdP software is installed; the user administrator accesses their own IdP in the cloud via a web interface to manage the users. The infrastructure is hosted in Italy, but it is also available for institutions outside Italy.

> ***Recommendation**: Use a managed service to operate your IdP, whether a commercial offering (such as OpenAthens, Gluu, Ping Identity and equivalent) or one offered by the NRENs (such as GARR's IdP in the Cloud).*

Although institutions and services are free to implement authentication processes as they wish, especially for users with limited technical know-how, authentication based on digital certificates should be avoided. As the grid world has demonstrated there are a number of usability issues related to digital certificates.

> ***Recommendation**: Avoid the usage of digital certificates; if services require a digital certificate (i.e. grid facilities), use solutions like the e-CSG to hide the complexity.*

> ***Recommendation**: The usage of social network identities should not be discarded; there may be applications for which a social network account is sufficient.*

Service providers should design their interface to be easy to use; particularly in the case of federated access it is important to follow accepted best practices to implement federated login in ways that improve user satisfaction and increase successful logins. The REFEDS group has produced guidelines to help login for federated access [Discovery Guidelines].

> ***Recommendation**: Applications should use simple graphic interfaces, rather than command line, to encourage wider usage.*

> ***Recommendation**: Especially if federated access is provided, best practice guidelines should be followed to improve user satisfaction*

The deployment of an e-Infrastructure for the digital cultural heritage domain requires significant investments, even if existing infrastructures are reused. The roadmap should offers inputs on how to engage with national and international decision makers to secure funding. Federated access is recommended although there may be cases (for instance if there are no plans to offer the service widely)

where local access can be a better option. Federated access works well for web-based applications. The technology used to date to support federated access for applications that do not run in a browser is still immature; this should be considered when deciding to provide federated access mechanisms.

> ***Recommendation**: Consider adding a cost/benefit analysis in the roadmap, which also includes considerations around reusing/sharing applications (federated access) versus managing services at an institutional level.*

The roadmap and any follow up project should engage more actively with the R&E federation operators and possibly budget should be allocated for the federation operators for support activities.

> ***Recommendation**: Engage more actively with national R&E federation operators and ensure that funding is allocated to the federations for support activities.*

# REFERENCES

AAA (2012). *Advancing Technologies and Federating Communities: A Study on Authentication and Authorisation Platforms For Scientific Resources in Europe (AAA).* http://www.terena.org/publications/files/2012-AAA-Study-report-final.pdf

Aitken, B., McCann, P., McHugh, A., Miller, K. (2012). *Digital Curation and the Cloud*. Final Report. Produced by the Digital Curation Centre for JISC's Curation in the Cloud Workshop, Hallam Conference Centre http://www.jisc.ac.uk/media/D/0/1/%7BD01C1CDB-AF99-4A20-A9BC-12E73DB224DD%7DCuration-in-the-Cloud.pdf

Ambacher, Bruce (2007). *Government Archives and the Digital Repository Audit Checklist*. In: Journal of Digital Information, Vol. 8 (2) http://journals.tdl.org/jodi/article/view/190/171

APARSEN (2012). D33.1A *Report on Peer Review of Digital Repositories*. APARSEN project http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D33_1B-01-1_1.pdf

ARA (2010a). *Storing Information in the Cloud*. Project Report

ARA (2010b). *Cloud Computing Toolkit. Guidance for outsourcing information storage to the cloud* http://www.archives.org.uk/images/documents/Cloud_Computing_Toolkit-2.pdf

Askhoj, Jan, Sugimoto, Shigeo, Nagamori, Mitsuharu (2011). Preserving records in the cloud. In: *Records Management Journal.* Vol. 21(3), 175-187

Atunes, Gonçalo, Pina, Helder (2011). *Using Grid Federations for Digital Preservation*. Paper presented at the iPRES 2011 conference, Singapore http://timbusproject.net/resources/publications/articles/70-using-grid-federations-for-digital-preservation

Berman, F., McDonald, R.H., Schottlaender, B.E. (2007). *The Need for Formalized Trust in Digital Repository Collaborative Infrastructure*. NSF/JISC Repositories Workshop http://www.sis.pitt.edu/~repwkshop/papers/berman_schottlaender.html

Castelfranchi, Cristiano, Falcone, Rino (2010). *Trust Theory: A Socio-Cognitive and Computational Model.* Wiley

Cloud Sweden (2011). Areas and problems to consider within information security and digital preservation during procurement and use of cloud services.Guidelines. http://cloudsweden.files.wordpress.com/2011/11/cloud_sweden_security-digitalpreservation_v1-1-1_english_final.pdf

Coleman, James (1990). Foundations of Social Theory. Harvard University Press

CPA/RLG (1996). *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information.* RLG

Day, M. (2008). *Toward Distributed Infrastructures for Digital Preservation: The Roles of Collaboration and Trust*. In: The International Journal of Digital Curation. Vol. 3 (1) http://ijdc.net/index.php/ijdc/article/view/60/39

DANS (2009). *Data Seal of Approval* http://www.datasealofapproval.org/

DCC (2010). *What is Digital Curation?* UK Digital Curation Centre http://www.dcc.ac.uk/digital-curation/what-digital-curation

DCH-RP (2013). D3.1 Study on a Roadmap for Preservation http://www.dch-rp.eu/getFile.php?id=114

DIN 31644:2012 *Information und Dokumentation - Kriterien für vertrauenswürdige digitale Langzeitarchive*

Discovery Guidelines http://discovery.refeds.org/

Dobratz, Suzanne, Schoger, Astrid, Strathmann, Stefan (2007). *The nestor Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification*. In: Journal of Digital Information. Vol. 8 (2) http://journals.tdl.org/jodi/article/view/199/180

Donaldson, Devan Ray (2013). Measuring Perceptions of Trustworthiness: A Research Project. Paper presented at the iPRES 2013 conference, Lisbon http://purl.pt/24107/1/iPres2013_PDF/Measuring%20Perceptions%20of%20Trustworthiness%20A%20Research%20Project.pdf

DRI (2013). *Caring for Digital Content: Mapping International Approaches*. Digital Repository of Ireland series, no. 2

e-Culture Science Gateway (e-CSG) http://ecsg.dch-rp.eu

eduGAIN Video http://www.youtube.com/watch?v=x1YhuFPxMz8

ENISA (2009). *Benefits, Risks and Recommendations for Information Security*

EUDAT (2013). D4.3.1: *Trust Establishment Report* http://www.eudat.eu/system/files/EUDAT-DEL-WP4-D4%203-Trust%20Establishment%20Report.pdf

Federated Access Management Video http://www.youtube.com/watch?v=wBHiASr-pwk

Harmsen, Henk (2008). *Data seal of approval - assessment and review of the quality of operations for research data repositories*. In: Proceedings of the iPRES 2008 Conference, British Library

Hofman, Hans, McHugh, Andrew, Ross, Seamus, Ruusalepp, Raivo (2007). *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA).* DPE/DCC http://www.repositoryaudit.eu/

ISO 14721:2003 Space data and information transfer systems -- Open archival information system -- Reference model

ISO 16363:2012 *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*

ISO/DIS 16919 *Space data and information transfer systems - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories*

ISO/IEC 27001:2013 *Information technology – Security techniques – Information security management systems – Requirements*

Lavoie, Brian (2004). The Open Archival Information System Reference Model: Introductory Guide. DPC Technology Watch Series Report 04-01. OCLC/DPC. www.dpconline.org/docs/lavoie_OAIS.pdf

McHugh, Andrew (2012). *A Model for Digital Preservation Repository Risk Relationships*. Paper presented at IFLA2012, Helsinki http://conference.ifla.org/past-wlic/2012/216-mchugh-en.pdf

NDSA 2014 (2013). *National Agenda for Digital Stewardship 2014*. Library of Congress http://www.digitalpreservation.gov/ndsa/documents/2014NationalAgenda.pdf

Nestor (2006). *Criteria for Trusted Digital Long-Term Preservation Repositories* - Version 1 (Request for Public Comment). edited by nestor - Network of Expertise in Long-Term Storage of Digital Resources and nestor Working Group on Trusted Repositories Certification, nestor materials 8 http://nbn-resolving.de/urn:nbn:de:0008-2006060703

nestor (2008). nestor-Kriterien: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. Kompetenznetzwerk Langzeitarchivierung / Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung. Version 2 http://www.nbn-resolving.de?urn:nbn:de:0008-2008021802

NISO (2011). *Cloud Computing Synopsis and Recommendations* of the National Institute of Standards and Technology

OCLC (2010). *Research Libraries, Risk and Systemic Change* OCLC Research http://www.oclc.org/research/publications/library/2010/2010-03.pdf

OCLC/RLG (2007). *Trustworthy Repository Audit and Certification (TRAC): Criteria and Checklist* http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Open Athens http://www.openathens.net/

Prieto, Adolfo G. (2009). From conceptual to perceptual reality: trust in digital repositories. In: *Library Review*. Vol. 58(8), 593-606

REFEDS https://refeds.org/

Riding the Wave (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data*. Final report of the High Level Expert Group on Scientific Data. European Commission http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

RLG/NARA (2005). *An Audit Checklist for the Certification of Trusted Digital Repositories*. Draft for public commenting. RLG http://www.oclc.org/research/activities/past/rlg/repositorycert.htm

RLG/OCLC (2002). *Trusted Digital Repositories: Attributes and Responsibilities.* RLG. http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf

Ross, Seamus, McHugh, Andrew, Innocenti, Perla, Ruusalepp, Raivo (2008). *Investigation of the potential application of the DRAMBORA toolkit in the context of digital libraries to support the assessment of the repository aspects of digital libraries*. HATII/DELOS

Rotter, J. B. (1967). A New Scale for the Measurement of Interpersonal Trust. In: *Journal of Personality*. Vol. 35, 651-665

Ruusalepp, R., Lee, C.A., van der Werf, B., Woollard, M. (2012). Standards Alignment. In: McGovern, N.Y., Skinner, K. (Eds.). Aligning National Approaches to Digital Preservation. pp. 115-166. Atlanta: Educopia Institute Publications

Schultz, Matt, Gore, Emily (2010). *The Importance of Trust in Distributed Digital Preservation: A Case Study from the Metaarchive Cooperative*. Paper presented at iPRES 2010, Vienna http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/schultz-39.pdf

Skinner, Katherine, Schultz, Matt (2010). *A Guide to Distributed Digital Preservation*. Educopia Institute http://metaarchive.org/sites/metaarchive.org/files/GDDP_Educopia.pdf

Skinner, Katherine, Zierau, Eld, Schultz, Matt (2013). Building a Framework for Applying OAIS to Distributed Digital Preservation. Presentation at the Aligning Digital Preservation across Nations Workshop, Amsterdam. http://digitalcurationexchange.org/system/files/skinner-et-al-oais-distributed_0.pdf

Trehub, Aaron, Halbert, Martin (2012). *Safety in Numbers: Distributed Digital Preservation Networks*. Paper presented at IFLA2012, Helsinki http://conference.ifla.org/past-wlic/2012/216-trehub-en.pdf

Walters, T.O, McDonald, R.H. (2008).*Creating Trust Relationships for Distributed Digital Preservation Federations*. Paper presented at iPRES 2008, London. http://www.bl.uk/ipres2008/presentations_day2/31_Walters.pdf

Williamson, Oliver E. (2006). *Calculativeness, Trust, and Economic Organisation*. In: Kramer, Roderick (ed.). Organizational Trust: A Reader. Oxford University Press, 48-81

Wittek, Peter, Darányi, Sándor (2012). Digital Preservation in Grids and Clouds: A Middleware Approach. In: *Journal for Grid Computing*. Vol. 10, 133–149

Yakel, Elizabeth, Faniel, Ixchel, Kriesberg, Adam, Yoon, Ayoung (2013). Trust in Digital Repositories. In: *International Journal of Digital Curation*. Vol. 8(1)
http://www.ijdc.net/index.php/ijdc/article/view/8.1.143/303

Zierau, Eld, Schultz, Matt (2013). *Creating a Framework for Applying OAIS to Distributed Digital Preservation*. Paper presented at the iPRES2013 conference, Lisbon
http://purl.pt/24107/1/iPres2013_PDF/Creating%20a%20Framework%20for%20Applying%20OAIS%20to%20Distributed%20Digital%20Preservation.pdf

# APPENDIX 1. POLICY, ORGANISATIONAL AND LEGAL RISKS IN A DISTRIBUTED DIGITAL PRESERVATION SERVICE

## Policy risks

| | |
|---|---|
| **Risk Identifier:** | R01 |
| **Risk Name:** | Vendor lock-in |
| **Risk Description:** | The organization becomes dependent on the services offered by the service provider, or is unable to change to another service provider without high switching-costs or losing assets. |
| **Example Risk Manifestation(s):** | Lack of standard technologies among the service providers to allow data portability (APIs, formats, procedures…).<br>Difficulties on migration from one provider to another or to in-house services (portability and interoperability issues).<br>Increase of data lock-in at the same rate as the amount of data stored in the e-Infrastructure if portability is not provided. |
| **Mitigation strategies:** | Negotiation of exit-strategy with the service provider.<br>Selection of provider with the most suitable options regarding interoperability with the organization.<br>Use of open standards, whenever applicable. |

| | |
|---|---|
| **Risk Identifier:** | R02 |
| **Risk Name:** | Loss of governance |
| **Risk Description:** | The organization cedes to the e-Infrastructure provider governing responsibilities over a number of issues concerning the assets stored in the e-Infrastructure. |
| **Example Risk Manifestation(s):** | The organization transfers the responsibility for issues affecting security to the e-Infrastructure provider and no reports/logs are shared with them, making it impossible to audit or control the assets.<br>Security procedures of provider are unknown, not agreed upon or are not aligned with the organizational ones.<br>Compliance challenges with regulatory or legal environment due to the lack of guarantees on the authenticity, integrity and reliability of information stored. |
| **Mitigation strategies:** | Ensure that Service Level Agreements (SLA), contracts or any other agreements are complete and clarify roles and responsibilities of each of the parties.<br>Reservation of rights by the e-Infrastructure provider should be analysed in detail and restrained when necessary.<br>Clarify the potential provision of services by third-parties and their compliance with the guarantees provided by the service provider. |

| | |
|---|---|
| **Risk Identifier:** | R03 |
| **Risk Name:** | Loss of ownership |
| **Risk Description:** | Organization cedes ownership of digital assets or related information to the e-Infrastructure provider. |
| **Example Risk Manifestation(s):** | Service provider takes control of assets due to a lack of transparency on the agreements.<br>Use of transactional and relationship information collected by the e-Infrastructure provider that might be revealing or commercially valuable. |
| **Mitigation strategies:** | Clear terms of contract and service, including statements on the ownership of the assets.<br>Clear roles and responsibilities in the contract. |

| | |
|---|---|
| **Risk Identifier:** | R04 |
| **Risk Name:** | Non-compliance with certification and accreditation requirements |
| **Risk Description:** | Stakeholders are not able to meet confirmation of the characteristics required to fulfil such certification and accreditation frameworks. |
| **Example Risk Manifestation(s):** | Standards not adapted to the use and characteristics of the e-Infrastructure infrastructures, thus there might be an impact on compliance or certification to them.<br>Control on the location of the data could be mandatory to fulfil the standards' requirements for accreditation, and in some cases it is unknown by the organization.<br>E-Infrastructure provider is not certified by standards that could increase the assurances on information security to the organization.<br>E-Infrastructure provider does not allow the organization to audit their processes. |
| **Mitigation strategies:** | Selection of e-Infrastructure providers accredited by relevant certification schemes.<br>Negotiation with e-Infrastructure provider on the requirements on communication, reporting and audit to ensure trustworthiness on their procedures and compliance with the SLA.<br>Selection of e-Infrastructure providers that allow selection of the jurisdictional areas to allocate the organizational assets. |

| | |
|---|---|
| **Risk Identifier:** | R05 |
| **Risk Name:** | Loss of service level or availability |
| **Risk Description:** | The e-Infrastructure provider fails in providing availability of the service or its quality levels are compromised. |
| **Example Risk Manifestation(s):** | E-Infrastructure provider does not reach levels of availability specified on SLA.<br>Service credit or insurance does not compensate for loss of service, income and reputation.<br>Planned downtimes are not included in terms of lack of availability.<br>Organization cannot control / measure availability levels and communication procedures / reporting not established.<br>Data loss and inaccessibility.<br>Business continuity and data recovery plans are not ensured.<br>Scalability expectations (either up or down) not met in a timely fashion by service provider.<br>Failure on backups leading to data loss. |
| **Mitigation strategies:** | Make sure that SLA is detailed enough on the availability of the service and possible downtimes.<br>Assess whether the compensation measures for downtimes adequately compensates for impact for the potential loss of service.<br>Establish procedures to get timely communication and reporting from the service provider.<br>Ensure that adequate plans for business continuity, data and disaster recovery or incident response are in place.<br>Understand resource provisioning policies and procedures of e-Infrastructure provider (resource overload, scalability, etc.) and assess whether it fits with organizational needs.<br>Agree upon a protocol for system updates and other planned maintenance activities to minimize impact on service. |

| | |
|---|---|
| **Risk Identifier:** | R06 |
| **Risk Name:** | Non-compliance with existing information management and preservation policies and procedures |
| **Risk Description:** | Organization policies and procedures to manage their digital assets throughout their life-cycles are not aligned with the requirements of using e-Infrastructure |

| | |
|---|---|
| | technologies. |
| **Example Risk Manifestation(s):** | Non-alignment with the OAIS reference model causes difficulties on transferring responsibility to an external party of some processes.<br>Preservation tasks not offered by the e-Infrastructure provider.<br>Removal actions are not transparent or appropriately carried out. There are severe technical difficulties around the destruction of records in the e-Infrastructure and its verification.<br>Disposal of public records not achieved as specified by the organization's retention and disposal schedule (multiplicity of locations, backups…).<br>Difficulties to put in practice retention policies for the assets stored in the e-Infrastructure.<br>Metadata mismanagement results in portability becoming possible.<br>Diminished level of metadata quality, losing usability for preservation.<br>Loss of control on provenance of the data.<br>Loss of integrity and authenticity of the data.<br>Lack of transparency on data migration and transformation actions. |
| **Mitigation strategies:** | Keep track of backup copies or any action that the service provider takes on the stored assets.<br>Detail policies and procedures including all aspects that might be affected by transferring parts or the whole workflow to the e-Infrastructure.<br>Seek compliance from the e-Infrastructure provider with policies and procedures to ensure bit and logical preservation.<br>Ensure complete record and control over the processes on the chain of preservation, to assure authenticity and reliability. |

| | |
|---|---|
| **Risk Identifier:** | R07 |
| **Risk Name:** | Difficulties in monitoring, auditing and reporting e-Infrastructure services |
| **Risk Description:** | E-Infrastructure provider does not allow the organization to monitor the service to check compliance with SLA in aspects such as information security and performance measurements. There are no guarantees that SLA is being fully accomplished or that the quality of service is adequate. |
| **Example Risk Manifestation(s):** | No access to logs provided to the organization.<br>E-Infrastructure provider does not provide standard audit documentation and reports.<br>E-Infrastructure provider is not audited internally or by external bodies.<br>E-Infrastructure provider does not provide the appropriate tools to the organization to allow auditing of performance.<br>SLA does not reflect in which ways performance can be measured and controlled. |
| **Mitigation strategies:** | Establish requirements on audit and reporting, and how these should be carried out by the service provider.<br>Ensure the correct level of readiness in the organization to fulfil these new tasks. |

| | |
|---|---|
| **Risk Identifier:** | R08 |
| **Risk Name:** | Non-compliance with organization’s security policy |
| **Risk Description:** | Security policies and procedures of the e-Infrastructure provider are not aligned with the organization’s own policies and fail to fulfil its needs. |
| **Example Risk Manifestation(s):** | Organization’s security policy has to be adapted to be aligned to those of the e-Infrastructure provider.<br>Access restrictions are not under the organization's control, and there is no assurance on unauthorised access. |
| **Mitigation strategies:** | Reach agreements with the service provider on particular conditions in SLA, contracts, terms of service.<br>Clarify roles and responsibilities of each party to avoid security areas not being covered. |

| | |
|---|---|
| | Establish adequate communication and reporting protocols that the provider must comply with.<br>Understand and agree upon authentication and access management policies to be carried out by the e-Infrastructure provider. |

| | |
|---|---|
| **Risk Identifier:** | R09 |
| **Risk Name:** | Limitation of liabilities on Service Level Agreements |
| **Risk Description:** | The definition of the levels of service includes limitations on the responsibility the service provider holds. |
| **Example Risk Manifestation(s):** | Unclear definition of roles and responsibilities in the agreements between organization and e-Infrastructure provider.<br>Low level of transferability of liability to e-Infrastructure provider. Legal and reputational implications faced by the organization. |
| **Mitigation strategies:** | Clarify roles and responsibilities of both parties on agreements.<br>Identify cases with no responsibility or obligation to compensation and assess whether the model fits with the organization's purposes. |

| | |
|---|---|
| **Risk Identifier:** | R10 |
| **Risk Name:** | Organization fails to revise its own policies and procedures |
| **Risk Description:** | Rationale and/or business activities and processes are not adapted to the new architecture of the service, leading to inefficiencies or contradictions. |
| **Example Risk Manifestation(s):** | New workflows not included in the organization's procedures.<br>Organization's security policy not updated or aligned with that of the e-Infrastructure provider. |
| **Mitigation strategies:** | Assess own policies and procedures and revise them according to the agreements reached with the service provider.<br>Adjust roles and responsibilities in the organization. |

## Organizational risks

| | |
|---|---|
| **Risk Identifier:** | R11 |
| **Risk Name:** | Lack of sustainability related to financial resources |
| **Risk Description:** | The organization dismisses cost implications of e-Infrastructure services in the long-term. |
| **Example Risk Manifestation(s):** | Organization does not own the resource, which implies on-going payment for the e-Infrastructure provider services due to usage-based pricing.<br>Cost implications of regular accesses or processes in the e-Infrastructure not considered.<br>Increments in the needs for bandwidth or storage significantly vary the costs.<br>Increase of computational expenses due to new operations (e.g. data/text mining) not affordable for the organization.<br>Additional costs might arise: hidden-costs, extraction process related costs, licensing costs, metadata updates, etc. |
| **Mitigation strategies:** | Clarify all possible additional costs and likeliness of increases.<br>Seek guarantees on the ability to switch between vendors, avoiding lock-in.<br>Ensure the level of budget.<br>Request additional funding or revise objectives when this is not possible.<br>Maintain contingency fund.<br>Review funding strategy. |

| | |
|---|---|
| **Risk Identifier:** | R12 |
| **Risk Name:** | Loss of business or service reputation |
| **Risk Description:** | Organization's stakeholders change their opinion about and lose confidence and trust in the service provided by the organization. |
| **Example Risk Manifestation(s):** | Lack of reputational isolation leads to a contagious effect due to negative activities on the part of co-tenants.<br>Lack of resource isolation in physical resources shared by multiple customers allowing unauthorized access or manipulation.<br>Negative stakeholders' perceptions towards the use of the e-Infrastructure to store data with privacy implications.<br>Lack of transparency on the use of cross-organizational authentication systems and perception of privacy infringing on end-users. |
| **Mitigation strategies:** | Comply with all relevant certification schemes.<br>Increase transparency towards end-users on the use of their personal data.<br>Ensure that possible vulnerabilities (e.g. hypervisor security model) are under control by the e-Infrastructure provider. |

| | |
|---|---|
| **Risk Identifier:** | R13 |
| **Risk Name:** | Role changes of organization's staff |
| **Risk Description:** | The use of e-Infrastructure computing requires different capabilities and modifications in the roles played by the organization's staff. |
| **Example Risk Manifestation(s):** | Management and maintenance tasks might differ or increase, if there is a need to manage and secure the operating system, applications and virtual instances.<br>Organization has to monitor e-Infrastructure services to check performance of SLA.<br>Difficulties for the staff to transition to an e-Infrastructure service. |

| | |
|---|---|
| **Mitigation strategies:** | Define new roles and profiles according to the new workflow/tasks.<br>Monitor performance and adapt plans after assessment.<br>Implement a training plan for the staff to improve competences and raise awareness on issues concerning the new system. |

| | |
|---|---|
| **Risk Identifier:** | R14 |
| **Risk Name:** | Staff skills become obsolete |
| **Risk Description:** | The introduction of new roles brings up the need for a whole new set of skills. |
| **Example Risk Manifestation(s):** | No training plans have been established before/after the introduction of the new systems. |
| **Mitigation strategies:** | Determine the organizational needs to address the new tasks and assess whether the staff members need additional training or there is a need for new members of staff.<br>Review performance regularly and implement training plans accordingly. |

| | |
|---|---|
| **Risk Identifier:** | R15 |
| **Risk Name:** | Resistance to change in the organization |
| **Risk Description:** | Perceptions of organization's staff towards the use of e-Infrastructure technologies do not contribute to the acceptance of the new model. |
| **Example Risk Manifestation(s):** | The change process is not well understood or followed within the organization.<br>Difficulties in implementation and failures in usability of the new systems. |
| **Mitigation strategies:** | Assess new organizational needs and identify staff expectations and experiences.<br>Establish change management plan. |

| | |
|---|---|
| **Risk Identifier:** | R16 |
| **Risk Name:** | Management failure |
| **Risk Description:** | Organizational management shortcomings produce a failure on the achievement of its objectives. |
| **Example Risk Manifestation(s):** | Insufficient allocation of resources considering the cost-models used by e-Infrastructure providers.<br>Organization does not have a business continuity plan to mitigate effects of a crisis involving critical processes or assets. |
| **Mitigation strategies:** | Design and adequate the policies and procedures according to the changes in the organization and establish review mechanisms.<br>Establish business continuity plans or any other mechanisms to mitigate and overcome the failure. |

| | |
|---|---|
| **Risk Identifier:** | R17 |
| **Risk Name:** | Business objectives not met |
| **Risk Description:** | Organization fails totally or partially to achieve the foreseen outcomes. |
| **Example Risk Manifestation(s):** | Preservation of the assets is not adequately achieved due to poor performance of operations in the e-Infrastructure.<br>Difficulties to prove authenticity and integrity of information preserved.<br>Personal data leakage/disclosure to third parties. |
| **Mitigation strategies:** | Ensure compliance with organizational policies and procedures.<br>Monitor and review service provider performance.<br>Review preservation policies and procedures.<br>Monitor business objectives and redefine them whenever necessary. |

| | |
|---|---|
| **Risk Identifier:** | R18 |
| **Risk Name:** | Enforced cessation of organization's operations |
| **Risk Description:** | Impossible to continue organization's activities. |
| **Example Risk Manifestation(s):** | Bankruptcy of service provider without an adequate strategy leads to the loss of the assets stored in the e-Infrastructure.<br>Technical failure affects the stored data causing an irreparable loss of the main digital assets.<br>Failure in outsourced critical business process.<br>Organization lacks succession plan for its digital assets. |
| **Mitigation strategies:** | Establish succession plans.<br>Establish exit strategy.<br>Establish policies and procedures ensuring security of assets. |

| | |
|---|---|
| **Risk Identifier:** | R19 |
| **Risk Name:** | Inability to evaluate organization's success |
| **Risk Description:** | Organization is not able to determine whether its objectives have been achieved or not and to what extent. |
| **Example Risk Manifestation(s):** | E-Infrastructure provider lacks transparency and does not provide the organization with sufficient information through audit reports about the state of the stored assets.<br>Organization has no mechanisms in place to monitor the performance of the e-Infrastructure provider.<br>Organization does not engage with stakeholders to determine satisfaction levels. |
| **Mitigation strategies:** | Establish means of assessment of both internal and external actions.<br>Use external certification to determine the degree of competence. |

| | |
|---|---|
| **Risk Identifier:** | R20 |
| **Risk Name:** | Difficulties in negotiating contracts and terms of service |
| **Risk Description:** | Organization lacks the ability to negotiate agreements with e-Infrastructure provider. |
| **Example Risk Manifestation(s):** | Organization does not have access to a legal counsellor able to determine the most suitable conditions for the organization.<br>E-Infrastructure provider has standard contract and additional agreements that cannot be negotiated.<br>Organization does not conduct due diligence assessment before entering into a contract. |
| **Mitigation strategies:** | Seek legal advice to give support on the negotiation of contracts and other agreements. |

## Legal risks

| | |
|---|---|
| **Risk Identifier:** | R21 |
| **Risk Name:** | Location and jurisdictional implications |
| **Risk Description:** | Location of the e-Infrastructure resource unknown, established in a different jurisdictional area to that where the organization is located. |
| **Example Risk Manifestation(s):** | Legal practices and regulations differ from those in the organization's jurisdictional area.<br>E-Infrastructure provider does not give the organization choices on the location of information.<br>E-Infrastructure provider does not provide timely information about changes of location of the e-Infrastructure resource. |
| **Mitigation strategies:** | Establish agreement with the e-Infrastructure provider about the jurisdictional areas where organization's assets can be stored.<br>Request notification on proposed changes of location.<br>Have control over the regulations of the jurisdictional areas that can affect organization's assets.<br>If possible, reach contractual agreement on the court and applicable law in case of eventual legal dispute. |

| | |
|---|---|
| **Risk Identifier:** | R22 |
| **Risk Name:** | Non-compliance with data protection laws |
| **Risk Description:** | Breach of regulatory requirements of protected data such as those containing personal or sensitive information. |
| **Example Risk Manifestation(s):** | Sharing protected information with e-Infrastructure providers might be non-compliant with privacy laws.<br>Records management and disposal laws may introduce limitations on the ability of government agencies to share information with e-Infrastructure providers.<br>Data stored in the e-Infrastructure is accessed by unauthorised people, intercepted or leaked to the public.<br>Breach of the limits on privacy set up by regulations on using personal information in a cross-organisational setting for the purpose of identity management.<br>Mismanagement of encryption keys leads to the loss of confidentiality of the information stored. |
| **Mitigation strategies:** | Ensure compliance through formal agreements with the e-Infrastructure provider and get assurance of its levels of liability for unlawful actions.<br>Select, on the role of controller of personal data, a processor with adequate guarantees on security measures.<br>Request e-Infrastructure provider's assurance on reporting on any data processing that they carry out<br>Organization should be informed on data security activities and the data controls e-Infrastructure provider has in place.<br>Get guarantees of a robust system for authentication, authorization and accounting.<br>Establish a hybrid model with highly sensitive data stored in a private e-Infrastructure.<br>Carry out an appropriate Privacy Impact Assessment (PIA) before entering into agreements with e-Infrastructure provider. |

| | |
|---|---|
| **Risk Identifier:** | R23 |
| **Risk Name:** | Non-compliance with IPR regulations |

| | |
|---|---|
| **Risk Description:** | Breach of regulatory requirements of copyright, patent infringement or other IPR-related misdemeanour. |
| **Example Risk Manifestation(s):** | The organization fails on fulfilling IPR laws requirements by using e-Infrastructure technologies to store protected materials.<br>The organization does not properly manage rights and restrictions of protected materials stored in the e-Infrastructure.<br>Information is not properly classified according to rights and restrictions.<br>Metadata mismanagement causes the loss of rights metadata and thus the lack of sufficient contextual information to identify the level of protection.<br>Actions taken for digital preservation of assets are not-compliant with IPR regulations. |
| **Mitigation strategies:** | Assess whether the assets stored in the e-Infrastructure are subject of IPR restrictions.<br>Seek legal advice to determine legality of the activity.<br>Establish the conditions and ensure e-Infrastructure provider's compliance with organization's requirements, without diminishing the quality of the service.<br>Establish and monitor agreements with the rights-holders, when necessary. |

| | |
|---|---|
| **Risk Identifier:** | R24 |
| **Risk Name:** | E-Infrastructure Provider disclosure obligations implications |
| **Risk Description:** | Legal requests enforce the e-Infrastructure provider to give access to the information under their supervision. |
| **Example Risk Manifestation(s):** | E-Infrastructure provider might be obliged to examine user records to find evidence of irregular activities. |
| **Mitigation strategies:** | Require from the e-Infrastructure provider information about procedures and conditions for disclosure and timely notification for any requested disclosure.<br>Demand guarantees on security of organization's data when co-tenants are subject of disclosure. |

| | |
|---|---|
| **Risk Identifier:** | R25 |
| **Risk Name:** | Unintentional disclosure in multi-tenant environments |
| **Risk Description:** | E-Infrastructure resource is shared by multiple tenants and isolation failure might allow third parties to access to organization's data. |
| **Example Risk Manifestation(s):** | Physical drives are shared with other tenants that are involved in a legal case and whose information disclosure is enforced. |
| **Mitigation strategies:** | Request guarantees on the full isolation of resource, even not sharing physical machines in the case of critical data.<br>Ensure sufficient levels of encryption and reliable key management. |

| | |
|---|---|
| **Risk Identifier:** | R26 |
| **Risk Name:** | Inadequacy of regulations and legislation to e-Infrastructure |
| **Risk Description:** | Regulations affecting the organization’s assets in the e-Infrastructure do not contemplate the challenges imposed by the use of e-Infrastructure technologies. |
| **Example Risk Manifestation(s):** | Contractual relationships are the only ones filling gaps within the regulation framework.<br>Impossibility of compliance and possible liability for infringement of regulations. |
| **Mitigation strategies:** | Seek legal advice to avoid non-compliance with legal framework. |

| | |
|---|---|
| **Risk Identifier:** | R27 |

| | |
|---|---|
| **Risk Name:** | Liability for infringement of legal requirements and regulations |
| **Risk Description:** | Organization is legally accountable for not fulfilling responsibilities or acting beyond the scope of what is allowed on the basis of legal and regulatory instruments. |
| **Example Risk Manifestation(s):** | Organization has to face legal consequences of the infringement of the laws protecting information even in the event of e-Infrastructure provider actions.<br>No clear delineation of liability has been set up between parties. |
| **Mitigation strategies:** | Monitor legal framework to ensure compliance of the actions, procedures, policies, agreements, etc.<br>Seek legal advice to determine legality of activities with respect to legislation.<br>Establish policies and procedures to follow in the event of legal challenge. |

| | |
|---|---|
| **Risk Identifier:** | R28 |
| **Risk Name:** | Exit-strategy deficient or not defined by the organization |
| **Risk Description:** | Lack of assurance that contractual, technological or planning resources are in place to move out or replace e-Infrastructure computing services. |
| **Example Risk Manifestation(s):** | The e-Infrastructure provider does not offer a standardised export procedure for information and the organisation needs to develop its own programme to extract its information.<br>Information retrieval requires a change in format, with possible consequences for authenticity, reliability or legal admissibility. |
| **Mitigation strategies:** | Analyse and document all the procedures and properties that are critical for the assets stored in the e-Infrastructure.<br>Define and establish an exit-strategy according to them. |

| | |
|---|---|
| **Risk Identifier:** | R29 |
| **Risk Name:** | Acquisition of e-Infrastructure provider |
| **Risk Description:** | The ownership of the service provider changes and operations and assets are transferred. |
| **Example Risk Manifestation(s):** | Policies, procedures and terms of service might change.<br>Unknown accountability or affiliations of new e-Infrastructure provider. |
| **Mitigation strategies:** | Include guarantees in the contract to keep the conditions agreed upon in the event of changes in provider ownership.<br>Establish exit-strategy. |

| | |
|---|---|
| **Risk Identifier:** | R30 |
| **Risk Name:** | E-Infrastructure provider ceases business |
| **Risk Description:** | The service provider goes out of business and ceases operations. |
| **Example Risk Manifestation(s):** | Limitations on retrieving data in the event that the provider ends business operations with little or no warning.<br>Difficulties in information and metadata portability.<br>There is no business-continuity strategy established by the cloud provider. |
| **Mitigation strategies:** | Conduct due diligence to get assurance on cloud provider accountability, maturity, viability, etc. before entering into a contract.<br>Establish exit-strategy. |

| | |
|---|---|
| **Risk Identifier:** | R31 |
| **Risk Name:** | Subcontract to third-parties by e-Infrastructure provider |

| | |
|---|---|
| **Risk Description:** | Third-party subcontractors provide the e-Infrastructure provider with parts of the service or infrastructure for the deployment of the service to the customer. |
| **Example Risk Manifestation(s):** | Third-parties subcontractors have different policies and procedures.<br>Subcontractor does not offer the same guarantees of service availability.<br>Subcontractor established in different jurisdictional area. |
| **Mitigation strategies:** | Acknowledge which services are subcontracted by third parties and establish the necessary procedures.<br>Get assurance that SLA with third parties does not diminish levels of service.<br>Ensure that their third party performance levels and security compliance are monitored by the e-Infrastructure provider.<br>Ensure that all organizational requirements are met in case of contracting services through an e-Infrastructure services broker. |

| | |
|---|---|
| **Risk Identifier:** | R32 |
| **Risk Name:** | E-Infrastructure provider's reservation of rights |
| **Risk Description:** | E-Infrastructure provider reserves certain rights on the use of customer's assets under their supervision. |
| **Example Risk Manifestation(s):** | E-Infrastructure provider changes its terms and policies unilaterally.<br>Secondary use of customer information by the e-Infrastructure provider. |
| **Mitigation strategies:** | Ensure transparency on agreements.<br>Establish the conditions that the e-Infrastructure provider should comply with to avoid unlawful actions. |

| | |
|---|---|
| **Risk Identifier:** | R33 |
| **Risk Name:** | Evidential value of information diminished |
| **Risk Description:** | It is not possible to prove authenticity and integrity of records stored in outsourced e-Infrastructure facilities. |
| **Example Risk Manifestation(s):** | No reliability in e-Infrastructure provider procedures on migration processes, backups, etc.<br>Information on security policies of e-Infrastructure provider is not documented or accessible.<br>E-Infrastructure provider does not update organization about issues concerning corruption, loss or data changes. |
| **Mitigation strategies:** | Monitor actions taken by the e-Infrastructure provider on the data stored in the e-Infrastructure.<br>Require transparency in e-Infrastructure provider's policies and procedures. |

| | |
|---|---|
| **Risk Identifier:** | R34 |
| **Risk Name:** | Liability for breach of contractual or licensing relationships |
| **Risk Description:** | Organization is legally accountable for not fulfilling responsibilities or acting beyond the scope of what is allowed in contractual relationships with stakeholders. |
| **Example Risk Manifestation(s):** | Protected materials are stored in a third-party storage facility without the consent of the right holders. |
| **Mitigation strategies:** | Monitor contractual relationships to ensure their terms are corresponded.<br>Seek legal advice to ensure no breaches of contractual relationships.<br>Establish policies and procedures to follow in the event of contractual challenge. |