TALLINNA ÜLIKOOL

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

UNIVERSITÀ DEGLI STUDI DI PARMA

Eva Montenegro Piñeiro

Organizational and Regulatory Risks of Distributed Digital Preservation

Supervisor: Raivo Ruusalepp

Master thesis
International Master in Digital Library Learning
2014

Abstract

The main aim of this study is to investigate the major organizational and regulatory risks in the context of distributed digital preservation in libraries and archives in particular, which will be referred to as memory institutions throughout this report. Furthermore, the study also aims to explore how the inter-organizational relationships emerged in the context of collaborative arrangements and with the use of third-party services are supported in terms of trust, considering the levels of vulnerability and threats to which the organizations involved may be exposed to.

Keywords: risk, distributed digital preservation, inter-organizational trust.

Table of contents

Acknowledgments

My sincere appreciation goes to all the participants on the data collection exercise. Without their time and generosity sharing their knowledge, this research would not be possible.

My sincere gratitude to my supervisor Raivo Ruusalepp, thanks to your *risk appetite* and all the new opportunities and inexhaustible support.

My heartiest gratitude to my professors in Tallinn University Sirje Virkus and Aira Lepik, for all your support, teachings and friendship.

My gratitude to all my DILL friends, classmates and professors for this adventure and to my colleagues at the National Library of Estonia for keeping me company and good conversations.

My love goes to my family, friends and various pen-friends. To those able to resist the time and distance, and to those who remained along the way.

Abbreviations

4C	Collaboration to Clarify the Costs of Curation
ALA	American Library Association
ALCTS	Association for Library Collections & Technical Services
APARSEN	Alliance for Permanent Access to the Records of Science Network
CCSDS	Consultative Committee for Space Data Systems
CRL	Center for Research Libraries
DCC	Digital Curation Centre
DDP	Distributed Digital Preservation
DPE	DigitalPreservationEurope
ENISA	European Union Agency for Network and Information Security
ISO	International Organization for Standardization
LOCKSS	Lots of Copies Keep Stuff Safe
NAA	National Archives of Australia
NARA	National Archives and Records Administration
NDIIPP	National Digital Information Infrastructure and Preservation Program
NDSA	National Digital Stewardship Alliance
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
PLN	Private LOCKSS Network
RLG	Research Libraries Group

Tables

Figures

Chapter 1. Introduction

1.1. Presentation and justification of the research theme

The main aim of this study is to investigate the major organizational and regulatory risks in the context of distributed digital preservation in libraries and archives in particular, which will be referred to as memory institutions throughout this report. Furthermore, the study also aims to explore how the inter-organizational relationships emerged in the context of collaborative arrangements and with the use of third-party services are supported in terms of trust, considering the levels of vulnerability and threats to which the organizations involved may be exposed to (Day, 2008).

Additionally, considering that despite several efforts have been addressing digital preservation from the organizational point of view, the urgent need of solving the technological aspects of the activity has resulted on a more prolific research output and discussion focused on the latest (Burda & Teuteberg, 2013; Sanett, 2013). Therefore, a secondary aim of this study is to contribute to address some of the gaps in the scope of the organizational requirements related to digital preservation.

To achieve those aims, the present study considers in the analysis different types of distributed digital preservation models currently in use by memory institutions. The cases that are under study are the use of third-party services to outsource functions or infrastructure needed for digital preservation, with particular attention to the use of cloud or grid technologies, and the cooperative arrangements among institutions partnering to fulfil some of their needs on the activity.

Collaboration and use of third-party services seem to be motivated by the difficulties that digital preservation involves for a single organization (M. Anderson, 2008). In a broader sense, those difficulties have boosted different forms of cooperation, ranging from partnerships for the discussion and set up of common policies and strategies to a more practical approach where distributed or collaborative organizational models, infrastructures or other joint projects have been developed.

Furthermore, the use of third-parties to fill in some of the needs of the memory institutions operating in the cultural heritage or academic sectors is not a novelty and has also been inherited in the digital preservation realm (Lindlar, Friese, Müller, Bähr, & von Trosdorf, 2013). In particular, the use of third-party services has become a common strategy especially to address the technical and technological aspects of digital preservation.

Nevertheless, handing over a core activity such as the preservation of digital assets (Walters & Skinner, 2010), whether it is the whole service or some of its components, has to be thoroughly analysed. There is a need to understand the possible impact on the outcomes of the memory institutions and whether digital preservation can effectively meet its commitments and overcome the challenges that may be posed in those scenarios. Furthermore, other potential consequences in the long term for memory institutions may also need to be thoroughly thought about. Skinner & Halbert (2009) have argued that outsourcing one of the memory institutions' core missions, which is that of preservation, “changes the equation of control of cultural memory in ways that are not ultimately advantageous for cultural memory organizations.” (p. 382).

The reasons for the increasingly popularity of cloud technologies among organizations reside in their low barrier to entry and flexibility (Dhar, 2012), having a broad adoption particularly in the business sector. Some governments have also been taking steps forward, and strategies and programmes for the adoption of cloud seem to be widespread. The United States Federal Cloud Computing Strategy[1] and the United Kingdom Government Cloud Strategy[2] and the G-Cloud framework[3] are good examples of the “cloud first” policies’ trend, which introduce the mandate that purchases through the cloud should be the first option considered in the public sector procurement of IT services.

With a similar approach, the European Commission developed in 2012 its own strategy aimed to “promote the rapid adoption of cloud computing in all sectors of the economy in order to boost productivity” (European Commission, 2012). In this sense, an increase on the easiness of the procurement of these IT solutions could be a factor increasing the levels of adoption in the public

[1] US Federal Cloud Computing Strategy http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/federal-cloud-computing-strategy.pdf

[2] UK Government Cloud Strategy https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/266214/government-cloud-strategy_0.pdf

[3] G-Cloud framework https://www.gov.uk/how-to-use-cloudstore

sector, including memory institutions. This strategy is also the reflection of the increasing awareness on the lack of mapping of some aspects of the use of cloud at the regulatory level.

Nevertheless, despite of the benefits that either outsourcing to cloud, other third parties or collaborating might bring in in terms of efficiency and economies of scale, institutions using these architectures for their services could be subject of a different set of vulnerabilities. Awareness is therefore needed about the threats that could affect them at the organizational level and the digital assets under their responsibility. In this context, this study assumes that risk management can be a useful tool for those organizations in order to identify and analyse the risks and to evaluate whether actions such as the treatment of the risk to fulfil their criteria, are required. The general goal of using these techniques is to allow organizations to be able to handle uncertainty on their organizational outcomes (ISO, 2009, p. V).

This assumption is based on the fact that risk management has been pointed out in many occasions as having a perfect fit with digital preservation as an activity with a large set of uncertainties due to the nature of the materials object of the preservation processes. Therefore, multiple issues concerning digital assets are likely to be identified and managed in terms of risks, and digital preservation be considered as a risk management exercise, converting the uncertainty about maintaining usability of authentic digital objects into quantifiable risks (DCC & DPE, 2007).

In the scope of digital preservation, several efforts have been accomplished by professional organizations, institutions and collaborative projects that have delivered guidelines or reports looking at the risks of outsourcing services of digital preservation, specially addressing technical and security risks of using cloud. In this given context, besides considering the cloud among the outsourced options using commercial providers as the main case in my study, the broader context of distributed digital preservation is also considered. Therefore, the experiences of institutions outsourcing to other types of third-parties or collaborating with other institutions are used to point out commonalities and differences with the cloud solutions.

Finally, the study is therefore an attempt to identify the organizational and regulatory risks, but also to examine the conditions that model inter-organizational trust, since the latest seem to be an important component on the perception of those risks (Walters & McDonald, 2008).

1.2. Aims and objectives

1.2.1. Aims

1. Investigate strategies used in the context of distributed digital preservation in libraries and archives.
2. Explore how the inter-organizational relationships are built in terms of trust.
3. Contribute to addressing some gaps in the scope of organizational requirements related to digital preservation.

1.2.2. Objectives

1. Identify risks at the organizational and regulatory level for the institutions with digital preservation responsibilities using distributed options, through the opinions of experts and their organizational practices.
2. Explore the avoidance and mitigation strategies for the identified risks used by the organizations under study.
3. Identify the benefits perceived with the use of distributed options for digital preservation.
4. Investigate the trust mechanisms that support inter-organizational relationships in the context of distributed digital preservation.
5. Explore the role of collaboration in digital preservation.

1.3. Research questions

1. What are the benefits perceived using distributed digital preservation?
2. What are the major organizational and regulatory risks for memory institutions with long-term preservation responsibilities using distributed digital preservation models?
3. What are the controls or mitigation strategies for the identified risks?
4. Which are the mechanisms of trust in inter-organizational relationships in the scope of distributed digital preservation?
5. Which are the major collaborative trends in the field of digital preservation?

1.4. Methodology

The present study has been developed under a qualitative approach. The main sources of data were semi-structured interviews conducted with experts from memory institutions collaborating with peer institutions or outsourcing functions or storage for digital preservation to third parties, and to representatives of some service providers. The selection of the participants was conducted through a purposive sampling of organizations strategically pointed out as they fulfilled a set of predefined characteristics related to distributed digital preservation. The criteria for the selection of the participants were mainly based in three aspects, the type of organization, the relationship established among them and a set of possible dimensions of the relationships established.

A total of sixteen interviews were conducted mostly in a synchronous way (eleven) using *Skype*, whereas the rest of them (five) were carried out through emails exchanges with the interviewees due to the geographic dispersion of the institutions involved. As additional source of data, the review of relevant documents was also done. A general qualitative analysis of the data was performed and the results are presented in a narrative way including quotations from the participants.

1.5. Outline

The present study is structured in five chapters. The first chapter introduces the topic under study and describe the aims and objectives of the research and the research questions that the study has been designed to respond.

Chapter 2. Literature Review, provides an account of different aspects framing the distributed digital preservation activities, such an overview of the digital preservation activities, requirements and main strategies. Distributed digital preservation, its meaning, scope and the major organizational and regulatory aspects to be considered are also discussed and an overview of actual implementations on the field are presented. A third block in the chapter describes the research on risk management, trust issues and control mechanisms relevant in the scope of distributed digital preservation.

Chapter 3. Methodology, provides with an account of the research design and methods used to accomplish the research objectives and a thorough description of the process followed.

Chapter 4. Research results, provides a complete reporting of the data gathered during the interviews conducted, under the lens of the researcher's analysis.

Chapter 5. Conclusions. Research results presented in chapter four are used to reply the research questions posed.

1.6. Limitations

Time constrains was a major limitation to be able to deepen as much as aimed in the early stages of the conception of the research project. Acknowledging the limitations of time set up as a framework for the master thesis allowed to adjust the scope and research design to be able to fit into the timeframe. Nevertheless, the use of additional sources of data such as interviews with additional experts in particular areas such as legal issues or the involvement of more participants from each institution was discarded from the research design.

Another limitation encountered was the novelty of some of the topics addressed in the scope of distributed digital preservation. Despite there is a growing number of experiences being put in practice, a reflection on the research and professional literature is still under development. Moreover, due to the short time since some of the phenomena under study have been implemented, long-term implications are difficult be analysed yet.

1.7. Significance

The study aims to contribute to the stream of research in topics related with distributed digital preservation. Furthermore, due to the scarcity and fragmentation on previous research in specific aspects such as those related with the risks from the organizational point of view for institutions in charge of digital preservation activities or the inter-organizational trust issues in this context, it attempts to fill in some of the gaps. In the scope of this contribution, the study is also aimed to inform institutions in the field and potentially contribute in a practical manner to serve as an aid in their decision making processes.

Chapter 2. Literature Review

2.1. Introduction

Digital preservation has been already presented as a complex task in presentation of the research theme in chapter 1; it is an activity that is subject of a large number of uncertainties, due to the type of materials that need to be preserved. But also, due to other factors such as the funding models or the legislative framework that surrounds the activity, it is complex and in continuous evolution. Therefore, this chapter addresses different concepts related with digital preservation such as collaborative approaches or the distributed architectures and their organizational and regulatory aspects that are subject of this study. But also the literature on risks management, which is the lens through which the research has been conducted and trust from the point of view of digital archives and the inter-organizational relationships.

The literature discussed in this chapter has been retrieved using online databases such as EBSCOHost, Emerald, Springer, ACM, Sage, Taylor&Francis and MUSE. The search terms were basically combinations of “digital preservation”, “digital curation”, “cloud computing”, “outsourcing”, “risk management”, “distributed digital preservation”, “trust”, “federated digital preservation”. Conference proceedings were a major source of information, and in some cases the papers were retrieved from their websites: PASIG, iPRES, Open Repositories or TPDL were especially relevant.

2.2. Previous research done in the topic area

2.2.1. Digital Preservation

2.2.1.1. Definition

Preservation of the materials entrusted to memory institutions has been one of their responsibilities along their history. Nevertheless, the preservation of digital assets, which keep growing in numbers in such institutions, entails a different set of complexities that are even nowadays difficult to handle with. As M. Anderson (2008, p. 5) asserts, “libraries and archives face a daunting task in their efforts to continue the tradition of preservation in the digital age”. How those institutions will be able to tackle this task in the future is still difficult to foresee.

For over more than a decade, efforts and investments have been directed in memory institutions to broaden their digital assets. This process has involved the digitization of a vast amount of materials belonging to their own collections; the collection of the research output of the academic institutions and its preservation and access through digital repositories; and the inclusion of new items in the collections or archival materials produced only in digital form. The lasting benefits from the investments in digitization and in the collection of digital content will require the support of digital preservation activities, including the development of information infrastructures and curation skills on researchers and information professionals (Beagrie, 2006). Moreover, considering that the preservation of digital objects is a complex activity and highly demanding on resources and know-how that requires an effort sustained over the long term (Giaretta, 2008), strategic view and planning seem to be also important requirements to succeed.

The reasons behind that complexity and the need of resources for digital preservation can be traced through its definitions in the professional and research literature. Digital preservation has been characterized as an activity that implies the challenge of an active management of the digital objects to the extent that actions need to be taken in order to maintain their conditions and requirements. And the second challenge pointed out is that this active management to ensure access to the digital assets has to be done over long periods of time. For instance, (Beagrie & Jones, 2008, p. 24) argue that digital preservation refers to “the series of managed activities necessary to ensure continued access to digital materials for as long as necessary.” Similarly, the Library of Congress[4] considers the term as the “active management of digital content over time to ensure ongoing access.” Barateiro, Antunes, Freitas, & Borbinha (2010, p. 5) claim that the aim of digital preservation is to “optimize the information life-cycle management, from the creation to the dissemination and use of the information objects, for long periods of time.” Caplan (2008) defines digital preservation as a “process of ensuring that a digital object remains accessible over the long-term” (p. 33), describing long-term as “to be long enough for changes in technology to threaten the usability of the object” (p. 33). Other definitions of digital

[4] http://www.digitalpreservation.gov/about/

preservation also stress the idea of long periods of time claiming that digital preservation seeks interoperability or communication with the future (Moore, 2008), in the sense that "digital objects must remain authentic and accessible to users and systems" (Barateiro et al., 2010, p. 5) maintaining their value in the long run.

In the organizational context, digital preservation needs to be supported by policies, strategies and actions (Barateiro, Antunes, Cabral, Borbinha, & Rodrigues, 2008) to ensure that the conditions for protecting the digital objects are stable for a long time, independently of the challenges introduced by component and management failures, as well as natural disasters or deliberated attacks (Barateiro et al., 2008).

2.2.1.2. Requirements and strategies for digital preservation

Successful digital preservation is characterized at a technical level by a "widely-accepted list of properties characterizing well-preserved digital objects" (Vermaaten, Lavoie, & Caplan, 2012). Vermaaten et al. (2012) describe those properties in their model for risk assessment *Simple Property-Oriented Threat (SPOT)*, pointing out the threats that would diminish the ability of the repository to achieve its objectives:

- Availability is the property through which long-term use of a digital object is possible, and could be threatened by decision-making related to long term value, property rights or physical barriers.
- Identity of an object implies that it can be discovered and retrieved through sufficient metadata able to reference to it.
- Persistence implies that the bit sequence of the digital objects maintain their state, usability and retrievability in their storage medium, and could be under risk in aspects such as physical media management and its related policies, as well as hardware migration or data security.
- Renderability means that the object can be used in a way that keeps its significant characteristics. It could be threatened by aspects such as format management workflows and policies, including preservation strategies and repository knowledge of its stakeholder community.

- Understandability implies that the content can be appropriately interpreted and understood by its intended users, and therefore it is important to acknowledge the characteristics of the communities of users, metadata and retention policies.
- Authenticity means that the digital object is what it purports to be, and relevant aspects that need to be adequately addressed are metadata collection and management practices, security procedures, and workflow documentation procedures and policies.

Taking into account the described properties, a broad distinction can be made between two types of technical strategies: bit preservation and functional preservation. The earliest must ensure that the bits remain intact and accessible being thus a starting point for further preservation actions. The latest, assures that the data remains understandable through further preservation actions, out of the scope of bit preservation (Zierau, Kejser, & Kulovits, 2010).

The National Digital Stewardship Alliance (NDSA) has developed a guideline called *Levels of Digital Preservation* that comprises a basic framework for developing and planning digital preservation activities, from a technical point of view (Phillips, Bailey, Goethals, & Owens, 2013, p. 1). The matrix is composed by five categories with four different levels of achievement, that identify technical functions and features to ensure long term access to digital content, being the earliest prerequisites for the latest (Phillips et al., 2013). The areas covered and the main strategies related to them are described as follows:

- Storage and geographic location: distribution of copies to ensure data redundancy and geographical dispersion to avoid data loss due to disaster threats; storage monitoring against obsolescence, documentation and plans to maintain accessibility.
- File fixity and data integrity: fixity checking and mitigation strategies to avoid or repair corrupted data.
- Information security: authorization to access and logs of actions performed in the data, ensuring accountability and transparency.
- Metadata: different levels of description starting from general inventory to administrative, transformative, technical and descriptive and preservation.
- File formats: normalization and use of open formats, creation of a register of the formats, monitoring obsolescence and perform actions such as migration or emulation.

2.2.1.3. Technological versus organizational aspects in digital preservation

A large and growing body of literature has investigated technical and technological aspects of digital preservation, such as the development of different preservation strategies or metadata schemas (Day, 2008), leaving aside until fairly recently those related with organizational aspects.

The challenges related to the technological issues in the scope of digital preservation are of major importance and still under development. Nevertheless, addressing the variety of challenges related to the ability of organizations to integrate the management of digital materials into their organizational structure is another significant part of the problem (Beagrie & Jones, 2008) that needs to be addressed. Furthermore, Lavoie & Dempsey (2004) argue that digital preservation is not an isolated process that can be tackled only from the technical point of view, but instead, a component of a broad aggregation of interconnected services, policies, and stakeholders that constitute the digital information environment.

Nevertheless, the systematic literature review conducted by Burda & Teuteberg (2013) with the aim to investigate how digital preservation is addressed in the research literature, pointed out the gaps on examining digital preservation through an organizational lens. Whereas the development and evaluation of preservation techniques or strategies have been the main issues discussed, there are unsolved organizational aspects such as, in particular, the lack of methods to support cost–benefit analysis or digital preservation decision making.

Similarly, the practice in the memory institutions has also been putting a great focus on the technological aspects. Sanett (2013) conducted a longitudinal study, gathering information on management practices in national archives developing digital preservation activities from 1999 until 2007. Through the surveys conducted during the two first stages of the study (1999-2003), awareness was raised on which areas of the digital preservation programs examined were underdeveloped. The absence of an effective managerial infrastructure to support and sustain these programs over the long term was specially noted and the key areas which remained behind the technological developments were policy, staffing and costs.

2.2.2. The collaborative approach to digital preservation

Sharing the burden of digital preservation as a way optimize the efforts made by the institutions responsible to be the long-term keepers of both cultural heritage and scientific output, has been subject of continuous discussion. M. Anderson (2008) claims that "preserving our cultural heritage is not a mission that can be accomplished by a single institution" (p. 5). The working group (RLG & OCLC, 2002) also addresses the issue of collaboration, remarking the importance of the creation of models for the establishment of cooperative archiving services, accompanied by more thorough understanding by information professionals of how cooperative digital repositories and networks can be implemented and managed. But it was already in 1996 when the *Task Force on Archiving of Digital Information* (Waters & Garrett, 1996) shed light to the idea that, to be able to tackle the organizational challenges of digital preservation there was a the need for infrastructure able to support a distributed system of digital repositories and other services (Day, 2008).

Based on this approach several institutions and initiatives have built their efforts towards collaboration with outcomes such as policies and strategies, partnerships between institutions or funding allocated on joint projects that have been fostering major developments, research and implementations in the field. Therefore, the collaborative approach has been developed at different levels, not only between individual organizations through partnerships, but also in a larger scale with national or even international initiatives. Additionally, as stated by Lindlar, Friese, Müller, Bähr, & von Trosdorf (2013), there are three main factors which usually play a role when looking for partners to collaborate with in the scope of digital preservation: geographical distance or association, organizational association, and collection factors.

Collaboration can help institutions to better address digital preservation challenges and to ensure the sustainability of the activity in the long run. Day (2008) propounds collaboration as a key part of the development of the organisational infrastructures that underpin institutional repository networks and digital preservation in general. Furthermore, Lavoie & Dempsey (2004) remark the need of an "ongoing, long-term commitment, often shared, and cooperatively met, by many stakeholders." And Beagrie & Jones (2008) argue that collaboration at different levels would “maximise the benefits of the technology, address issues such as copyright, and also to overcome the challenges cost-effectively” (p. 38).

In this sense, a number of benefits of collaboration in the scope of digital preservation can be pointed out. Some of them are economic aspects such as the improvement of cost-effectiveness, sharing of resources, tools and expertise (Lindlar et al., 2013; Trehub & Halbert, 2012). Collaboration as a way to strengthen the community-based stewardship maintaining control over the digital assets (Trehub & Halbert, 2012), engaging in efficient workflows (Lindlar et al., 2013), generating managerial efficiency; consecution of economies of scale (Lavoie & Dempsey, 2004; Evens & Hauttekeete, 2011; Lindlar et al., 2013); the leveraging of existent expertise and infrastructure (Jordan, Kozbial, Minor, & McDonald, 2008); the reduction on the dependence on limited resources or the achievement of common approaches to meet the goals of the partnership (Downs & Chen, 2010).

In particular, the issue of economic sustainability is an essential factor to maintain the activity of digital preservation in an institution. Nonetheless, getting guarantees of a stream of funding in perpetuity seems a task difficult to achieve (Downs & Chen, 2010; Giaretta, 2008; Jordan et al., 2008). The development of “institutional commitments on the scale and timeframe” (p. 6) assumed by libraries or other institutions engaging in digital preservation are therefore needed (Jordan et al., 2008). However, even a long-term commitment does not give enough assurance on getting support for the activity as discontinuity of the organization or changes on priorities may happen to any institution, including governments, memory institutions, private sector organizations or any other funding source (Downs & Chen, 2010). The issue of economic sustainability in digital preservation will be discussed in section 2.2.6.3.

Sharing the responsibility of digital preservation through the development of collaboration networks could be an optimum strategy to support sustainability (Downs & Chen, 2010; Giaretta, 2008). But it also introduces new concerns as it is more complex than just a “simple chain of preservation consisting of handing on the collection of bits from one holder to the next” (Giaretta, 2008, p. 113). Pooling digital preservation in a sustainable manner needs thus to be accompanied by organizational structures and planning able to support the commitment, although difficulties might also arise to introduce structural cooperation within the daily management of the institutions (Evens & Hauttekeete, 2011).

2.2.3. Distributed Digital Preservation

2.2.3.1. Meaning and scope

Despite distributed digital preservation has become a common phrase in the field, what it is actually designating does not seem to be so straightforward. Zierau & Schultz (2013) draw our attention on the variety of initiatives that have been labelled as distributed in digital preservation, as well as the multiplicity of reasons for adopting those approaches. They suggest that there is a lack of a commonly accepted definition, despite the widespread use of distribution in the scope of digital preservation. In the context of their project aimed to create a framework to adapt OAIS to distributed digital preservation, they use this term "to emphasize the practice of applying distribution in intentional ways, both organizationally and technically, for accomplishing digital preservation, for example through geographic distribution, infrastructure heterogeneity and organizational diversity" (p. 1) or in a more concise way "the use of replication, independence, and coordination to address the known threats to digital content through time to ensure their accessibility" (p. 1).

Walters & McDonald (2008) echoing the use of the phrase distributed digital preservation federations explain that it is used to describe "cooperatives of geographically-dispersed institutions who are banding together to form solutions to the digital preservation problem" (p. 1). They consider the term federation appropriate to name this type of relationships in the sense that whereas a number of organizations come together to solve a common need, they still keep the control of their own internal affairs.

Caplan (2008) also claims that the distribution of responsibilities for digital preservation among different agencies and applications is a way to bring in advantages for the activity, pointing out two alternative ways of using the distributed model. On one hand, consortia models running facilities for digital preservation and in the other, the use of third-party services, both for and not for profit. In the second case, distributed systems for digital preservation would be therefore based on the adoption of a 'disaggregated' approach of the digital preservation activities, "where the various components of the preservation process are broken apart into separate services distributed over multiple organizations, each specializing in a focused segment of the overall process" (Lavoie & Dempsey, 2004).

2.2.3.2. Typology of distributed digital preservation

Zierau & Schultz (2013) classify the spectrum of distributed digital preservation in three main categories: organizational, geographical and systems-based. Each of the categories are interrelated and can overlap, as it will be examined in the following paragraphs.

a) Organizational

From an organizational point of view, using a distributed approach for digital preservation can range from collaborative associations between peer institutions to the establishment of contractual relationships with third-parties.

Skinner & Halbert (2009) make a distinction between decentralized and centralized approaches. Consortia and cooperation models are decentralized structures used by memory institutions either in particular projects or through partnerships with longer commitment to coordinate their activities and put together the efforts required for digital preservation. Halbert (2009) suggests that cooperative arrangements are a good option for memory institutions that are not in the position to afford the requirements to preserve their digitized or digital-born materials in a distributed manner allowing for instance, the replication of materials as a strategy for preservation. At the operational level, long-term inter-institutional agreements would typically translate into distributed grids shared by partnering institutions (Wittek & Darányi, 2012).

Nevertheless, sustaining distributed digital preservation infrastructures has been suggested to be more challenging considering the organizational aspects rather than the technical ones, as stated by Halbert (2009). In the case of collaborative approaches, the author also argues that challenges might appear because of the lack of institutional experience on tasks such as participating in networks, deciding about the foundational requirements and the analysis of business and cost models, carrying out strategic planning or effectively running the network in a jointly manner. Halbert (2009) also claims that competitive factors related to institutional prestige or *de facto* leadership taken by the largest institution in a network might cause malfunctions in unincorporated partnerships in opposition to cooperative models.

In centralized approaches, there are different possible scenarios. For instance, digital preservation responsibilities can be hosted by centralized institutions external to the memory institution. In this case, the third-parties are usually either another public organization in the sphere of memory institutions or academia, or commercial providers to which outsource the digital preservation activities through contractual arrangements.

The use of third-party service providers to distribute the responsibility required to fill in the functions needed in digital preservation (RLG & OCLC, 2002) has been a common strategy for many institutions. Moreover, digital preservation functions have been in several occasions assumed as a responsibility of a third-party rather than being part of the organizations' role. Day (2008) illustrates this situation with the case of institutional repositories, where the responsibility of preserving the content was not clearly stated to be of the institution owner of the resource and it has been generally understood as a need for inter-institutional collaboration and third-party services specially dedicated to long-term preservation services. The services offered by third parties will be explored more in detail later on in section 2.2.5.2.

b) Geographical

Long-term preservation of information has historically been possible by gathering copies of content in secure archives geographically distributed (Halbert, 2009).

The distribution of copies of digital assets to geographically dispersed locations has been pointed out as a responsible practice for any digital preservation system (Wittek & Darányi, 2012). Redundancy and geographical distributed copies of the digital objects has become indeed a very common strategy to reduce the risk of data loss.

For instance, this strategy would be helpful in cases such as the corruption of the bit-stream of a particular object, inadequate handling due to human errors or natural disasters in a particular data centre. Moreover, the distribution of copies would possibly ensure that the digital assets do not get lost in case of financial or organizational failure of the institution responsible for their preservation.

One of the potential difficulties of using this type of architecture is that a single organization with preservation needs might not have enough capability to handle geographically dispersed and secure infrastructure (Wittek & Darányi, 2012). Therefore, the collaboration among institutions or the use of external providers, earlier mentioned in this paper, has become widespread in this type of infrastructures.

c) Systems-based

There is a diversity of types of storage provision, hardware or software components, protocols or ways of implementing the processes needed in the digital preservation architectures. The use of the heterogeneity within those systems is also considered a strategy for distributed digital preservation. Looking at actual implementations (Schultz & Skinner, 2014) carried out a comparative analysis of three distributed digital preservation systems, taking in consideration aspects such as ingest, data models, storage, monitoring, security, recovery, scalability and costs. The systems in question were *Chronopolis* using *iRODS*, University of North Texas using *Coda* and *MetaArchive* using *LOCKSS*.

Some of the main features observed were that in terms of storage, servers were sourced from multiple vendors and provisioned differently at the various replicated storage nodes of the networks, as heterogeneity was considered a priority for all of them. Recovery possibilities ranged from restricted to circumstances of data loss in the cases of dark archives, to returning any data upon request to the data provider. Scalability on collections, replication, and organizational expansion was implemented through scheduled and as-needed consultations with partners and management, including storage or staff provisioning. In terms of costs, the three initiatives use different cost models specific to their missions and operations. There are different security levels, ingest methods and data models in use, but it is remarkable that the three systems use strategies for documenting changes and monitoring technologies depending on their organizational and technical environments.

The use of cloud and grid technologies in digital preservation will be further explored in the section 2.2.5.

2.2.4. Landscape of collaborative efforts in digital preservation

As mentioned earlier, collaborative initiatives are a common way to articulate digital preservation activities. Some of the initiatives will be presented in this section, categorizing them in three groups: development of strategies and policies, development of preservation infrastructures and an overview of certain remarkable implementations of distributed digital preservation.

2.2.4.1. National or international cooperation efforts for strategy and policy development

The organizations in this group are mostly focused on high-level developments such as policies and strategies. Additionally, in some cases they provide critical funding for the initial development of services and infrastructures. Some initiatives or institutions with remarkable outcomes are *Digital Preservation Europe* (DPE), *Digital Preservation Coalition* (DPC), *Network of Expertise in long-term Storage* (Nestor) or *the National Digital Information Infrastructure and Preservation Program* (NDIIPP).

In particular, the NDIIPP, which is based at the Library of Congress, has had a very influential role on the developments in digital preservation through its network. Some of the partners focus on the development of services for long-term digital preservation, performing work that is aimed to be useful to different entities involving the actual deployment of those services and not only resting on expertise and information sharing. Some of the projects that were born under NDIIPP umbrella are as remarkable as Dspace, Fedora or LOCKSS (M. Anderson, 2008).

2.2.4.2. Preservation infrastructures and services

This group of initiatives are focused on the development of collaborative infrastructures or third-party services that are able to assume the functions needed for the preservation of digital assets on behalf of memory institutions. A remarkable example from the point of view of enabling the development of services, is the EU-funded project “Preservation and Long-term Access through Networked Services” (PLANETS) has had an important role. One of its outcomes was the development of a testbed, which is an integrated environment to support experimentation with preservation strategies, from the design and testing to the actual deployment. Through a variety of end-user applications, the testbed allows preservation experts to conduct experiments using an ample number of preservation services. The system is built into a distributed research infrastructure, integrating existing content repositories, preservation tools, and services (Schmidt et al., 2009).

2.2.4.3. Distributed digital preservation strategy implementations

There are several implementations of distributed digital preservation strategies currently functioning. In this section, an account of only some of the most remarkable collaborative implementations will be explained.

- LOCKSS

Based on Stanford University and created originally under the sponsorship of the NDIIPP, LOCKSS (*Lots of Copies Keep Stuff Safe*)[5] is an open-source software that has evolved into an international community initiative that provides tools and support to institutions with responsibilities on preserving digital assets. In a typical LOCKSS network there are two types of stakeholders, that could converge: content providers, which are those whose content is being crawled and ingested by preservation nodes and a number of institutions that partner together to administer the infrastructure and to preserve this content collaboratively and in a distributed manner (Skinner & Halbert, 2009).

The LOCKSS public network comprises a large number of institutions, preserving content of interest for its members (Reich & Rosenthal, 2009). Similarly, CLOCKSS[6] is a joint project between publishers and memory institutions to keep dark copies of scholarly publications in digital format. Through the use of LOCKSS, the participating institutions collaborate among themselves keeping several copies of digital objects in a geographically distributed manner, avoiding data loss in case of events occurring in one of the nodes of the network.

- Private LOCKSS Networks.

Private LOCKSS Networks (PLNs) are one of the possible implementations of LOCKSS, providing institutions with a technological tool that a low cost and community-run distributed

[5] http://www.lockss.org/
[6] http://www.clockss.org/

method for preserving digital content. They are usually conformed by a group of institutions with equivalent missions that partner together to preserve the digital assets they are responsible for. One of the most well-known implementations is the *MetaArchive* Initiative[7] which is a cooperative organization. Skinner & Halbert (2009) describe the initiative as an organization operating a technical infrastructure using LOCKSS for the preservation of digital assets of memory institutions in a geographically distributed framework. They also remark that its aim is to empower the institutions part of the cooperative to handle their own preservation solutions in collaboration with their peers, instead of outsourcing it, as it is assumed as one of the core missions of memory institutions. This idea is indeed at the core of the cooperative's philosophical approach, putting forward a solution to enable institutions to take responsibility on the preservation needs, instead of losing capabilities to fulfil their mandate through the outsourcing of the tasks related to digital preservation.

- Chronopolis Digital Preservation Initiative

Chronopolis[8] data grid framework for digital preservation was developed by the San Diego Supercomputer Centre and the University of California San Diego Libraries (UCSDL), among other partners[9]. *Chronopolis* has generated its own business model, evolving from a project funded only by the *NDIIPP* into a broader-reaching fee-for-service model (Minor et al., 2010). The service was certified by the Center for Research Libraries (CRL) as a trustworthy digital repository in 2012.

The mission statement is paraphrased in CRL (2012) and defines *Chronopolis* as “a preservation data grid and its supporting human, policy, and technological infrastructure” which does not provide preservation services beyond the bits deposited by the client. The four partners that constitute the organization contribute with expertise and infrastructure to create an archive system characterized for its geographical distribution, heterogeneity and redundancy (CRL, 2012).

[7] http://www.metaarchive.org/

[8] http://chronopolis.sdsc.edu/

[9] San Diego Supercomputer Center (SDSC) at UC San Diego, the UC San Diego Libraries (UCSDL), National Center for Atmospheric Research (NCAR) in Colorado and the University of Maryland's Institute for Advanced Computer Studies (UMIACS).

The geographical distribution in different locations and grids seems to be the main strength of the system, because it is "essential to minimize risk, provide high performance access across the nation, support distribution, and balance load" (Moore, 2004).

- Danish Bit Repository

The Danish Bit Repository[10] is a platform that allows bit preservation in a shared environment. Zierau & Schultz (2013) assert that the Royal Library of Denmark is "a pioneer in proposing this model, which is an approach to achieving reliable, auditable distributed preservation." (p. 2). They also point out the challenges the system cause to maintain independence among copies of data and how to use bit preservation solutions for different requirements of bit integrity, confidentiality and availability.

- Digital Preservation Network (DPN)

Using a federated approach to preservation, DPN[11] was formed to ensure the long-term preservation an access of scholarly record that the higher education community is storing in their digital repositories. Currently in its start-up phase, is building a digital preservation backbone that connects five preservation-oriented repositories: the *Academic Preservation Trust* (*APTrust*), *Chronopolis*, *HathiTrust*, *Stanford Digital Repository* (SDR), and the *University of Texas Digital Repository* (UTDR). The five nodes introduce to the network the benefits of geographical, systems and organizational and financial diversity (Hilton, Cramer, & Minor, 2013).

The network will link the first node where the content is deposited with other repositories and replicate the digital objects in at least three nodes, using dark copies for preservation purposes and applying processes related to bit preservation[12]. DPN also counts with the participation of *DuraSpace* and it is currently membership is over fifty members across the United States.

The network aims to build a "sustainable ecosystem in which digital preservation can scale and evolve" and claims benefits such as resilience, succession, economies of scale, efficiency, extensibility and security (Hilton et al., 2013).

[10] https://sbforge.org/display/BITMAG/
[11] http://www.dpn.org/
[12] DPN specifications can be accessed at https://wiki.duraspace.org/display/DPNC/Specifications

- Goportis

The Leibniz Library Network[13] for research information is formed by the collaboration the German National Library of Science and Technology, the German National Library of Medicine and the German National Library of Economics. As Lindlar et al. (2013) describe, the three partners have been conducting a digital preservation project since 2010 to fulfil their mandate of archiving and the responsibility for maintaining the long-term-access. Therefore, the collaboration is aimed to build "a sustainable trustworthy digital preservation system for the three National Subject Libraries in Germany" (p. 2). Some of the objectives of the network related to digital preservation are the optimization of workflows, gaining expertise in the field and implementing a joint infrastructure for a digital long-term archive.

- Data-PASS

The *Data Preservation Alliance for the Social Sciences* (Data-PASS)[14] is "a voluntary partnership of organizations created to archive, catalogue and preserve data used for social science research."[15] The collaborative activities started as a program funded by the NDIIPP of the Library of Congress and it is formed by large academic institutions in the United States. The partnership works to archive social science data, maintain a shared catalogue, maintain replicated preservation of archived collections and advocate best practices in digital preservation (Altman et al., 2009).

2.2.5. Outsourcing digital preservation

2.2.5.1. Outsourcing in memory institutions

The American Library Association (ALA, n.d.) defines outsourcing in the context of libraries as an activity that "involves transfer to a third party, outside vendor, contractor, independent workers, or provider to perform certain work-related tasks involving recurring internal activities that are not core to the mission of the library."

[13] http://www.goportis.de/

[14] http://www.data-pass.org/

[15] Ibid

One of the aspects to take into consideration when outsourcing archiving services is the transfer of custody and therefore the responsibility from the memory institution to the supplier of the service, including in some cases keeping physically the records (Dečman, 2007). In this context, transparency and reliability in the relationship become critical factors for trusting the service.

Whereas ALA did not consider in its definition of outsourcing the possible externalization of core missions of memory institutions when using third-party services, this is exactly one of the arguments that advocates of collaborative systems are showing a consensus about as a possible source of threats of outsourcing to commercial providers. Digital preservation functions as a core mission of memory institutions in a context of the cultural memory becoming increasingly digital, does not seem a good choice to be handed to commercial providers in their opinion (Skinner & Halbert, 2009).

Some of the consequences that have been argued are the loss of control (Halbert, 2009); the loss of expertise in digital preservation assessment, workflows, and technologies; the impact on the quality of the preservation work; the detriment of the value proposition and viability as institutions (Walters & Skinner, 2010). The restructuration of the memory institution's sector and centralization of the functions of preservation in few specialized corporations (Skinner & Halbert, 2009) with objectives and mission oriented to the generation of profit; lack of transference of memory institution's mission into that of the commercial providers (Walters & Skinner, 2010), are also potential consequences that have been stated.

Nevertheless, Lavoie & Dempsey (2004) claim that whereas memory institutions will keep their role as stewards of the digital collections, they consider that it may not be possible for every institution responsible for digital preservation to have the whole set of resources and expertise to implement the entire process locally. Therefore, it is not unlikely that part of the responsibility is handed to third-party services.

In addition, the benefits of outsourcing have been also examined. Dečman (2007) argues that outsourcing the archiving functions to a service provider "can eliminate upfront capital costs, offer a predictable cost structure, and provide valuable expertise to help an organization stay compliant, while getting the service up-and-running very quickly" (p. 138). Moreover, it is also suggested that public-private partnerships can establish those services in a "fast, efficient and inexpensive way" (p. 142) compared to a possible higher capital and revenue cost if the service is implemented by the organisation itself. Scale, specialization and technical cost-efficiency are the vantages of using third-party services pointed out by Lavoie & Dempsey (2004). Similarly, Lambert, Hein, Bazzanella, Proell, & Strodl (2014) argue that preservation institutions can benefit from sharing and be able to scale and increase robustness through those services.

2.2.5.2. Services for digital preservation

Services can be understood as an aspect of sustainability of the activity of digital preservation. (Lambert et al., 2014) define them as "activities with a specific function that can help organisations preserve their digital holdings against threats such as changes in hardware, software, environment and designated communities, and that have a scope and applicability wider than a single digital repository" (p. 6). High-level services are defined by the authors as those designed to give an answer to the potential need of services in digital preservation, rather than the actual provision in the field.

To identify those high-level services in the chain of digital preservation, the report published by the APARSEN project (Lambert et al., 2014) uses OAIS functional entities. Lavoie & Dempsey (2004) also propose the deconstruction of a digital preservation system in functional layers, allowing therefore fulfilling the needs of the functions through separate but interoperable services, combined depending on the needs of the repository and the digital materials.

For instance, long-term archiving services, cloud storage for preservation or full repository service are some of the identified services with actual implementations in the market.

2.2.5.3. e-Infrastructures and grids in the scope of digital preservation

e-Infrastructures or cyberinfrastructures, being the first the preferred term in Europe and the latest in the United States, have been gaining momentum in the scope of digital preservation. Stewart et al. (2010), through the compilation of different definitions of the term state that cyberinfrastructure "consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible."

e-Infrastructure collaborations with memory institutions for digital preservation have been subject of research and remarkable deployments, specially related to the management of research data and its preservation. The infrastructure of supercomputing and data centres has been found to provide new usefulness in digital preservation environments and the endeavours of memory institutions, funding agencies, and infrastructures can be taken as complementary for digital preservation (Jordan et al., 2008).

The e-Infrastructures were originally created to provide computational capabilities for research. Nevertheless, the idea that they would fulfil other functions was early developed. For instance, in the US a NSF report (Atkins, 2003) and other NSF activities were relevant on making clear that the functions of those infrastructures were wider (Jordan et al., 2008). One of the examples of active collaboration in relation with digital preservation has been the San Diego Supercomputer Center (SDSC) in relationship with the Library of Congress or National Archives and Records Administration.

Other remarkable initiatives have been the project *InterPARES 2, which* investigated the creation of preservation environments using data grids. The project DCH-RP (Digital Cultural Heritage Roadmap for Preservation), which besides aiming to harmonize data storage and preservation policies in the digital cultural heritage sector, also tries to make advances in a dialogue and integration among institutions, e-Infrastructures, research and private organisations. Furthermore, the project's aim is also to identify models for the governance, maintenance and sustainability of the integrated infrastructure for digital preservation of cultural content[16].

Wittek & Darányi (2012) describe grids as infrastructures potentially relevant to fulfil services needed in digital preservation. They already store data that must be preserved and also provide a set of functionalities required by digital preservation systems such as redundancy. They can be federated, which is a feature that may allow different grids with from different institutions to interoperate and share data.

Some of the benefits of the collaboration for digital preservation reside on the combination of the technological side offered by e-infrastructures, with data grids that have the capabilities for

[16] http://www.dch-rp.eu/

replication and distribution of data[17]. Secondly, those centres have experience on bit preservation and seem suitable to develop more expertise on technological aspects of digital preservation. And lastly, their network availability and capabilities for access (Jordan et al., 2008). Nevertheless the development of partnerships with memory institutions and domain experts turns to be a need to support other actions required for preservation and to guarantee the understandability and accessibility of the data over the long term.

2.2.5.4. The use of cloud technologies for digital preservation

Using cloud has become a widespread approach for the provision of computing services in all kinds of sectors, public or private, from business to government. This type of deployments seem to have realized completely the idea of "utility computing", serving computing services on-demand on a "flexible, efficient and readily-accessible manner" (Bradshaw, Millard, & Walden, 2010, p. 3). Kyriazis (2013, p. 1) refers to cloud as "a paradigm building on a set of combined technologies, it enables service provision through the commoditization of IT assets and on-demand usage patterns."

Bradshaw et al. (2010, p. 6) argue that "cloud computing provides flexible, location-independent access to computing resources that are quickly and seamlessly allocated or released in response to demand. Services (especially infrastructure) are abstracted and typically virtualised, generally being allocated from a pool shared as a fungible resource with other customers. Charging, where present, is commonly on an access basis, often in proportion to the resources used."

It differs on traditional outsourcing in the way that organizations hand over the responsibility to service providers in the fulfilment of their needs in terms of the contracts and agreements established among them, raising different legal issues. In the particular case of outsourcing to cloud, Dhar (2012) refers to "a paradigm shift to an asset-free provision of technological resources" (p. 670), that allows to increase the flexibility, efficiency and to introduce cost reductions on the provision and the service compared to traditional IT outsourcing. Some other differences are the lack of up-front costs using cloud as a result of the leasing of resources, less level of customization in the cloud or what can be more on the downside and potentially increase

[17] For instance, iRODS data grid software was created to support the policy and management needs of long-term digital repositories (Jordan et al., 2008)

the risks of the activity, “a lack of provisions for compliance, business continuity, security, and privacy of data” (p. 670).

Cloud computing can be used with different service and deployment models, that are often characterised in terms of the type of service that is being offered (Bradshaw et al., 2010) and are usually known as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). Those same services can be deployed in four different modalities: public, private, community or hybrid clouds. Public clouds exemplify the paradigm of the already mentioned utility computing, owned by a service provider, granting organizations with access and use of a generally share resources or technology on a pay-per-use basis, leveraging economies of scale and elasticity of the resource. Whereas private clouds are those in which “the relevant infrastructure is owned by, or operated for the benefit of, a single large customer” (Bradshaw et al., 2010, p. 4).

In the scope of digital preservation, the initiative *Long Term Digital Preservation Reference Model* (LTDP)[18] differentiates between three types of services in relation to the use of cloud: a cloud digital archive service, a cloud digital preservation service and cloud storage. A cloud digital archive service is different of a storage cloud, because of the additional features and services characteristic of the former. Additionally, when a digital archive includes services to support digital preservation, it becomes a cloud-digital preservation service. Moreover, a digital preservation service may use other cloud-based services to augment its own features (Peterson, 2011).

There are different opinions about the adequacy of the use of cloud for digital preservation. For instance, Rosenthal & Vargas (2013) studied the possible use of cloud storage with the PLNs, running an experiment in which they implemented LOCKSS boxes in Amazon’s cloud service. They concluded that cloud storage was not as cost-competitive as local disk storage for long term preservation in general and LOCKSS in particular.

Whereas, on the other hand, Chowdhury (2013) states that using technologies such as cloud computing may reduce the economic and environmental impact of digital information.

[18] http://www.ltdprm.org/

Nevertheless, he also argues that there are a number of social and user related issues that should be considered when using cloud-based IT systems, such as those related with sensitive content, user behaviour or institutional and user culture and practices.

Guidance in the use of cloud technologies in digital archives has been developed providing especially with information on the risks that the use of those can introduce in aspects such security and from the technological side. The National Archives have recently published a guidance with recommendations on the use of cloud storage by public archives in the UK (Beagrie, Charlesworth, & Miller, 2014). Using a set of case studies, the guidance explores different aspects concerning cloud in relationship with the needs of digital preservation, from the dimensions of security, legal issues and costs. Additionally, they provide a comprehensive risk register on legal requirements and key contractual and SLA issues to consider when engaging in this type of services.

Similarly, the Minnesota Historical Society (Toussaint & Rounds, 2013) also published a report analysing the main concerns that they considered relevant in for digital preservation using cloud services. The report provides an overview that sets up the scene through a set of requirements mostly related with the capacity of the systems to ensure the properties of the digital objects, but also considering issues related with the ownership, data portability and costs. They also include an appraisal of a selection of providers in the market, assessing them against the pre-defined criteria.

A relevant study on the use of cloud in the scope of digital curation was produced as a final report of the JISC’s Curation in the Cloud Workshop (Aitken, McCann, McHugh, & Miller, 2012). The report provides an overview of pilot and functioning experiences using either private or public cloud services, and includes an analysis of aspects such as sustainability, costs or technology. Furthermore, they provide a synthetized account on the major issues to be considered areas of risk for the activity of digital curation.

The Archives & Records Association UK & Ireland also produced a study and a toolkit for institutions as a guidance for outsourcing storage to cloud (Convery, 2010a). The report of the study (Convery, 2010b) especially encourages to the institutions entering into arrangements with cloud providers to consider a risk assessment for the organization and assets outsourced to cloud, the procurement rules and policies.

More general guidance and risk registers on using cloud technologies have been produced by organizations such as ENISA (2009) or Cloud Sweden (Lindström, 2011).

A general overview of the literature reveals a variety of risks with consequences to regulatory or organizational aspects, such as the possibility of vendor lock-in, non-compliance with certification or legislative frameworks, liabilities for infringement of data protection regulations (Aitken, McCann, McHugh, & Miller, 2012; ENISA, 2009), the loss of governance or ownership of the digital assets (Dečman & Vintar, 2013; Lindström, 2011; Convery, 2010), disclosure (Convery, 2010; Gellman, 2009), loss of evidential value of information (NAA, 2014), loss of reputation (Dečman & Vintar, 2013; Convery, 2010), loss of levels of service or availability (Aitken et al., 2012).

2.2.6. Organizational and regulatory issues relevant for distributed digital preservation

2.2.6.1. Institutional setting and digital preservation policies

Organizational and regulatory aspects underpinning digital preservation can be located in two main strands: the institutional setting and external regulations influencing the activity. The institutional setting in which the preservation activity takes place includes several of the high-level factors with impact in the preservation planning as described by Becker et al. (2009). For instance, the authors consider the mandate of the repository, the designated community, legal, operational and preservation policies, organizational procedures and workflows, contracts and agreements as relevant from this point of view.

Similarly, the TRAC checklist (RLG-NARA Task Force, 2007) considers several elements as part of the organizational infrastructure, such as governance; organizational structure; mandate or purpose; scope; roles and responsibilities; policy framework; funding system; financial issues, including assets; contracts, licenses, and liabilities and transparency.

Additionally, regulations such as copyright and data protection legislation need to be considered to ensure compliance with the external environment of the institutions.

2.2.6.2. The relevance of reference models

Formal reference models provide with common frameworks and terminology for the design of digital preservation architectures, serving as a foundation for setting up organizational structures, auditing the performance of the systems or benchmarking the common practices (Lavoie & Dempsey, 2004). Beagrie & Jones (2008) also point out their usefulness to break down the processes of the digital lifecycle management to calculate and allocate the costs involved in the activity.

The reference model for an Open Archival Information System (OAIS) was made available in 1999 by the Consultative Committee for Space Data Systems (CCSDS) and in 2003, the model was adopted as ISO 14721:2003. OAIS constitutes the most broadly accepted reference model for archival systems in the scope of digital preservation (Becker et al., 2009), and is very often claimed by many institutions as the framework to which their archival systems comply with. The main goal of an OAIS is to "preserve information for a designated community over a long period of time" (p. 149), implying that preservation is a major feature of archival systems, considered at the same level as other functions and activities of the archive.

As a functional framework, OAIS serves to identify and define the main components, features and data flows within a digital archival system, but not recommending any particular type of implementation. OAIS establishes a common set of terms and concepts and three models for the preservation of digital assets that include a minimum number of responsibilities in each case (Wittek & Darányi, 2012). The information model, with different information packages[19] related to the digital object to be ingested in the archive. The functional model is composed by seven functional entities that represent the archive's functions. And the environment model, which is the one defining the generic roles interacting with the archive: producers, consumers and management. Figure 1 shows how the different parts of the model relate to each other.

[19] Archival Information Package (AIP), Submission Information Package (SIP) and Dissemination Information Package (DIP).

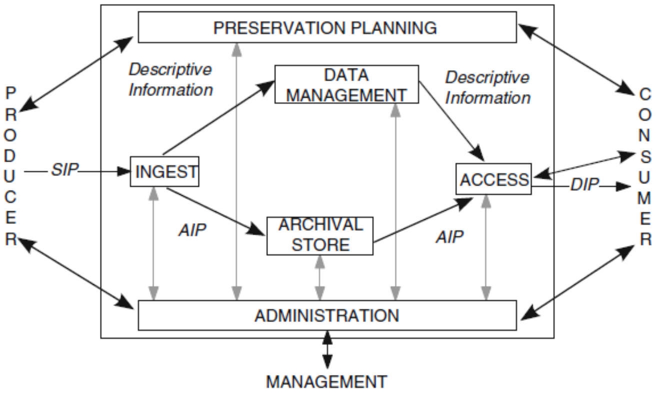


Figure 1 OAIS Functional Model (Wittek &Darányi, 2012)

Nevertheless, OAIS does not seem to map completely in the case of distributed architectures. Ruusalepp, Justrell, & Florio (2014) assert that the implementation of OAIS is designed for a single archival infrastructure; they claim that other formal reference models describing distributed digital preservation services do not exist yet and the reason seems to be related with the fact that practices of distributed service architectures are still emerging at the current moment.

Askhoj, Sugimoto & Nagamori (2011) argue that distributed digital preservation require that services can be abstracted in different layers, allowing different functions to be outsourced to a third-party. Nevertheless, OAIS functional entities are interdependent, making it difficult to isolate them and entrust them as services to external parties without creating overlapping processes in the model. Askhoj et al., (2011) propose a framework using the concepts present in OAIS, mapping them to a structure of services in the cloud. The model ranges from a lower level representing the underlying infrastructure to different levels where the management and access to the digital object would occur. Figure 2 shows how the layered model using cloud services would map with OAIS functions.

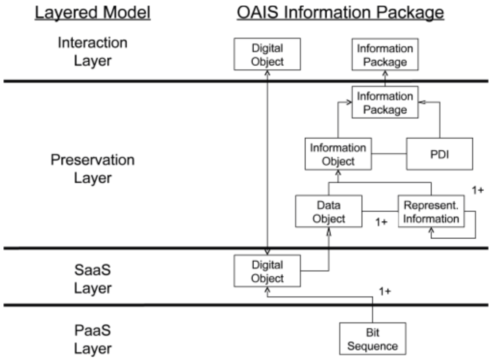


Figure 2 Layered framework for adapting OAIS to cloud (Askhoj et al., 2011)

Having evidenced the lack of conceptual models and common vocabulary specific for distributed digital preservation and the difficulties on the application of OAIS to such models, Zierau & Schultz (2013) have been working on the development of a *Framework for Applying the Reference Model for an OAIS to Distributed Digital Preservation*. The aim of such framework is to serve as an aid for future analyses and assessment of repositories performing digital preservation in a distributed manner, and thus as a base to build “effective, reliable, and auditable distributed preservation environments” (p.1).

Other initiatives working in the same line are projects such as LTDP, SHERPA DP or SHAMAN. *Long Term Digital Preservation Reference Model* (LTDP)[20] is developing a reference model for distributed architectures for digital preservation by gathering the different practices that organizations use to extend OAIS. The projects SHERPA DP and SHAMAN have been investigating the design of shared or distributed preservation environments to enable preservation activities to be outsourced to third parties, using a framework based on OAIS. In the case of SHAMAN, the purpose was to create a framework for preservation that could be verifiable, open and extensible (Innocenti et al., 2009), using a layered model that includes features not present in OAIS.

2.2.6.3. Economic sustainability, costs and funding models

[20] http://www.ltdprm.org/

Economic sustainability

Chowdhury (2013) states that economic sustainability has remained a major challenge in the digital preservation research and that few research projects have identified economic aspects of digital preservation as the major area to date.

The Blue Ribbon Task Force (2008, p. 19) defines economically sustainable digital preservation as a "set of business, social, technological, and policy mechanisms that encourage the gathering of important information assets into digital preservation systems, and support the indefinite persistence of digital preservation systems, enabling access to and use of the information assets into the long-term future." The requirements to achieve economic sustainability stated by this report, focus on the recognition of the benefits by decision-makers and incentives for them to act in the public interest; the need of collection development for long-term retention; and the mechanisms to secure an ongoing, efficient allocation of resources as well as an appropriate organization and governance of digital preservation activities.

In the final report (Blue Ribbon Task Force, 2010) there is an account of proposed actions to address the structural challenges of long-term sustainability of the preservation business. The challenges identified that affect digital preservation strategies are long time horizons, diffused stakeholders, misaligned or weak incentives, and lack of clarity about roles and responsibilities among stakeholders. With the aim of addressing those risks and making the digital preservation activity able to persist over time, the report's main finding states that "sustainable economics for digital preservation is not just about finding more funds. It is about building an economic activity firmly rooted in a compelling value proposition, clear incentives to act, and well-defined preservation roles and responsibilities" (Blue Ribbon Task Force, 2010, p. 7).

Chowdhury (2013) proposes an integrated model based on the three forms of sustainability in the scope of digital information services, including the social, economic and environmental aspects of the digital information. He argues that the target of economic sustainability is to ensure that the digital information service provides "cheaper, easier and better access to information" (p. 605) and therefore, an indicator to measure the success would be the reduction of costs. The dimension of social sustainability implies ensuring equitable access measured through the use and impact of information; and from the point of view of environmental sustainability, reductions on the impact on the environment are considered (Chowdhury, 2013). Figure 3 shows a representation of the model, including internal and external factors that may influence the different aspects of sustainability.

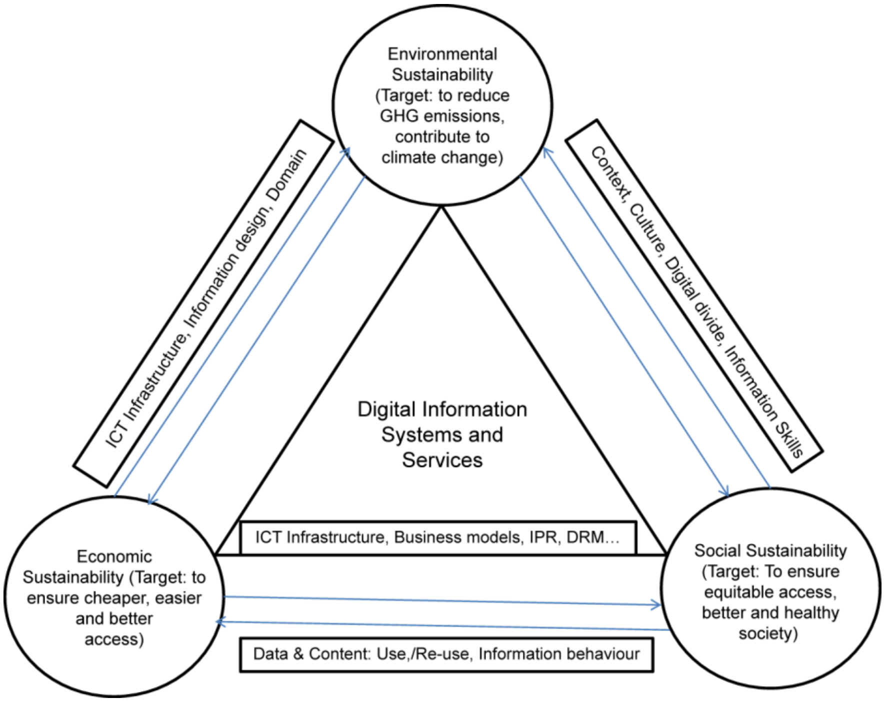

Environmental Sustainability (Target: to reduce GHG emissions, contribute to climate change)
ICT Infrastructure, Information design, Domain
Context, Culture, Digital divide, Information Skills
Digital Information Systems and Services
ICT Infrastructure, Business models, IPR, DRM…
Economic Sustainability (Target: to ensure cheaper, easier and better access)
Social Sustainability (Target: To ensure equitable access, better and healthy society)
Data & Content: Use,/Re-use, Information behaviour

Figure 3 Sustainability model for digital information systems and services. (Chowdhury, 2013)

Funding models

A major aspect of economical sustainability depends on how the allocation of resources for digital preservation is done. The activity of digital preservation requires a reliable long-term support for the ongoing programs that the soft-funding scenario used to finance the digital preservation activities is not able to provide (Sanett, 2013).

Funding models in digital preservation have been often characterized for allocating funds in a temporary basis supporting the activity through grants or special projects, instead of ensuring an ongoing stream of fund over the long-term (Lavoie & Dempsey, 2004). This idea is also stated by Sinclair et al. (2011) based on the type of budgets assigned by memory organisations to the activity. The prevalence of the use of capital-only budget compared to revenue-only reflects the fact that "they operate under funding models where it is easier to obtain grants for individual projects than a long-term commitment from a funding body to support on-going investment." (p.

279). Moreover, Sinclair et al. (2011) also state that it may be a reflection in the first case of organisations starting their activities of digital preservation, and therefore in the need of a high capital expenditure to put the solution in place, whereas revenue budgets should therefore increase over time to support on-going maintenance.

Two strategies for funding the activity are stated by Lavoie & Dempsey (2004). One of them would be the institutional commitment to budget an ongoing supply of funds, whereas the other would be based on the idea of digital preservation as self-sustained activity, generating revenues.

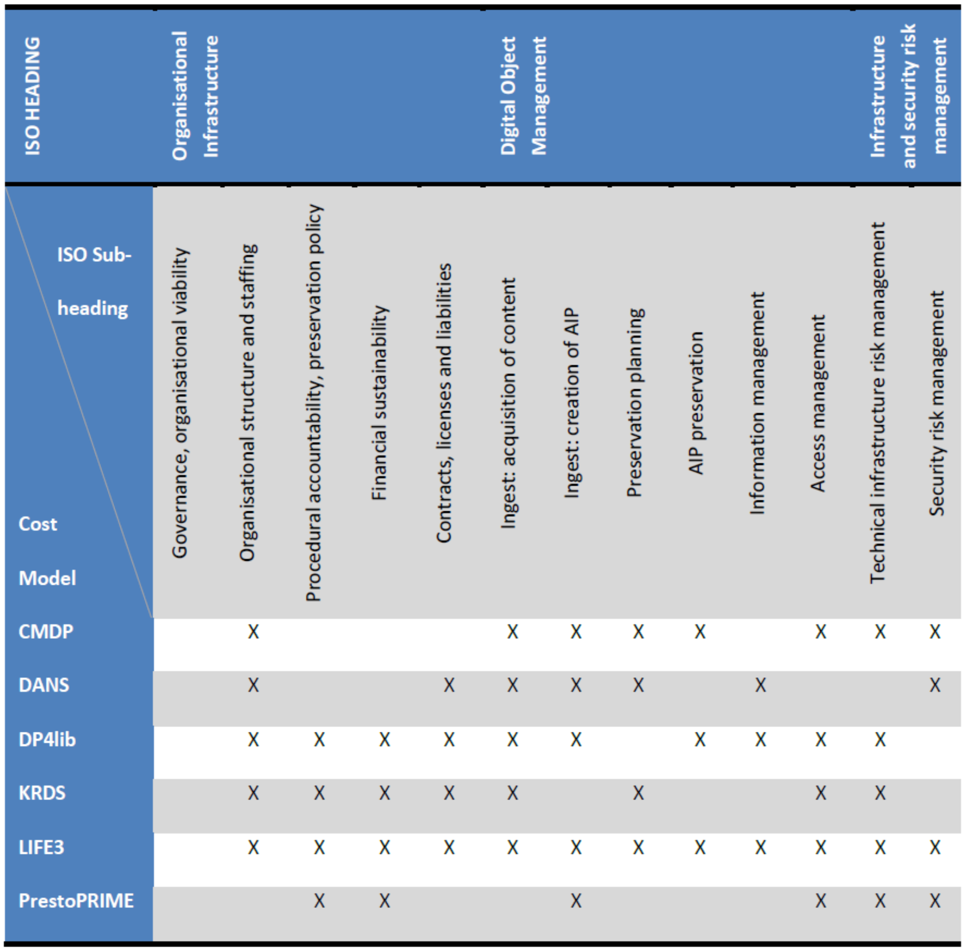
Cost models

Besides allocation of resources, building strategies for economic sustainability requires an empirical basis, and data on the costs of digital preservation is needed (Lavoie & Dempsey, 2004). In the field of digital preservation, aspects related with costing digital preservation have been developing substantially in the last few years (Rosenthal et al., 2012), and there are different frameworks that have been designed to allow institutions to establish the costs of the different elements that are needed in digital preservation.

The APARSEN project conducted a series of studies to evaluate and test cost models in use in digital preservation activities. To do so, they conducted a survey among research libraries in Europe, reviewed the published cost models and map them with the International Standard on Audit and Certification of Trustworthy Digital Repositories (ISO 16363) in order to "show how cost model parameters are concentrated or where areas of activity are not included within a particular cost model" (Kaur, Herterich, et al., 2013). The study points out that OAIS principles are the basis for the analysed models, and tailored afterwards to the needs of the institution creator of the model, which make them difficult to re-use (Kaur, Herterich, et al., 2013).

The gap analysis conducted revealed possible areas to expand the models, particularly in aspects related to organizational infrastructure or risk and security (Kaur, Darby, et al., 2013). In the case of organizational infrastructure, none of the models reviewed addressed aspects such as governance and organizational viability, and other aspects as described by ISO 16363 were mapping partially. Table 1 displays the cost models used in the analysis and the gaps identified.

Table 1 Summary of Cost Models and gaps identified (Kaur, Darby, et al., 2013)



ISO HEADING	Organisational Infrastructure					Digital Object Management						Infrastructure and security risk management	
ISO Sub-heading Cost Model	Governance, organisational viability	Organisational structure and staffing	Procedural accountability, preservation policy	Financial sustainability	Contracts, licenses and liabilities	Ingest: acquisition of content	Ingest: creation of AIP	Preservation planning	AIP preservation	Information management	Access management	Technical infrastructure risk management	Security risk management
CMDP		X				X	X	X	X		X	X	X
DANS		X			X	X	X	X		X			X
DP4lib		X	X	X	X	X	X		X	X	X	X	
KRDS		X	X	X	X	X		X			X	X	
LIFE3		X	X	X	X	X	X	X	X	X	X	X	X
PrestoPRIME			X	X			X				X	X	X

The 4C Project carried out an evaluation of cost models in the field of digital curation (Kejser et al., 2014), aiming to investigate how well those models meet stakeholders' needs in order to calculate and compare financial information. This evaluation was used in the study to point out the gaps that needed to be bridged, allowing a better uptake of cost and benefit modelling as well as cost comparison in different scenarios.

In order to determine the costs of the activity the use of concepts such as benefit and value is also relevant. Kejser et al. (2014) describe a benefit model as a representation describing the benefits and value incurred by digital curation activities. The benefits in this case may typically be, on one hand, financial benefits such as those that can be expressed in monetary values and might generate revenues from the activity. In the other hand the report considers non-financial benefits, which would be those in the form of an organisations' increased reputation or reduced business risks.

2.2.6.4. Contracts and agreements with third-parties

The establishment of a relationship with third-parties entrusting to them the provision of applications, infrastructure or services, implies that some elements of control are also handed over to the service providers (Vincent, Hart, & Morton, 2011). In the case of cloud services, contracts have turned to be the one of the key determinants for the regulation of the relationship between the parties involved.

Contracts regulate the terms and conditions of the relationship among the stakeholders involved. Moreover, Jordan et al. (2008) also points out the need for specific agreements to govern the processes in case the collaborative relationship dissolves, such as data transfer to and from the partner institution.

Service Level Agreements (SLA) are one of the most common instruments besides contracts in the relationship between the service providers and the organizations using the services because they establish how the service will be deployed. SLAs set the ground for the expectations and obligations around the service. Kyriazis (2013) details some of the main attributes of a SLA, despite there is not a fix framework for what the agreements should include which are usually related with the levels of service and protection, responsibilities assumed in case of non-compliance and additional conditions to modify the service provided.

The Quality of Service (QoS) attributes detail the goals of the service and the expectations through the description of actions that need to be taken in order to deliver the service. Qualitative and quantitative measures can be included when appropriate. The Quality of Protection (QoP) attributes detail privacy and protection constraints. The responsibilities may be specified through the inclusion of obligations of parties including penalties and exclusion terms (Kyriazis, 2013).

A critical aspect around SLAs is the fact that despite they provide a framework for expectations and responsibilities, verification of compliance and mechanisms for the enforcement might be a need for the customers. Kyriazis (2013, p. 2) claims that there is a "need for supporting tools and mechanisms used during different phases of the SLA lifecycle, such as monitoring of service execution adherence to the agreed terms and enforcement through triggering of actions to support emerging requirements."

2.2.6.5. Copyright and related right issues

Intellectual property rights and specially the regulation on copyright in particular, have a significant impact on the digital preservation activities. It is therefore a key issue to understand the requirements and to develop mitigation strategies against potential legal risks (D. Anderson, 2013) avoiding possible breaches of the regulation or even facing liabilities. Copyright legislations are currently under review in several countries, yet the ambiguity of copyright laws in relation with digital materials (Lavoie & Dempsey, 2004) has commonly been a source of impediments for digital preservation and the authors stated that achieving a balance between the interests of content providers and the institutions may be achieved through the use of formal agreements.

Legislation on copyright varies from country to country, even showing a lack of consistency within the EU member states (D. Anderson, 2013) that are subject to directives of harmonization of internal regulations. A study conducted in 2008 (Besek et al., 2008) analysed the challenges for digital preservation within legislative framework on copyright in different countries: the United States, the United Kingdom, Australia and the Netherlands. Despite copyright legislation is being updated in countries such as the UK, some of their conclusions are still valid in the current context. The authors acknowledge that although copyright was not the only barrier to digital preservation, it introduced challenges in the four countries. Although all of them had exceptions on their laws that allowed reproduction of protected works, those exceptions did not accommodate well to the needs of digital preservation. In this framework, different strategies have been taken by the memory institutions surveyed and for instance, it was reported as a very common one the establishment of collaborative agreements with the right holders.

Notwithstanding, strategies to ensure digital preservation involve processes and actions that might not have a good fit with the legal framework in terms of copyright protection. For instance, the processes might involve the creation of additional copies for different purposes, alter the content in some way, such as through the migration of formats or disaggregating the original object in different parts (Lavoie & Dempsey, 2004). In the case of emulation, copies would be done for the purpose of media transfer, moving the files from their original storage medium to a managed storage system put in place by the memory institution (D. Anderson, 2013) The distribution of copies in different storage locations may be a potential source of regulatory breaches.

2.2.6.6. Data protection and the right to privacy

Data protection laws regulates how information related to individual persons is processed including their collection, storage or dissemination. One of the aspects protected is the right to privacy, concerning personally identifiable information that is collected and stored, whether it is in digital form or otherwise (Jøsang, Fritsch, & Mahler, 2010).

Bygrave (as cited in Jøsang et al., 2010) identified a set of basic principles that can be found in most data protection regulations, both at international or national level. Data subject of fair and lawful processing; purpose specification, which means that personal data must be collected for specified, explicit and legitimate purposes and not further processed for other purposes; minimality on the collection and storage of data, limited to what is needed for the purpose; ensure quality of the data; data subject participation and control; limitation of fully automated decisions; disclosure limitation related to third parties; information security limiting unauthorized access, alteration, destruction or disclosure; and processing certain categories of especially sensitive data is subject to a stricter control than other personal data.

The use of cloud technologies has raised concern in relation with the potential lack of compliance on some of those principles. Guidance and risk registers, especially those related with security issues, such as de case of the report published by ENISA (2009) addresses the main concerns in the area. Nevertheless, the report published by World Privacy Forum (Gellman, 2009), specially focused on data protection issues provides a very detailed account of potential threats of using public clouds.

2.2.7. Risk management in digital preservation

In the scope of digital preservation, risk management concepts have been used as a managerial tool to assist on the decision making processes through gaining self-awareness on what needs to be improved (McHugh, 2012), but also to assess repositories and their organizational framework to gain the assurance of their trustworthiness.

A major development for risk management was the publication in 2009 of the ISO standard 31000 (ISO, 2009). The text defines risk as the “effect of uncertainty on objectives” (p. 1) and risk management as “coordinated activities to direct and control an organization with regard to risk” (p. 2). Nowadays, many disciplines are adopting this approach and so is doing digital preservation also.

In the field of digital preservation, the project ERPANET was the pioneer on the introduction of risk management with the publication of a risks communication tool. ERPANET (2003) understood that the activity of digital preservation being framed by a continuous technological change implies a constant risk. The nature of digital assets themselves brings in a large number of uncertainties, increased by the perspective of the commitment over the long-term.

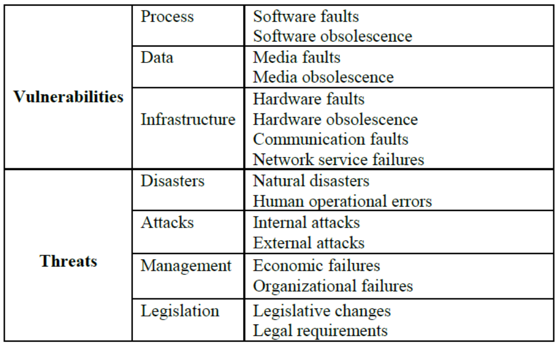
Nevertheless, one of the major contributions on the scope of risk assessment for digital preservation was the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA), completed in 2007 by DigitalPreservationEurope (DPE) and the Digital Curation Centre (DCC). The methodology’s aim is to guide digital repositories on a self-assessment to assist them on the transformation of “controllable and uncontrollable uncertainties into a framework of manageable risks” (DCC & DPE, 2007, p. 11).

The use of risk management techniques have been used to develop other tools to support decision-making processes. For instance, the EU-funded project called TIMBUS has defined a risk management process whose aim is to align the activity of digital preservation with risk management, using the ISO 31000 as the foundation to build upon.

Barateiro et al., (2010) propose a risk management approach to assist on the assessment of digital preservation solutions, but also on the design of new ones. The rationale for this approach resides on the idea that digital preservation aims to protect digital objects from the various threats affecting their future interpretation, which means reducing the risks of those threats. In their study, they not only suggest a process for risk management, but also they survey the main requirements for digital preservation, define a taxonomy of threats and vulnerabilities and detail a set of techniques that can be used for the treatment of risks. The taxonomy of vulnerabilities and threats to digital preservation is shown on Table 2.

The proposed risk management process requires the definition of context and requirements, the identification of threats and vulnerabilities in order to address potential threats and vulnerabilities (Barateiro et al., 2010).

Table 2 Taxonomy of vulnerabilities and threats to digital preservation (Barateiro et al., 2010)



Vulnerabilities	Process	Software faults Software obsolescence
	Data	Media faults Media obsolescence
	Infrastructure	Hardware faults Hardware obsolescence Communication faults Network service failures
Threats	Disasters	Natural disasters Human operational errors
	Attacks	Internal attacks External attacks
	Management	Economic failures Organizational failures
	Legislation	Legislative changes Legal requirements

An example of the application of risk assessment applied to the field of digital preservation can be found in Zierau et al. (2010). They carried out a study for the development of a methodology to evaluate bit preservation strategies for different bit repository solutions, including distributed architectures, defining the requirements in terms of relative importance of the risk preventions. Nevertheless they acknowledge that there would be alternative ways to formulate requirements considered for their study, for instance taking technical and organizational levels more in detail and including standards like TRAC, ISO or DRAMBORA.

2.2.8. Trust in the scope of distributed digital preservation

2.2.8.1. Trust in digital archives

Walters & McDonald (2008) claim that trust in relation to digital preservation in memory institutions means that they are trusted by their stakeholders “to maintain the digital library or archives to sustain the information deposited in it, and that this information remains authentic, reliable, and unchanged over time and across technologies” (p. 1). Dobratz & Schoger (2007) argue that “trustworthiness of a system means that it operates according to its objectives and specifications (it does exactly what it claims to do).”

The ability to demonstrate trustworthiness of digital archives and the assets they host is a major concern in the scope of digital preservation. Operational aspects such as quality and security of the digital objects need to be maintained and proved by the archives, in terms of their reliability, authenticity, integrity, interpretability, provenance, confidentiality or availability (Dobratz & Schoger, 2007). But despite operational issues are of a major importance organizational aspects cannot be overlooked.

In the described context, (Dobratz & Schoger, 2007, p. 212) defined several groups of stakeholders for whom trustworthiness might be relevant:

- repository users who want to access trustworthy information – today and in the future,
- data producers and content providers for whom trustworthiness provides a means of quality assurance when choosing potential service providers,
- resource allocators, funding agencies and other institutions that need to make funding and granting decisions, and
- long-term digital repositories that want to gain trustworthiness and demonstrate this to the public either to fulfil legal requirements or to survive in the market.

Assessment and certification processes have been pointed out as a basis to demonstrate trust and there are already a few initiatives available that will be detailed later on in this report. One of the earliest claims was made by the Task Force on Archiving Digital Information (Waters & Garrett, 1996, p. 9), stating that “repositories claiming to serve an archival function must be able to prove that they are who they say they are by meeting or exceeding the standards and criteria of an independently administered program for archival certification”. Different investigations followed this idea and different set of criteria to establish trustworthiness as well as risk assessment methodologies suitable for digital preservation were defined (Dobratz & Schoger, 2007).

2.2.8.2. Inter-organizational trust

As was already discussed earlier in this report, the need for sustainable models for digital preservation has been increasing the cooperation of different types of organizations in the pursuit of effective and efficient means for the storage, preservation and curation of unique digital information (Walters & McDonald, 2008). Some may use cooperatives and some others engage in contractual relationships with third-party service providers to fulfil the needs of the activity to ultimately develop strategies and systems to preserve digital information.

In this context and to allow viability and healthiness of relationships between the organizations involved in the partnerships, confidence among them seems to be a decisive factor. Walters & McDonald (2008) claim that it is critical for designing digital preservation systems to consider the concept of trust as an essential element, particularly considering inter-institutional relationships. It seems reasonable that institutions partnering with others in the preservation of their digital assets need an answer to whether the actions taken by their partners are trustworthy or not. Day (2008) argues that, trust creates the foundation for a successful co-operation, understanding trust as "a concept that is typically defined in terms of confidence in the actions, intentions or goodwill of other parties within a given context" (p. 21). Therefore, trust might determine the success or failure of those inter-institutional relationships.

It is clear that a partnership or contractual relation with other organizations introduce some level of risk or vulnerability. The acceptance of vulnerability in exchange for the potential benefits, the development of trust over time (Day, 2008) and risk assessment and management (Walters & McDonald, 2008) are common strategies used by the parties involved in those relationships.

To allow institutions to build those relationships based on trust, Walters & McDonald (2008) outline a trust model for distributed digital preservation, based in the governance and trust models developed by previous studies focused on federated relationships. They identify two levels needed to support trust. On one hand, frameworks and models supported by formal mandates between the organizations involved. And on the other, business models that would develop trust relationships within partnerships. The establishment of contracts, evidence based practice and organizational structure analysis are relevant supports to trust in inter-organizational relationships.

2.2.9. Control mechanisms in digital preservation

Castelfranchi & Falcone, (2000) argue that control mechanisms can themselves build trust. Following this approach, the use of control mechanisms has been therefore a central issue for the recognition of trust in digital preservation.

To understand what does control refer to in this particular context, (Day, 2008, p. 21) defines it as “processes that are used to monitor and enforce activities, e.g. through things like governance structures, contracts or adherence to standards”. He also differentiates between two types of control mechanisms: formal control mechanisms, such as rules, policies and procedures, supported by monitoring and measurement of organization’s processes or outcomes; and informal value-based control, supported by the creation of shared organisational cultures that encourage certain behaviours and outcomes.

The main trends on the use of control mechanisms in the digital preservation domain to support trustworthiness on the activity have been certification and self-assessment. But it is relevant to notice that most discussions about trust have focused on the development of criteria for the evaluation of repositories and other preservation services (Day, 2008).

The need for a certification process was early detected on pioneer reports starting discussions on the field such as the Task Force on Archiving of Digital Information or the Working Group RLG-OCLC on Trusted Digital Repositories (RLG & OCLC, 2002). They suggest the need of a certification process on digital archives that would help to establish a climate of trust on the digital preservation realm.

In the case of using third-party services the working group (RLG & OCLC, 2002, p. 9) argues that service providers may gain the trust of memory institutions “through a combination of proven reliability, fulfilment of contractual responsibilities, and demonstrated sensitivity to community issues.” Despite these attributes can be measured, engagement with third-parties does not seem likely to happen if the reliability is not proven a priori. Therefore, they suggest that a program for certification might be a useful basis over which build trust upon.

2.2.10. Audit and assessment methods in digital preservation

Several instruments have been made available in the field of digital preservation to assist managers to conduct self-assessments, identify weaknesses and help to improve capabilities over the long-term (Downs & Chen, 2010). Risk management is the base for tools such as DRAMBORA, and the assurance of trustworthiness of systems and organizations has been the basis for the development of a number of other frameworks to which evidence-based practice is essential (Ross & McHugh, 2006).

2.2.10.1. DRAMBORA

The *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA) process focuses on risks, and their classification and evaluation according to the activities, assets and contextual constraints of individual repositories (Barateiro et al., 2010, p. 7).

Is was conceived as a tool for self-assessment on the risks that may serve to support effective management (McHugh, Ross, Innocenti, Ruusalepp, & Hofman, 2008), such as directing resources specifically to identified areas of concern. Moreover, it can be used as a control mechanism that may be also helpful in the identification of the particular risks of working with third-party services within collaborative networks or to prepare the organization for external audits (Day, 2008).

The methodology is organized in a series of steps that the organizations have to follow fulfilling the requirements established in each of them through the completion of structured exercises. Organizations are required to “expose their organization, policies and infrastructures to rigorous scrutiny” (p. 131) achieving a complete registry of their most pertinent risks, as the final output of the exercise (McHugh et al., 2008).

2.2.10.2. TRAC

A major work in the quest for establishing trust in digital repositories is the *Trustworthy Repositories Audit & Certification: Criteria and Checklist* (TRAC), released in 2007. It was developed by the *Digital Repository Certification Task Force* constituted by OCLC's Research Libraries Group (RLG) and the National Archives and Records Administration (NARA).

The major aim was to produce certification criteria and delineate a process for certification applicable to a range of digital repositories and archives, including academic repositories, data archives, national libraries or third-party digital archiving services, to name a few of them (RLG-NARA Task Force, 2007).

The identified criteria help to identify potential risks to digital content held in those archives. TRAC is built using a top-down approach using the Open Archival Information System (OAIS) Reference Model as foundation (Smith & Moore, 2007) and to benchmark against it to determine how successful a repository is in terms of trustworthiness (Barateiro et al., 2010).

The checklist has 84 requirements divided into four categories related to the organization supporting the digital repository; repository functions, processes and procedures; the designated community and usability of information in the repository; and technologies and technical infrastructure of the repository (Smith & Moore, 2007). A set of methodologies are provided to determine the "soundness and sustainability of digital repositories" (Barateiro et al., 2010, p. 7).

Some criticism on the representation of TRAC criteria has been also argued as constraining digital preservation to be analysed in silos (Becker, Barateiro, Antunes, Vieira, & Borbinha, 2011) limiting the possibilities of digital preservation to give holistic and multidimensional view of the problems and to respond therefore to the concerns of multiple stakeholders in the organization.

Based on the TRAC checklist, ISO 16363:2012 "Audit and certification of trustworthy digital repositories" was also developed under the ISO family of standards, by a working group at the Consultative Committee for Space Data Systems (CCSDS). Additionally, the ISO 16919 "Requirements for bodies providing audit and certification of candidate trustworthy digital repositories" was also carried out.

2.2.10.3. nestor

The *Catalogue of Criteria for Trusted Digital Repositories* (2006) was elaborated in Germany by a working group of the *nestor* initiative. The *nestor* catalogue was created over the foundations of previous works such as the RLG-NARA checklist, adapting to the requirements specific to the German context (Dobratz & Schoger, 2007), but also feeding back for further development of the TRAC methodology (Day, 2008).

The comprehensive list of criteria was defined at an abstract level (Dobratz & Schoger, 2007), and similarly to TRAC, the criteria was organized in three groups: organizational framework, object management and infrastructure and security. For the development and application of the criteria the following concepts detailed by (Dobratz & Schoger, 2007) are of relevance:

- Accordance to OAIS terminology, to define concepts related to the digital repositories, objects; from processes to the life-cycle.
- Abstraction, to allow a variety of archives to use it and over the long term.
- Documentation of concepts, goals, specifications and implementation in a proper manner.
- Transparency, towards stakeholders and internally, relying on documentation.
- Adequacy to the particular context on which the assessment made.
- Measurability and reliability in indicators of trustworthiness.

The work developed in *nestor* was also transferred to create the standard DIN 31644:2012 “Information and documentation - Criteria for trustworthy digital archives.” And based on the standard, *nestor* also offers the *nestor Seal*, an extended self-assessment with a process of revision that would certify if repositories are indeed trustworthy based on the criteria established (Harmsen et al., 2013).

2.2.10.4. Data Seal of Approval (DSA)

The Data Archiving and Networked Services (DANS) in the Netherlands also published sixteen guidelines, categorized in three groups related to data producers, repositories and data consumers. The process consists in two stages, a self-assessment and a peer review process[21].

Additionally, a framework for audit and certification of digital repositories has been developed since 2010 after the signature of a Memorandum of Understanding. It consists of three different levels of assessment, starting from a basic certification granted to repositories that achieve the DSA; an extended certification for those archives that have been granted with the basic certification and that additionally perform a externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644. Moreover, formal certification would be granted to

[21] http://www.datasealofapproval.org/en/

repositories which in addition to a basic certification obtain full external audit and certification based on ISO 16363 or equivalent DIN 31644[22].

2.3. Conclusion

This chapter has presented an overview of the main issues considered relevant to contextualize the activity of digital preservation. From a more general overview of what difficulties digital preservation entails and therefore motivates collaboration and distribution of responsibilities, an overview of the services that currently in use, to the major issues related to organizational and regulatory aspects that affect digital preservation and the concepts of risk and trust, including inter-organizational trust as they are the lens through which distributed digital preservation has been investigated in the present study.

[22] http://www.trusteddigitalrepository.eu/

Chapter 3. Methodology

3.1. Introduction

This chapter presents the research approach and methods used for the present study. Moreover, it also gives an account on how the data analysis has been performed and other aspects such as how trustworthiness of the research is claimed, limitations and ethical considerations.

3.2. Research Approach and Strategy

3.2.1. Qualitative methodology

To understand how the research has been conducted, it is useful to contextualize it within the research perspective used as a lens to approach the issues investigated. Using the approach stated by Mason (2012) the ontological position adopted in this research has the underlying assumption that “people’s knowledge, views, understandings, interpretations, experiences, and interactions are meaningful properties of the social reality” (p. 63) explored. Moreover, the epistemological position adopted allowed this research to consider that “a legitimate or meaningful way to generate data on these ontological properties is to talk interactively with people, to ask them questions, to listen to them, to gain access to their accounts and articulations, or to analyse their use of language and construction of discourse” (p. 63-64).

Therefore, following those ontological and epistemological principles a qualitative approach was followed in this research, considering that the best way to understand the phenomena under study was to put the emphasis in gathering the perspective of the insiders (Lapan, Quartaroli, & Riemer, 2012). Strauss & Corbin, (1990, p. 12) state that “qualitative research allows researchers to get an inner experience of participants, to determine how meanings are formed through and in culture, and to discover rather than test variables”.

The main features of qualitative studies are the capability to implement a flexible design (Robson, 2002), that will be evolving along with the data collection; the presentation of multiple realities; the researcher as an instrument of data collection and a focus on participants’ views

(Creswell, 2007). Pickard (2007) claims that the emergent nature of qualitative research is not the best fit for a detailed plan before the beginning of the research. However, she also acknowledges that it is possible to developed one accordingly to the iterative nature of the study.

Another significant feature is that the data gathered using qualitative methodology is commonly in the form of words, non-numerical. Therefore, qualitative researchers usually find their way to gather the multiple perspectives through the use of research methods such as interviews and observation (Robson, 2002).

Summing up, considering all the stated above, this study took the qualitative approach. The main source of data was interviews with experts from organizations dealing with the topic under study. The chosen method allowed for a degree of flexibility in designing the data collection and analysis that was desirable, given the domain under study (risk analysis in cloud-based preservation services) that is only emerging.

3.3. Data Collection and Processing

3.3.1. Sampling

3.3.1.1. Purposive sampling

The selection of participants for the data collection exercise in this study was undertaken in a strategic way, thus neither accidental nor in an attempt to be representative of a wider universe (Mason, 2012). The aim was to reflect the diversity within the phenomena under study and not to select typical cases (Barbour, 2008).

A priori criteria sampling was one of the techniques chosen because it was found to be useful to establish a predefined framework attending to a predefined criteria, as will be explained further below. Pickard (2007, p. 64) describes this type of sampling as a “trade-off between a totally emergent design and a more structured a priori design but it also allows for an element of inductive design within the framework that is created.”

Therefore, the selection of institutions and the experts within their boundaries was attained attending to particular characteristics (Mason, 2012; Pickard, 2007), in order to be able to address the aspects set as aims and research questions of this study and to establish cross-contextual comparisons. In fact, as Barbour (2008) acknowledges, the potential for the analysis and the establishment of comparisons of the data gathered is determined by the sampling.

3.3.1.2. Sampling criteria

With the purpose of seeking a relevant range of participants and "to ensure that each new research participant contributes characteristics differing from preceding participants" (Pickard, 2007, p. 14), the first step taken was to establish the set of requirements of each type of institutions that should be included. The criteria taken in consideration were based in three groups of characteristics.

- Type of organization, considering as that the participant's organizations were either service providers or memory institutions in the domain of digital preservation.
- Type of relationship. The requirements to consider that the participants are in the scope of distributed digital preservation involve relationships of contracting third-party services or collaborating with other institutions.
- Dimensions giving form to the relationships established between memory institutions with a service provider or with their peers. The options considered where organizations taking part of a cooperative, a centralized service provider for a network, an institution using cloud or grid technologies, a repository outsourcing digital preservation functions or a cloud or grid service provider offering digital preservation functions[23]. Table 3 summarizes the criteria used and Figure 4 displays the potential relationships between the organizations using distributed architectures for their digital preservation activities.

The main strength of the sampling design is the fact that it considers the different perspectives of the actors involved when the digital preservation activities are distributed, as is the case of service providers and customers. Therefore, the selection of participants was done with the idea of obtaining a complete variety of perspectives, being able to gather contradictory or overlapping perceptions and nuanced understandings that the different individuals interviewed may hold (Rubin & Rubin, 2005).

[23] This group of characteristics was adapted from: Ruusalepp, Justrell, & Florio (2014).

Table 3 Criteria and characteristics used for the sampling

Criteria	***Characteristics***
Type of organization	
	Service providers
	Memory institutions
Type of relationship	
	Memory institutions using third-party services
	Collaboration between memory institutions
Dimensions of the service	
	Cooperative model
	Centralized service provider for a network
	Use of cloud / grid storage
	Repository outsources digital preservation functions
	Cloud or grid service provider offers digital preservation functions

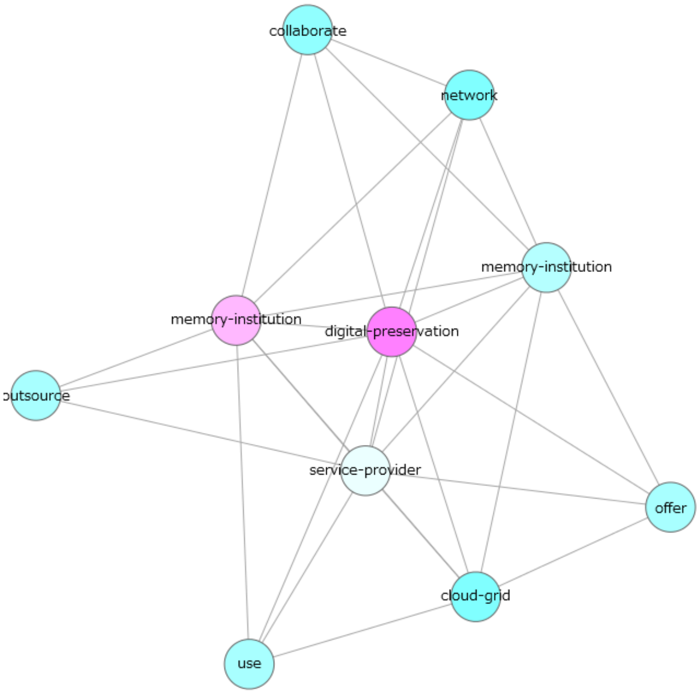


Figure 4 Potential relations between the categories used for the sampling.

3.3.1.3. Geographical area

A specific geographical area was not initially targeted for the selection of participants. Some of the issues addressed in the research questions, such as the regulatory aspects related to the activity, were initially a motivation to narrow the focus into a specific country. Nevertheless, after getting into the field and understanding the complexity of the relationships arising from the collaboration or contractual relationships between the organizations subject of study, the idea was discarded.

Within any of the organizational relationships considered in this study, partnering institutions or contractual relationships could, and actually are established in the same or among different geographical contexts. Establishing very tight criteria on this aspect would eventually have led to missing the richness of the spectrum of experiences.

The institutions selected are based in Europe, the United States and Canada. Attempts were made to include institutions also from New Zealand and Australia, but the endeavours did not succeed. The resulting geographical coverage used for the data collection, despite being broad, seems to reflect the fact that major developments in the digital preservation field are being conducted in those countries. And also reflects an informal and sometimes official network of interactions among the geographical areas specified earlier.

3.3.1.4. Size of the sample

Regarding the number of experts or representatives from institutions to interview, following Mason (2012)'s approach, it was not my first concern to make the sample size of a dimension representative of a total population and the focus was put into the achievement of inclusion of the categories defined as relevant for the data collection. Rubin & Rubin (2005) claim that not a vast number of participants is needed, but enough to obtain different points of view to provide a complete picture. Patton (1990, p. 184) also notes that "there are no rules for sample size in qualitative inquiry. Sample size depends on what you want to know, the purpose of the inquiry, what's at stake, what will be useful, what will have credibility, and what can be done with available time and resources."

The number of organizations contacted was a total of forty-eight, including service providers, libraries, archives and consortia of memory institutions, and the final number of participant organizations was a total of sixteen.

As the emphasis on the selection of participants was made on the achievement of completeness of the sample (Rubin & Rubin, 2005), in relation with the categories established during the design phase, the process of contacting and selecting participants kept progressing alongside data collection, until it was considered that an acceptable level of theoretical saturation was reached.

Nevertheless, the time constrains were a factor limiting further involvement of participants after the data analysis was completed. Despite the data collected through the interviews were considered sufficient, it could have been interesting for the study to have had additional time to fill in some of the gaps in the knowledge of the participants with complementary views from other members of their own institutions. This issue will be further explained in the section Limitations of the Research.

3.3.1.5. Identification of the interviewees

After the criteria were identified, a grid was created and each of the resulting cells were represented in the final sample, as recommended by Pickard (2007). To populate each cell, a theoretical sampling strategy was taken. This type of sampling allows the researchers to use their own judgement for selecting the cases that are considered more useful (Bloor & Wood, 2006). Additionally, in a few cases snowball sampling was also used where the identified individuals pointed out more suitable participants from their organization. Despite this combination of sampling methods can be seen as a limitation for the flexibility and emergent design characterizing qualitative studies (Pickard, 2007), the nature of this study recommended the use of a sample planned in advance.

Participants were identified and selected because of their role as experts on the field that was addressed in the study. It was considered critical that they have relevant first-hand experience on the topics. The process to do so came hand by hand with the review of the literature and additional research looking for organizations relevant in the field through collaborative projects, conference proceedings, personal contacts in the field and service providers' websites.

3.3.2. Interviews

Interviews were chosen as the primarily data collection technique because of the power to obtain in-depth qualitative information from experienced and knowledgeable people (Rubin & Rubin, 2005). In this case, the research designed relies for the gathering of the data on what Gillham (2009) calls elite interviews, which involve interviewees who are in a position of authority, expert or authoritative, capable of giving answers with insight and comprehensive grasp of what is being researched. Furthermore, due to the strengths of direct interaction, interviews allow to gather extended responses and even the disclosure of sensitive material is more likely to happen (Gillham, 2009), compared to other data collection techniques such as questionnaires.

Interviews are suitable for small samples that include participants whose opinions are highly relevant and none of them can be afforded to be lost (Gillham, 2009). They allow "interactional exchange of dialogue" (p. 62), through one-to-one conversations, that were the case of this study (Mason, 2012). Under the assumption that "knowledge is situated and contextual" (p. 62), there was an intention to bring the focus to the relevant topics so knowledge could be produced during the interaction. A thematic approach was taken, allowing participants to elaborate from a number of starting points of discussion (Mason, 2012).

3.3.2.1. Interview design

The interviews were conceived as semi-structured with open-ended questions, leaving room to the interviewees to elaborate. Designing the interviews as semi-structured was based on the idea that such a design enhances the "capacity of interviews to elicit data on perspectives of salience to respondents rather than the researcher dictating the direction of the encounter, as would be the case with more structured approaches" (Barbour, 2008).

Nevertheless, an interview structure was prepared in advance. A different set of main questions was established for the three types of participants, and same or similar questions were asked within each group sharing a similar role. The first decision was to make a differentiation between memory institutions and service providers. Additionally a third group was established and memory institutions partnering on their digital preservation activities had a slightly different interview structure. The decision of keeping a similar structure was made with the analysis of the data in mind, in the sense that it will allow the comparison of the data from the different perspectives of each group.

The selection of the interview questions was done based on the previous research on the topic and the main theories already described in Chapter 2. Literature Review. The first question was about the benefits perceived by the memory institutions distributing the digital preservation functions. The main aim was to explore which type of benefits the institutions achieve, whether they are financial or not (Kejser et al., 2014), and how that perception creates conditions for distributed digital preservation.

Three blocks of questions follow the initial one. A first block was related to the organizational and regulatory risks, based on the ideas gathered through the literature detailed on the sections of Risk management in digital preservation and Organizational and regulatory issues relevant for distributed digital preservation; a second one, about trusting the partnership or the service provider, building upon the theoretical approach stated in the sections Trust in the scope of distributed digital preservation and Control mechanisms in digital preservation of the literature review. Additionally a third block of questions was about the changes for the memory institutions after entering into the relationships to distribute digital preservation, also building upon the approaches stated in the section on Organizational and regulatory issues relevant for distributed digital preservation of the literature review. The interview questions are provided in Appendix 1: Interview structures.

Additionally, a checklist including the major risks already described in the previous research literature was created for my personal use, serving as a guide for the conversation and to introduce follow-up questions and probes when necessary during the conversation (Rubin & Rubin, 2005). This instrument helped to further explore issues that were not arising during the conversation.

3.3.2.2. Gaining access

The process followed to get access was done as described by Robson (2002), despite some of the steps were not so relevant in this particular case. Gate-keepers and participants were usually the same person, and gaining access was dependent mostly on their willingness to collaborate or time constrains, rather than other factors that would eventually need more negotiation. Despite that, in some cases due to the fact that some of the questions were sensitive in terms of confidentiality, additional negotiation was required with those organizations.

1. After selecting the relevant institutions according to the categories defined, potential participants or individuals that could provide access to the institutions were identified. To be able to manage appropriately the contact details of participants, interactions and information that was generated after those interactions, a database was created.
2. The next step was to prepare an outline of the study, sent by email to the targeted individuals having two purposes of collaboration in mind. The main aim was to request their participation in the study, but additionally this communication was done also with the purpose of getting some feedback about the outline itself and whether the focus of the study and the research questions were considered relevant for experts in the field. Some of the responses included resources or comments that were of high level of interest. Together with the presentation letter, a link to my *curriculum vitae* was provided in order to give them additional information about my background.
3. Through the exchange of emails with participants, the permission was granted and the conditions and scheduling were established.
4. The study was explained and discussed with those gate-keepers or participants that showed doubts about it or about the process, mostly related with confidentiality due to their relationships with third-parties subject to maintain secrecy obligations in some cases. An informed consent form was sent to the participants explaining how data collected was going to be treated.
5. Some aspects of the study, mostly those related with the instruments for data collection were adapted in the light of the discussions.

3.3.2.3. Collection of interview data

The collection of data was scheduled according to the participants' availability. After the communications and negotiation of access, two weeks were initially allowed for the data collection, although an extension of that time was needed to match some changes on the initial schedule. The period of time for the interviews finally lasted from May 13 until June 3, 2014.

The interviews were held online using *Skype* and through an exchange of emails, when the first was not possible. Synchronous interviews were recorded with the purpose of note-taking. A total of sixteen interviews were conducted mostly in a synchronous way (eleven) using the online voice-over-IP service, whereas the rest of them (five) were conducted through an exchange of emails with the participants. The interviews were one-to-one, except for one of the synchronous interviews which was a group interview with two participants from the same institution.

The planned length of the interviews was of a maximum of forty minutes. Despite that, the duration of interviews was kept flexible, depending on the availability and willingness of the participants. None of the interviews held lasted for more than one hour.

The tentative interview schedule was shared with each participant before the interview to allow some reflection on the topics that were going to be discussed. Despite the broad areas addressed in the interviews, they were conducted as semi-structured with a number of follow-up questions to gather as much data as possible.

3.3.3. Documents' review

The review of relevant documentation was also performed to broaden the information sources and as evidence of some of the comments provided during the interviews. Relevant documents were provided by the participants, such as internal documents like business cases or service level agreements not publicly available. In those cases, confidentiality was guaranteed.

Other documents provided were certification reports, in the case they went through self-assessment or certification processes. Some digital preservation policies were also examined, as well as the information publicly available for stakeholders in the websites of the institutions and commercial providers.

Additional documentation reviewed were papers presented to conferences explaining their experiences in aspects related with this study, and description of their cases publicly available.

3.4. Data Analysis Methods

Rubin & Rubin (2005, p. 201) state that analysis "entails classifying, comparing, weighing and combining material from interviews to extract the meaning and implications, to reveal patterns, or to stitch together descriptions of events into a coherent narrative."

With the objective of producing this coherent narrative that the authors allude to, the analysis of the data collected through the interview went basically through two steps that enabled to give to the interviewees opinions and descriptions my own interpretation (Rubin & Rubin, 2005). The first step was the transcription, identification of themes, classification and coding of what was said by participants. And the second consisted on the comparison and combination of the identified themes and concepts across the sources of evidence gathered. With the finalization of this two-step analysis, I was in a position to draw the major conclusions about the research carried out. A detailed description of the process is presented below.

3.4.1. Transcription of Interviews

The first step after conducting the interviews was to transcribe them in the following hours or days, to allow reviewing what was said and whether modifications on the way subsequent interviews should be conducted were necessary.

A full version of the recordings was made available into a text version[24] to allow the process of content analysis (Gillham, 2009).

3.4.2. Categorization of Data

The process of coding involves the selection of statements and the extraction of categories from them in order to give answers to each of the research questions posed. As stated by Gillham (2009), coding implies an iterative process of comparing the statements and categories and continuously modifying wording and even adding or deleting categories.

[24] Express Scribe software was used as an aid for the transcription http://www.nch.com.au/scribe/index.html

In order to complete this process, a qualitative data analysis tool, *Nvivo*[25], was used to assist with the development of a coding scheme and to assign the codes to meaningful statements in the texts analysed. An initial coding scheme was developed based on the literature review and topics emerging during the interview process. Followed by the stage of recognition, which is meant to identify the concepts, themes, events and topical markers in the interviews themselves (Rubin & Rubin, 2005). The codes were subsequently applied to the interview transcripts and additionally, the codes grouped and reduced to categories relating them to the research questions and other emerging topics with the objective of facilitating the analysis and reporting of findings.

3.4.3. Content Analysis

The content analysis was done following a general qualitative approach. The coding process helped to break down into data units the comments made during the interviews and enabled blocks of information being examined together (Rubin & Rubin, 2005). The next stage of analysis continued with the identification of substantive statements, discarding repetitions, digressions or irrelevant material and combining data units on the same topic, both within single interviews and across the entire set of interviews (Gillham, 2009).

3.4.4. Data Presentation

The data was presented in a narrative manner, giving descriptions on the patterns identified and exemplifying with and including as much variety of perspectives given by the participants as possible to gather the richness of their insights. Quotes from the interviewees' comments were included to support and complement the statements and as evidence of what it was being stated.

The interviewees' personal names and their institutions names were anonymized. To identify them, a set of codes were used (Int#1, Int#2, etc.). Names of other institutions and companies were stated as used by the interviewees.

Despite Int#10 was a group interview with two participants the presentation of their comments were kept under an unique identifier, as they were complementary and it was not considered relevant for the purpose of the research to differentiate among them.

[25] *Nvivo* QDA software http://www.qsrinternational.com/products_nvivo.aspx

3.5. Trustworthiness of the Enquiry

To ensure trustworthiness of the research, advice from Pickard (2007), Robson (2002) and Shenton (2004) was followed to fulfil the criteria of credibility, transferability, dependability and confirmability of the study. These aspects are commonly used in qualitative research to demonstrate the rigor of the research, in opposition to the concepts of reliability and validity more commonly used in quantitative approaches. Nevertheless, recommendations given by Yin (2003), despite done from a more positivist point of view and aimed to increase reliability and validity, were also followed to test the data collection exercise.

3.5.1. Credibility

To fulfil the criteria of credibility, the use of multiple sources of evidence (Pickard, 2007; Yin, 2003), such as the data collected through the interviews and document review, was performed increasing the robustness of the results through triangulation. Nevertheless, semi-structured interviews were used as the primary corpus of data, because of their appropriateness for in-depth exploration. As already mentioned in the section 3.3.2, interviews were recorded and transcribed to guarantee the accuracy of the data collected. Triangulation via data sources and sites, using a wide range of informants could be also argued, since the array of interviewees and organizations was also intended to allow verification against others, for instance, in the case of vendors and memory institutions outsourcing the service (Shenton, 2004).

Additionally, the engagement with participants was prolonged through a series of interactions by email, facilitating a climate of trust between the participants and the researcher and also attention was put on the "development of an early familiarity with the culture of participating organizations" (Shenton, 2004, p. 65) before the interviews were conducted.

The adoption of research methods well established in qualitative research and in particular and also more precisely in the previous research the area can be also considered a strength for the credibility of this study. Additionally, to increase the chances of honesty in the responses, the interviewees were informed that they could refuse to participate anytime, their anonymity was ensured, and an iterative questioning was put in practice during the interviews. Debriefing sessions were also held during the research seminars and meetings with my supervisor to discuss different potential approaches for the research (Shenton, 2004).

3.5.2. Transferability

Sufficient contextual information about the interviewees' organizations was provided in order to situate the readers of the study in a position through which they could find similarities with other situations, and therefore to allow them to make the transfer of the findings. Moreover, the boundaries of the study have been provided and issues such as the number of organizations and the location where they are based, the number of participants, the data collection methods used, the number and length of data collection sessions and the time period over which the data collection was done (Shenton, 2004).

3.5.3. Dependability

Research design was stated in detail in the present chapter, allowing the reader to "assess the extent to which proper research practices have been followed" (Shenton, 2004, p. 71). The strategic and operational levels of design and implementation of the research design and the limitations observed during the data gathering have been detailed to ensure dependability.

3.5.4. Confirmability

Considering the objective of confirmability as the ability to trace back the results to the raw data of the research, despite the subjective views of the researcher (Pickard, 2007), the data collected was carefully gathered and treated and transparency on the audit trail (Shenton, 2004) was stated along Chapter 3. Methodology. Study databases were also created (Yin, 2003). The first database was created as a support for the research design and to maintain track of the evolution of the preparations for the data collection. Additionally, a second database was created with the aid of *Nvivo* software, which allowed gathering all the sources of evidence through a single access point with the aim of performing the analysis of the data. In order to maintain a chain of evidence (Yin, 2003) the reporting of the data includes representative quotations from the interviewees' comments to link the recorded data with the interpretations and analysis performed.

3.6. Ethical Considerations

Participants were reported through an informed consent form that the data collected was going to be treated with confidentiality and anonymity. It was guaranteed that the data collected was going to be used exclusively for academic purposes only, and in the context of the master thesis of DILL. Participants were also informed that they could leave the study anytime and therefore, their opinions would not be considered in the final analysis.

3.7. Limitations of the Research

Approaching the research using multiple case study as a strategy was considered. The underlying idea was that a more in-depth study of each of the organizations and their activities, including multiple perspectives from different actors and examination of further related documentation, could be beneficial to understand the phenomena. Due to the limitations on time, addressing a large variety of cases in-depth was not possible. Therefore, considering this point of view, the study may be lacking views from administrative or managerial units of the organization that may be more knowledgeable in some of the aspects under study.

Another aspect related with time constrains was the impossibility for some institutions to participate in the research during the data collection exercise period specified. Moreover, for the organizations, sensitive data was not possible to be disclosed due to the agreements that are subject of confidentiality. Nevertheless, sufficient data was gathered to understand the phenomena.

Some of the interviews had to be conducted in an asynchronous manner, through the exchange of emails. Despite most of the participants sent detailed responses and were open for additional interaction, the opportunity that synchronous interviews offer to exchange opinions or for the spontaneous elaboration by the interviewee, may be lost.

Moreover, in some cases the questions posed during the interviews seem to be subject of misinterpretation by the interviewees, and additional explanations were required. This aspect was taken in consideration and the questions were expressed in a different way after those episodes.

Chapter 4. Research results and discussion

4.1. Introduction

This study aims to investigate the risks at the organizational and regulatory levels to which memory institutions engaging in relationships with other parties in the scope of digital preservation are subject to. Secondly, the study explores how those organizations deal with the potential vulnerabilities and threats emerged with the distribution of their responsibilities in digital preservation, especially in the light of the benefits perceived and the inter-organizational trust relationships established. To accomplish those aims, the study has been guided by the research literature in the field of digital preservation, specially looking at aspects related with outsourcing and collaboration, risk management and establishment of trust, both in digital archives and inter-organizational, which are relatively new and emerging fields.

Therefore, this chapter presents the results of the data analysis conducted using the data gathered through the sixteen interviews carried out as described in the methodology chapter. The chapter starts with a profile of the participants involved in the data collection exercise, providing a description to allow the reader to understand the roles of the organizations and their relationships in the scope of distributed digital preservation. The roles of the interviewees within those participant organizations and information about the geographical area are also provided.The results of the analysis are presented in five major categories, namely: Perceived benefits for the organizations outsourcing or partnering for digital preservation. Major risks perceived. Controls: risk avoidance and mitigation strategies. Trust mechanisms used in the inter-organizational relationships . Collaborative trends in the scope of digital preservation.

The conclusions will be presented in Chapter 5.

4.2. Profile of participants

The interviewees were selected as already mentioned in 3.3.1 following a set of predefined criteria for the institutions of which they are part of. Due to the limitations set up to fulfil the confidentiality requirements, detailed descriptions about the institutions cannot be included.

Nevertheless, brief descriptions about the institutions involved will be provided and a summary of the characteristics of the institutions and the interviewees can be found in Table 4.

Table 4 Summary of the participants' profiles

profile	*type of institution*	*country*[26]	*id.*	*Role of interviewees*	*storage*	*functions*	*cloud*	*Grid*	*Collaborative activities*
Cooperative									
	University Library	CA	Int# 16	Digital preservation officer	x	x			x
	Consortium of Universities	ES	Int# 10	Director of Consortium / Head of IT Department	x	x			x
Institution outsourcing									
	University Library	UK	Int# 3	Library Systems Manager	x		x		
	Archive	UK	Int# 1	Digital Preservation Officer	x	x	x		
	Archive	UK	Int# 5	Digital Preservation Officer	x	x	x		
	State Library	DE	Int# 6	Digital Preservation Officer	x	x		x	
	Consortium of Libraries and Archives	UK	Int# 14	Head of Digital Preservation	x	x	x		x
	Consortium of Universities	US	Int# 9	Director of Curation Services	x	x	x	x	x
	State Archive	AT	Int# 8	Digital Preservation Officer		x			
Centralized service									
	State Library	DK	Int# 4	Head of Digital Resources	x	x			x
	e-Infrastructure	PL	Int# 7	Projects lead Digital & Preservation Services	x	x		x	
	e-Infrastructure	US	Int# 2	Program Manager & Director of Digital Preservation	x	x		x	x

[26] Country codes taken from ISO 3166 https://www.iso.org/obp/ui/

				Initiatives					
Service provider									
	Commercial Provider	PT	Int# 11	Innovation Director		x			
	Commercial Provider	UK/US	Int# 15	Operations Director	x	x	x		
	Open-source service	CA	Int# 12	University Librarian & President / CEO		x	x		x
Private cloud									
	University Library	UK	Int# 13	R&D Project Manager	x	x	x		x

Although a primary role of each institution was considered when selecting the institutions during the research design, each organization presents different roles that can overlap. For instance, if an institution is outsourcing some preservation functions and storage, it can also be engaged in a cooperative relationship with other peer institutions, as in the case of Int#14.

In relationship with the roles, the interviewees in memory institutions were the ones dealing with the digital preservation duties in the organizations. There is some diversity on the roles of the participants, although most of them, especially in the libraries and archives could be grouped as digital preservation officers. Some of them with unique dedication to digital preservation and others with responsibilities in relationship with IT or other aspects related with the general planning on the development of collections and additionally, high-level managerial positions. Despite the different positions, all of them expressed being involved in the decision-making processes related with the solutions for digital preservation.

In the other hand service providers' interviewees were the individuals in charge of the development of projects related to the services of digital preservation in their organizations, which implied that a high-level overview in relationship with their customers was provided and not from the technical point of view of the solutions.

Cooperative models

Cooperative models engaged by Int#16 and Int#10 are both based in a relationship established through consortia using a Private LOCKSS Network (PLN) as a strategy to ensure the distribution of copies of their digital assets through different geographical and organizational environments. In the case of Int#16, the consortia of universities organized their own PLN among the members based in the same country; whereas in the case of Int#10 as a consortium of universities they are participating in a wider international PLN. None of them expressed the need of having large local technical expertise or capacity, due to the type of service covered and the easiness on the set-up of the systems. Nevertheless, the need for expertise has been expressed at the policy and organizational levels in relationship with the network.

Memory institutions outsourcing

Memory institutions outsourcing their needs for digital preservation are the biggest group and therefore the one that shows major heterogeneity. They range from an academic library and a consortium of them (Int#3 and Int#14); archives (Int#8, Int#1 and Int#5); a state library (Int#6) and a consortium of libraries and archives (Int#14). All participants are outsourcing their storage either to cloud or grid services; only Int#8 did not outsource this feature. All participants, but for Int#3 are outsourcing functions of digital preservation.

The needs of outsourcing seem to be linked with the expertise and technological capabilities held locally, rather than being only dependent on the size of the institution. The fulfilled needs range from being the main infrastructure in place for the service in the case of institutions that expressed lacking the resources in-house, which is the case of Int#1, Int#5, Int#14 who, at the same time, were using cloud technologies both for outsourcing storage and the services.

Other institutions outsource a part of the digital preservation needs (Int#6, Int#3, Int#8), due to their capacity, as their expressed having enough expertise in the own premises to cover some the duties not externalized. Similarly, outsourcing using cloud as a control strategy to avoid potential risks (Int#9) seeks fulfilling part of the needs that cannot be done locally.

Centralized services

The category of centralized services includes institutions in the scope of cultural heritage or academia providing services of digital preservation to other institutions. This type of services are typically provided by large institutions with capacity in terms of expertise and technological development enough to build a business model and fulfil other institutions' needs besides the own ones. Two e-infrastructures (Int#7, Int#2) and one large state library (Int#4) are grouped in this category. Moreover, two of these organizations stated their collaborative relations in a larger network with other institutions to provide the services (Int#4, Int#2).

Similarly, in the case of Int#13's institution which is a large academic library the main role considered in the analysis was the development of a private cloud to support their own digital preservation activities locally and potentially offer services to other institutions. One of the objectives in this case was to gain insight on the differences between the private and public cloud uses, but also about their collaborative activities.

Service providers

The organizations in the group of service providers are commercial providers, two of them offering services related to open source software for digital preservation (Int#12, Int#11), and the third one providing both services and storage in the cloud (Int#15). Nevertheless, in the case of Int#12 insights have been also provided from the point of view of his responsibilities as university librarian responsible for digital preservation activities, in aspects such as collaboration.

4.3. Perceived benefits for the organizations outsourcing or partnering for digital preservation

Outsourcing to third-parties or collaboration among institutions have been discussed earlier in this study (see 2.2.2 and Outsourcing digital preservation) as options for distributing the burden of digital preservation. Thus, a first general question for all participants was intended to find out what are the perceived benefits of those options for the memory institutions. The replies obtained gave different perspectives, from the point of view of the service providers and memory institutions either using services provided by third parties or collaborating with peer institutions.

Commoditization of IT infrastructure and using cloud as a utility (Dhar, 2012) was perceived as an advantage by several interviewees, as it implies that the service can be used without the burden of sorting it out within the organization's premises.

> Int#1 "If you can use cloud storage, then all that infrastructure is already provided really elsewhere. So you are just into a service really."

Moreover, leasing IT resources was reported to bring in the vantage of flexibility (Bradshaw et al., 2010), allowing customers to modify the conditions on the infrastructure needs depending on the requirements.

> Int#1 "the flexibility that the cloud storage providers provide, because we have a good estimate (…) in terms of what's the content that it is going to be ingested and preserved."
>
> Int#5 "The flexibility and scalability of storage means we pay only for what we use, and we can choose between storage options according to our access needs."

Using the cloud or other services was described as a more cost-efficient solution than in-house and expected reductions of costs of the service (Dečman, 2007) have been commented by three different interviewees.

> Int#14 "They produce everything in the cloud so we might go with them because the prices are coming down considerably."
>
> Int#1 "We've already done like cost-analysis basically and it works out more, much more cost-efficient to actually do it using cloud storage providers than doing it all internally."

Nevertheless, Int#12 stated that public cloud might not be yet the most cost-effective solution depending on the size of the collection (Rosenthal & Vargas, 2013), although due to the cost reductions, it is getting closer.

> Int#12 "If you have a lot of data, like you are in hundreds of terabytes, then I think that commercial cloud storage can still be not quite as cost effective as local data centre storage. But I think it is increasingly a viable option."

Another source of cost reductions reported by interviewees was a consequence of leveraging economies of scale through the use of digital preservation services, as pointed out by Lambert et al. (2014). Memory institutions would have access to lower costs either using third-parties or if service providers might join partnerships with larger groups themselves, as reported by one of the participants.

> Int#2 "It is also going to help us to provide cheaper, more inexpensive services, just because it is a larger group and so we can share those costs among ourselves."
>
> Int#4 "They can obtain secure bit-preservation at a low cost."

Costs have been reported to be clearer outsourcing the service compared to in-house infrastructure (Dečman, 2007), simplifying the administrative burden to financial units in the organization. Moreover, in the case of open-source solutions Int#11 asserted that organizations using those options through commercial providers benefit from avoiding hidden costs in which they may incur if they take care of the system by itself.

> Int#3 “with outsource it is quite clear the costs. You know what you are paying, you know the service you are getting. When it is a local service, many of the costs are hidden.”
>
> Int#11 “if you take the system by yourself, there might be the risk of this hidden cost of developing and maintaining the system as it goes along. There are also commercial services that take this out of your hand without hidden costs.”

In the case of labour, some participants reported that outsourcing the service reduced the need of increasing the expertise in the tasks involved in digital preservation, as well as the need to staff the areas related to those activities. Moreover, outsourcing aspects related with technology was pointed out as a factor that allows memory institutions without expertise in digital preservation to meet their needs enabling them to keep the focus on the activities of the institution on which they are experts.

> Int#2 “They get their digital preservation needs met without having to hire anyone in their own organization (…). And they can continue instead of focus on what their expertise is and what they do care about.”
>
> Int#5 “Using a third party service means we do not have to rely on existing IT resources within the Council – these are limited and cannot be guaranteed for any length of time.”
>
> Int#1 “We require a certain amount of support. We rely quite heavily upon [institution’s] IT but we need that extra layer of support.”

Therefore, smaller institutions without infrastructure or expertise would benefit from the use of outsourced services. It was pointed out by Int#15 that contracting out services using cloud in particular would be an affordable way for those institutions to obtain the same services than bigger institutions. Although the downside of this aspect has been suggested by Skinner & Halbert (2009). On their opinion, the expertise on the functions related to digital preservation may end up only in a number of specialized corporations if the services are outsourced to commercial providers. Similarly, Walters & Skinner (2010) argue the potential loss of expertise in digital preservation by memory institutions outsourcing.

> Int#15 “larger institutions tend to have their own existing IT infrastructures and IT staff to run them, they have that kind of expertise, while the smaller customers maybe don't”

Nevertheless, some of the interviewees also commented that, without outsourcing or partnering in the technological aspect, it would not be possible to support the activity of digital preservation neither evolve in other aspects related to the service.

> Int#1 “So it just allowed us to evolve our digital preservation set up here in the [institution] really. It's been pretty key.”

Low barriers to entry and short time to set up compared to the use of in-house IT infrastructure, were also pointed out in general by the interviewees using cloud as benefits of outsourcing to cloud services. Besides that, lower expenses using cloud storage was also reported by one of the service providers (Int#15) as a critical factor to allow smaller not well-endowed institutions to be able to afford services of digital preservation similar to those used by bigger institutions.

> Int#3 "I think one of the benefits is speed in the sense that if we do it in-house we are going to set up everything ourselves, set up the hardware, infrastructure, the systems, and it can take 18 months, maybe, to achieve. Going into a company, either a cloud hosting solution, or other that can perhaps deliver some hardware in the institution or host externally as well. With this we can do it in about 6 months."
>
> Int#5 "Using *Preservica Cloud* meant we could procure and get started with digital preservation very quickly at an affordable cost."
>
> Int#15 "(…) services that are offered in the cloud allows our customers to get the benefits of a digital archiving solution without having hardware. And that allows smaller institutions (…) to get as much as the same service as we offer to our bigger customers. With a very easy start-up process."

The low level of investment needed was also reported by some interviewees that are members of partnerships between memory institutions (Int#16, Int#10). They stated that they had been able to satisfy their needs by leveraging their own infrastructure and expertise to meet the goals of the partnership without a needing a big individual effort by the institutions.

> Int#16 "The software runs on regular computer machines so you don't need big servers or expensive equipment. It is regular equipment you can use to run your LOCKSS boxes."
>
> Int#10 "The first effective low cost solution that could give some preservation to a collection of data that we have which is that of the thesis."

Another benefit of partnering for digital preservation was pointed out by Int#4, who argued that the projects done in partnership with other institutions are a more sustainable choice than those tackled by an individual organization. Lambert et al. (2014) state similar ideas about memory institutions sharing their preservation services increasing sustainability in the sense that it allows those institutions to benefit from scale and increase robustness.

> Int#4 "When it comes to sustainability I guess we believe that projects between several partners are more sustainable simply because they are harder to just shut down (compared to a project run by only one institution)."

Outsourcing and partnering have also been referred to as the outcome of a risk management exercise, which is a common way for decision making in the field of digital preservation as has been discussed in the literature review (see 2.2.7). Therefore those options were considered a control put in place to avoid or mitigate the risks of keeping digital preservation locally.

- Difficulties to economically sustain keeping the digital preservation set-up in-house.

> Int#1 “balancing the costs and the risks, I think the cost will simply be too prohibitive to do all this within [institution] so cloud storage has provided a really good option.”

- Use of third-parties aims to act as a control strategy storing data in redundant architectures, reducing the likeliness of some risks (see Redundancy of data).
- The guarantees offered by third-parties providing the services of digital preservation were also pointed out. For instance, Int#3 commented that guarantees that third-parties are able to provide (e.g. geographical redundancy, escrow agreements or taking legal liabilities), give them more confidence about outsourcing than developing the service locally.

> Int#3 “(…) we get a lot of confidence, in fact more confidence than looking into it locally. (…). I think in many ways, outsourcing is more beneficial, brings more guarantees for us than a local solution.”

- Common satisfactory solution for all the members in the partnership through the cloud storage and cloud installation (Int#14). The risk identified was the use of open-access software not fully released that could not be installed in some of the local systems.

> Int#14 “To be honest we could only see the cloud, because we had so many issues with local installed hardware and the inability to get local IT support (…).”

So far, the benefits perceived by those using cloud were stated on one hand, from the point of view of what the technology itself is capable to offer using it for digital preservation systems; on the other, the ideas stated were related to using third-parties or outsourcing. Nevertheless, other approaches are possible depending on the type of deployment of the cloud as technology. For instance the use of a private cloud by a single institution and the configuration of a community cloud for a partnership of universities were also discussed. The benefits perceived in those two cases were associated with the long-term sustainability of the projects of digital preservation. Taking advantage of some of the characteristics of the cloud, and even using similar technologies these large academic institution and network of universities described those options as more suitable for the activity of digital preservation in the long-term.

> Int#13 “Essentially, I guess the reason we go for a private cloud is, to be honest, we are looking at long term plans and a critical factor for us, (…) is we expect to be around an awful lot longer than other people, just based on history.”

> Int#12 “I think one of our primarily methods of cloud storage and preservation will be an emerging regional and national storage cloud service. (…) a group of universities that is creating their own kind of cloud storage similar to Amazon (…) very important to me for long term sustainability is defining that regional and national cloud storage service.”

Table 5 Summary of benefits of using third-party services and partnering for memory institutions

Third-party/Cloud services	Partnership
Cost-effectiveness	Cost-effectiveness
Mitigation strategy vs. (only) local	Mitigation strategy vs. (only) local
Lower the costs of infrastructure	Low cost of the shared solution
Lower the costs of labour	Leverage of own infrastructure
Cloud as utility	Leverage of own expertise
Expected decreases of costs	Sustainability over the long term
Economies of scale	
Easier identification of costs	
Lower barrier to entry, short time to set-up	

4.4. Major risks perceived

As a framework to organize the risks discussed by the interviewees, the first group of criteria of the TRAC checklist (RLG-NARA Task Force, 2007) referred to organizational infrastructure, is used. The aim is to match the risks identified with the criteria related to organizational attributes defined in the framework of trustworthy repositories. The risks presented are grouped in five sections: the Financial management and sustainability; the Organizational structure and staffing; the Governance and organizational viability; the Procedural accountability and policy framework; and the Contracts, licenses, and liabilities.

4.4.1. Organizational level

4.4.1.1. Financial management and sustainability

Outsourcing to cloud has been pointed out as having implications in the structure of the budgets used by the organization contracting the services; ongoing payment for the service is required instead of doing the investments as capital expenditure. As stated by Sinclair et al. (2011) most of the European institutions surveyed through their study were assigning capital-only budgets, not suitable for maintaining ongoing payments. Two of the interviewees commented on organizations' potential concerns about this issue.

> Int#5 "There is an on-going debate about how to fund cloud services from a budget traditionally used only for capital expenditure."

> Int#12 "I can imagine in some institutions where will be a very challenging to convince the organization to spend a certain amount of money every year for a long time (…). I think it is still something new for many organizations, so I'll say it can be a challenge to for many organizations to account for cloud storage in the budgetary context."

Economic sustainability and ensuring the levels of funding are concerns for any organization involved in long-term preservation activities (Blue Ribbon Task Force, 2010). One of the main difficulties arising is generally related with the long-term dependency on funding that the activity involves for the organization, as described by Downs & Chen (2010), Giaretta (2008) or Jordan et al. (2008), in opposition of being a self-sustained activity (Lavoie & Dempsey, 2004). How to keep the streaming of funding needed, and how to cope with the ongoing payment for the services handed to third-parties may be difficult over the long term. As stated by Aitken et al. (2012) in relationship with public clouds, the computing resource is never owned by the user, requiring therefore an ongoing payment.

> Int#2 "One of the big challenges for digital preservation since it is long term it should be for years and years and years is the challenge of funding that, paying for it for that long. It's often very difficult and it can be very expensive to do that. Often, just providing backups is very cheap and so people assume that preservation is very cheap and it's not. So that whole budget and economic process is very much a challenge."

> Int#14 "it is obviously something that you have in the back of your mind, sustainability, how we are going to keep going."

Some stream of funding could be gathered from research projects funded by the research councils, there is however an amount of research that is not being funded and the expenses of preserving its inputs or outputs shall be included among the operational costs of the library, in the case of Int#13. This idea is connects with what was stated by Lavoie & Dempsey (2004), in the sense that funding models in digital preservation have typically allocated resources in a temporary basis.

> Int#13 "there is a big amount of research that is being done unfunded, especially in the Arts and Humanities, so we need to also have a way of providing them with archival storage."

Moreover, the libraries or archives responsible for the digital preservation activities could lose the stream of funding coming from their parent institution or other instances like the governmental organizations, since, as argued by Sinclair et al. (2011) the funding models in digital preservation operate in such a way in which it is easier to fund individual projects than obtaining long-term commitments from funding bodies to support ongoing investment. Therefore the viability of the activity cannot be guaranteed under those circumstances.

> Int#12 "University could pull their support for ongoing funding for hardware and storage pretty much any time and then we will be facing that issue on, in the absence of our institutional support, how do we ensure the long term viability and sustainability of our collections."

> Int#5 "We currently only have funding for a one year pilot to explore digital preservation using *Preservica*. Another major risk to the project is lack of funding, although this is not related to the use of the third-party service."

> Int#7 “The preservation activities related to digital content are not explicitly funded by the government; therefore lack of financial resources is one of the risks in this context.”

Likewise, public funding arriving to the memory institutions seems to be critical for the establishment of collaborative activities among them. But once they are set up and functioning, the institutions taking part of them would be responsible for providing the funding needed to continue with the activity. Further discussion about funding collaborative activities will be stated in relation with the collaborative trends section, later on in this chapter.

> Int#14 “they have a funding program (…) really does assist with our collaborative activities (…) we can't rely on funding over time.”

> Int#6 “(…) it is a very close cooperation. In two projects, (…) funded by the German Research Foundation (DFG) it became possible to create the technical infrastructure and address the organizational matters.”

Institutions unable to estimate the growth of the collection in the medium to long term might incur in unexpected expenses. On the other hand, even using a private cloud as in the case of Int#13, operational funds in the organization might not be sufficient to cover the expenses if they rise sharply due to the need of the service. This assertion is also claimed by Aitken et al. (2012) in relation with community and private clouds, in the sense that the institution/s involved have to sustain by themselves the costs of maintaining the underlying infrastructure and the consecution of economies of scale is less likely to happen; or Dečman & Vintar (2013), who stated that long-term financial resources for personnel, hardware, software, management, and other activities need to be guaranteed.

For instance, Int#1 commented that one of the important issues raised in connection to the use of cloud storage was the potential growth of the collection. He mentioned that putting figures in the short term was possible “but in the medium to long term it is more uncertain so we'll just going to have to see how it goes.”

> Int#3 “But certainly you have to increase the revenues to actually increase the data.”

> Int#9 “cost issues or cost concerns have arisen when we discussed putting more than data, qualitative data files, quantitative data files in the cloud so we have 60 to 80 to hundred+ TB of video data or video files and those are not in the cloud simply because of the size of the cost, I think associated with replicating.”

Increasing the volume of data stored in outsourced cloud involves the need of increasing the expenses and therefore, a higher allocation of financial resources to the activity, as reported by Int#3 and Int#1. Storing large research data files may imply very large costs and therefore the need of increasing the financial support for the activity. Yet, those large data files will also increase the costs of performing actions on the data, such as stated by Int#9, when talking about the cost of replicating and supported by Convery (2010a) who stated that the total cost involved in moving information and processes into the cloud should be considered.

Therefore, pricing models in cloud services can be a source of additional expenses to the institutions outsourcing to service providers. The cloud up-front payment may be cheaper than other options but additional charges may occur when there is a need to perform actions to the digital objects, as was described by Convery (2010a), but also by Aitken et al. (2012), in relation with data that needs to be accessed frequently and higher bandwidth needs.

On the other hand, service providers not using cloud have reported different models to charge for their services, such as Int#2 or Int#11 with a single payment including a complete catalogue of services, which in some cases could initially imply a higher price, compared to cloud services that may imply extra charges to perform additional actions after the up-front payment.

> Int#2 "You pay for get it out or you pay for network charges or you pay (…) if you need to do any kind of computation, a checksum or verification you pay for everything. Whereas for us, our single cost is much higher but there are no other charges."

> Int#12 "If I have some kind of disaster and I need to recover my data from Amazon Glacier then I'll pay for that, but at least is fairly secure and safe in that container."

> Int#3 "(…) Amazon and Google you can store vast amounts of data but they charge you when downloading data. (...). Which makes them not particularly ideal for distributing data and make it available for everyone."

In the particular case of the partnerships using LOCKSS, independently of the expenses that they might incur related to the collaboration activities, an ongoing payment is required to use the system. Int#16 comments that it might be challenging for certain institutions to sustain the cost of the fee in the long run.

> Int#16 "LOCKSS software is from Stanford University, so they charge a yearly membership fee. So institutions like [ours], we are paying for that every year and then we can be part of any of the PLN initiatives but other smaller institutions they might find it difficult in the future to pay every year the LOCKSS fee."

Lock-in practices have been described by Aitken et al. (2012), NAA (2014) and ENISA (2009). Costs related to vendor lock-in practices may arise when a memory institution wants to leave a service provided by cloud vendors. In the case of an institution wanting to switch to another service provider, the commercial entity may create some difficulties. Examples of such situations were reported by Int#3 and Int#9:

- Vendors do not communicate between each other, unless the institution pays for it as a service.

  Int#3 "If you make a deal with the second outsourced company these two companies won't talk to each other or you have to pay for them to talk to each other because they are business so they might insist on giving the data back to you before you can give it to someone else."

- Current service provider will only return the data to the institution and not transfer it to a second service provider. If the institution does not have infrastructure in place in its own premises, temporary hardware will have to be purchased.

  Int#3 "makes it very difficult, you might have to buy a lot of hardware temporarily just to store the data and bring it back again."

- Cost of claiming the content back can be very high so in case of having more copies if the organization wants to terminate the service, the copy in the cloud will be left in the cloud service or be deleted.

  Int#9 "If we want to take the content back, (…) copies that we put on Amazon I don't believe we'll worry about bringing the content back from them, I think that can be expensive so what we would do is essentially move on, just leave it or have them deleted, because that is not the only copy, it is just one copy."

Table 6 Summary of financial management and sustainability risks

Commercial/Cloud service	Centralized service	Partnership
Inadequacy of ongoing payment for budget structure	-	-
Insecurity on long-term funding to pay for the services	Insecurity on long-term funding to pay for the services	Ongoing payment to partnership
Pricing models and additional charges due to increase of data stored, the size of the files or the actions performed	Greater up-front payments	In-house expenses due to the partnership
Inability to estimate collection growth related with pricing models	-	-
Additional costs due to lock-in practices in case of data portability needs	-	-

4.4.1.2. Organizational structure and staffing

Lack of expertise can be seen as part of a structural problem in the organization, to a large extent. Larger organizations with IT teams seem to be better positioned than others to assume digital preservation activities independently. As reported by Int#13 their projects of digital preservation can be built using their own on-site expertise. The University Library supported by the IT services, not only from the library but from the University, a super-computing centre, plus a number of research centres and departments that might also be involved in the projects provide them with the capacity to run projects such as the private cloud they are using.

> Int#13 "At the moment we are looking to basically use our own on-site expertise. But this is probably because of the nature of the organization we are and we do have the capacity and the expertise to do that. (…) To some extent I feel that it is our job to be experts as National Libraries to be experts in long-term preservation of data."

Organizations participating as members in partnerships require staff involved in the tasks of digital preservation with skills and expertise. Sometimes, as pointed out by Int#16 advance knowledge is required.

> Int#16 "So you need LOCKSS plug-ins which are not available, and writing those plug-ins is also very difficult."

A similar situation was reported when some of the interviewees explained their concerns about using open-source solutions. They considered that their institution could not support those alternatives with the staff and level of expertise currently available in the institution.

> Int#5 "Although favourable in terms of up-front costs, the open source option would be dependent on in-house maintenance and development of the software. We had neither the technical capacity within our own service, or any guarantee that the Council IT department's resources would allow a long-term commitment to such a project."

> Int#1 "open source was looked up, so we are talking about repository environments just like Fedora and e-Prints and all that kind of stuff, but within [institution] itself we couldn't support that really with the developers requirements."

On the other hand, capacity and resources are also needed for the organizations to be able to contribute to the partnership. As Halbert (2009) argues, challenges may arise because of the lack of expertise in the institutions on participating in networks. Some participants mentioned the limitations of some of the organizations to contribute to the partnerships, for instance with resources such as dedication of time, financial contributions or labour.

> Int#14 "completely diverse capacities and some people were able to do something but they just didn't have the resources or the time to be able to contribute as much as they would have liked."

> Int#13 "talking at the moment with the economy in the state as it is, there is a number of institutions that really can't afford contribute very much even thought they would have a particular need."

> Int#10 "by transferring the costs I don't mean only the fee, which is our case, but often costs that may be of labour."

Impact on staff was only reported by some participants. In the case of outsourcing services to cloud, Int#1 and Int#10 reported that there were no significant changes such as loss increase of staff members. Int#12 argued that the easiness of using the systems did not require new skills or expert knowledge. On the other hand, participants Int#3 and Int#9 also reported that some new capabilities and knowledge were required to have a better understanding of the new technologies and that adaptation to the new workflows or procedures was also needed, which is consistent with what was expressed by Aitken et al. (2012) related with potential role changes and costs associated with staff in institutions moving their infrastructure to cloud.

The involvement of more staff in some duties or the creation of new roles because of the new workflows was reported by Int#3 and Int#10. Nevertheless, this new need seems to be connected to the fact that the activities related to digital preservation services under discussion did not exist prior to those changes in the organization, rather than depending on the setup after outsourcing or joining the partnership, respectively. Additionally, Int#3 also reported that staff in the different units involved in those workflows is used to work in changing environments.

> Int#3 "Some technical skills we have, but we need to understand how the technology works (…) but beyond the technical we have to understand the processes to some extent."

Moreover, Int#2 reported misunderstandings on the processes involved in digital preservation activities by organizations using their services.

> Int#2 "they often don't think about a lot of the same things that you have to do in preservation and so when we come in and we say well you have to do this and you have to do that and you have to do this other thing and then they say 'well, are you sure? we don't think that's important'"

The change of the type and length of relationships that are more common in outsourced projects over the short term can introduce challenges to maintain levels of relationship between customer and service provider. Maintaining the appropriate communication and contact levels in long-term was also reported as a potential threat.

> Int#2 "that is one of the big challenges I think we are all going to face over time, maintaining that relationship, that appropriate kind of contact with the right people. It is very different than a short one or two year relationship."

Mismatches between the preservation and access services, whether those are in the organization or provided by a third-party, were reported by some interviewees. An issue mentioned during the interviews with Int#14 and Int#2 was that the use of different processes and technologies in each of the services and a lack of coordination in the strategies of the different actors involved.

> Int#14 "link anything in the cloud with your cataloguing system (…) that is another risk that we are identifying, how to make sure we can integrate within our existing catalogues."

> Int#2 "the other risk that people face is being able to connect their preservation services with their access services, because they are usually two very different processes and so people who often care about preservation they are not the people who care about access and so they have different goals and strategies they often have different technologies."

In relation to access, a major limitation stated was the level of development of the infrastructures in the geographical context where the organization of Int#13 is located. The bandwidth infrastructures are not advanced enough to allow the use of external service providers to store large amounts of data when they need to be accessible.

> Int#13 "within the UK and Europe that level of infrastructure doesn't exist yet. So just the volumes mean that in a lot of cases, certainly for what I'd call online accessible data, we need to keep that in-house."

Table 7 Summary of risks for the organizational structure and staffing

Commercial/Cloud service	Centralized service	Partnership
Need of new capabilities and knowledge	-	Lack of skilled staff or expertise
Adaptation of staff to new workflows and procedures	Misunderstanding of the processes involved	-
-	Difficulties on maintaining levels of relation on the long term	-
Mismatches between preservation and access services	Mismatches between preservation and access services	-
-	-	Lack of capacity and resources to contribute to the partnership

4.4.1.3. Governance and organizational viability

Loss of reputation as a threat for an organization outsourcing to cloud was mentioned by Int#1, although considering that would only happen in a worst case scenario. This aspect has been addressed also in the literature such as NAA (2014) that raises concern on the loss of service reputation or ENISA (2009) that mentions potential loss of business reputation due to co-tenant activities. Other factors commented by Convery (2010a) are problems with the availability and reliability of services that might cause loss of service, income and reputation.

> Int#1 "Obviously you keep the reputational element in the back of your mind as well, but that's a worst case scenario and that's why we do so much work to actually reducing those risks the best as possible."

Vendor lock-in as a menace to the memory institutions was already commented in relation with the potential costs that may appear if the organization decides to interrupt business relation with the service provider. But the difficulties that leaving the service could bring in may be also understood as loss of governance (ENISA, 2009). For instance, it was already commented in the case of Int#9 that would rather prefer to leave the copy of the data stored in the cloud provider for them to delete it or to decide over it because of the high cost that discontinuing the relationship could introduce.

On the other hand, the fact of being the assets in the cloud data seems to increase the risks related to lock-in, that commonly increases at the same rate as the data does.

> Int#3 “for the experience with smaller systems we know that it is problematic to leave a company. There is a process to discontinue business with them but it is much harder when they have data and have to move it back so that's going to be complicated.”

Changes on the strategy or conditions of the service provider were also identified as risks by the interviewees. Changing the strategy could result on the loss of support to the services in use by the institution or result into different arrangements in relation with the service. The particular case of the cloud was also mentioned by Int#12 that expressed concern on the potential unilateral changes of conditions by the commercial cloud providers, that could make the institution not interested in or able to or use it anymore.

> Int#11“Not just ceasing to exist but changing the objectives of the company and drop the support to these solutions.”
>
> Int#7 “Obviously changes in the strategy can affect the whole setup.”
>
> Int#12 “The danger with the commercial cloud services is that they can change the rules anytime.” “The risk is that sometime in the near future, Amazon discontinues their Glacier cloud service or they raise the price, for me from 10 to 30 thousand dollars or change the service in such a way that is no longer feasible for me to use it.”

Service provider going out of business or bankrupt, in the case of outsourced services to commercial companies, was also mentioned by several participants as a potential threat and is commonly identified as such in the research literature (Gellman, 2009; NAA, 2014; Convery, 2010a) Some of the interviewees emphasized this aspect especially in relationship with the long-term commitment that digital preservation involves, and the fact that companies cannot guarantee its existence indefinitely. The size of the company could be a factor influencing this risk, considering that a smaller company could be more vulnerable to cease its activity, as pointed out by Int#11.

> Int#1 “at the end of the day these are all commercial organizations and they are all subject to the winds of all commercial organizations, they could go out of business”
>
> Int#9 “and in the past there have been providers that shut down, and that happens all the time with web companies.”
>
> Int#3 “Certainly it is more predictable but of course you need to have the trust that the company is going to be around in 10 years.”

In the case of collaborative arrangements, there could also be dependencies between the institutions involved. Int#16 reported that the system that they collectively use does not work without the minimum number of participants. Therefore, if an institution leaves the partnership they could incur in additional expenses or in the need of seeking for new partners.

> Int#16 "if the number of nodes are below six, then the LOCKSS software cannot work so it means that we must need at least six or seven nodes in order to run this preservation system. So if some of the member institutions quit then we are in trouble."

Sustainability of the services was especially discussed in the cases where the organizations were using or providing open-source software solutions. Dependency on the original creators of the solution for the project to evolve and lack of plans ensuring business continuity may be a risk for the organizations using those services.

> Int#16 "it is an open software, but there are no guarantees. It is still run by one or two persons there."

Diversity on the mission, goals and strategy can be a source of conflict among the institutions within a partnership. Some interviewees mentioned different levels of alignment among the partners.

> Int#14 "the partners are very diverse in terms of their capacity and what they want to achieve." "There is a bit of tension possibly between what the Library is doing and what the Archive is doing but we try to align and work together on that."

> Int#4 "You need to have common objectives to be able to steer common projects in the same direction."

Interviewees in partnerships also commented about the different involvement of the members in the decision making processes of the partnership, depending on which types of common structures have been established. The different approaches will be commented later on in the section dedicated to mitigation strategies, but in a broad sense, the examples ranged from common agreements in all decisions in the case of Int#14 to only high-level decisions in common and delegation of operational decisions in the case of Int#9. On this regard, the capacity of the network to be able of strategically planning and run the network effectively is essential to avoid potential challenges (Halbert, 2009).

> Int#9 "we have this council that meets three times a year and provides insight and votes on certain things and I think for the small details we don't run them by council all the time."

Leadership role in the partnerships could be also formal or informal, depending on the cases. Int#14 reported on the informal leadership role taken by the National Library despite their decentralized formal structure of the partnership. Related to this aspect, Halbert (2009) stated that having the largest institution of the network as *de facto* leader, may be a source of malfunctions .

> Int#14 “the library does lead because the project manager is based here and I'm based here and we have access to infrastructure and skills to a greater extent.”

Table 8 Summary of the risks on governance and organizational viability

Commercial/Cloud service	Centralized service	Partnership
Loss of reputation	-	-
Loss of governance due to lock-in	-	Informal leadership role taken by the larger organization
Changes on strategy or conditions	Changes on strategy or conditions	Diversity of mission, goals and strategy
Service provider out of business	Service provider out of business	Minimum number of members in the partnership, dependencies

4.4.2. Regulatory level

4.4.2.1. Procedural accountability and policy framework

The selection of the collections to be preserved in systems that are part of partnerships or centralized services offered by memory institutions for a network, has been mentioned to be a challenging policy issue by the interviewees in those kind of situations. Decision making process around the collections has to be done collectively in the case of partnerships. The relevance of collection development for long-term retention is one of the requirements to achieve economic sustainability stated by the Blue Ribbon Task Force (2008). Therefore, the scope of these decisions on the policies may be crucial for the long-term viability of the partnerships.

Some relevant aspects that have been pointed out by the interviewees partnering for digital preservation were the differences on the size and typology of the digital objects in the collections (e.g. formats); the different capacities on the infrastructure and expertise available in the organizations; the different goals established by the organizations regarding their collections; the different procedures and actions that need to take place in relation with the collections (e.g. content ingest rates); different needs in terms of data protection.

Different interviewees made an approach to what may be behind these regulatory concerns. Thus, Int#16 commented that problems might arise if one of the members does not have the technological capacity or willingness to add the content that other members want to. Therefore, deciding around collections is one of the difficult questions with which Int#16's institution is particularly still struggling with.

Int#3 commented that one of the issues that could be a problem partnering with regional peer institutions is the different nature of research and, consequently, the different types of data that need to be preserved. Establishing common policies and procedures for the variety of requirements in terms of formats, size or confidentiality, might be challenging as well. Similarly Int#14 also remarked that the type of collections and the requirements within the partnership on which her institution is part of are very diverse. The institutions range from academic libraries to small archives, and therefore from research data to all kinds of textual documents. In this sense, Int#4 as a centralized service provider for other institutions stated that due to the different requirements of collections and the differences with their own internal collections, they cannot accept to provide services to other institutions in those occasions.

> Int#10 "I don't think it would be a matter of cost, but of finding a collection that would make sense (…) [a whole repository] is too big and there are all kinds of things. A much higher activity than the thesis' one"
>
> Int#16 "there are 10 members in our network and if one member (…) comes up with a large amount of terabytes of data and want to save it in the PLN, then the problem is what if other institution do not want to add that."
>
> Int#3 "I think where the difficulties will be is the different styles and different types of research. (…) So they are using different formats storing their research data, their research data is very large in size, and also is very confidential."

Entrusting control over the digital assets to third-parties is the usual practice using their services. Nevertheless, getting assurance that the properties of digital objects are properly maintained by the service provider is a matter of major importance in terms of meeting business objectives for the memory institutions, such as committing to a successful digital preservation (Vermaaten et al., 2012). Relying in contracts and SLAs might not be sufficient to get guarantees (Kyriazis, 2013), despite that one is the current practice reported by some of the interviewees.

Monitoring performance is one of the options that memory institutions outsourcing may use to control those aspects. Nevertheless, lack of access to adequate reporting might diminish the possibilities of monitoring the state of the digital assets handed to those third-parties or the level of compliance with the agreements, and therefore facing the risk of losing control at different levels.

> Int#1“[we] rely upon SLAs and contracts (...) so we don't actively check. Once is in the cloud we rely on the commercial organizations to do it but that is an issue that we have to look down in terms of how do we'll do.”
>
> Int#15 “we don't measure it independently but they say they have never lost of our files.”
>
> Int#9 “with the cloud provider they can tell you limited information, so if you send them information they will say, ‘yeah, we got it’, and ‘yeah it is here’, but even when they send you those reports you still have to trust that they run their reports correctly.”

Effects caused to the institutional policies related to digital preservation were also a common concern. In this context, policies and transparency on the actions are particularly relevant to prove trustworthiness in digital preservation, as stated in the TRAC checklist (RLG-NARA Task Force, 2007).

The interviewees provided some ideas on whether using third parties or collaboration have been a source of changes to their institutional policies related to digital preservation. In particular, seven of the interviewees commented the situation of their memory institutions and three of the service providers delivered some perceptions about the topic. In general, interviewees stated different stages of development of policies for digital preservation. In the cases of Int#1, Int#13, Int#9 and Int#6, policies were complete and already in place. Other cases reported that their policies were still under development, such as Int#5 and Int#3; and Int#10 stated that as a consortium they were planning to develop some common policies.

Int#1 asserted that when they set up the policies and strategy they did not have yet decided whether they were going to outsource or chose other options. Therefore, the specifics were left open in those documents, and what they detailed were the best practices tied to the organizational goals. Int#13 underlined that their digital preservation policy basically states that “the manner on the preservation depends on the material in question” (Int#13), related to their institutional repository but also to the different external initiatives or repositories with which they collaborate or where the data from the institution is located. Int#9 declared that their policy suffered changes updating aspects related with outsourcing to cloud such as the number of copies and where the copies are kept and tacitly the level of acceptability of using a cloud provider. Int#6 institution’s policy for digital preservation details different aspects of their collaboration with a third-party e-Infrastructure.

The service providers Int#2 and Int#11 expounded that many institutions to which they were providing their services did not have a preservation policy in place before starting their collaboration, as they were not active in digital preservation. Upon joining the service, the requirements and boundaries became clearer, on their opinion. And therefore, policies were easier to be built.

> Int#11 “normally what happens [is] that the policies are created, refined and updated once this new system is put on.”

The interviewees mentioned about different levels of involvement or acknowledgement by their stakeholders in the decision making process to configure the setup for the digital preservation activities. Int#10 stated that, as a consortium, they worked internally in the decision and afterwards presented it to the universities part of the consortium setting a clear expectative about the project and getting feedback from them. Moreover, the decision of joining the cooperative was also publicized to the rest of the community members.

Int#1 judged that their users would not usually be aware of the technological architecture, in this case using cloud. Other stakeholders in their institution, which is a public service organization, might have shown some concern about using cloud, but the interviewee argued that the process has been done with transparency and according to the rules established in the organization.

> Int#1 “within the [institution] there will be always certain reservations about the use of the cloud (…) it's been a long going process to lay those concerns or issues that people may have internally”

Similarly, Int#9 noted that despite they publicized the procedures and information about the infrastructure in use, such as storing some copies in the cloud, the stakeholders might either not be aware or just fully trust them in the decisions made.

> Int#9 “the majority are unaware of these issues or it just doesn't bother them, so they don't think about long-term preservation. It just doesn't concern them and they are fully happy to hand over us and just implicitly trust us.”

Due to the changes on the requirements of research data management, Int#13 and Int#3 remarked that the researchers were more aware of the needs and services for digital preservation. Int#13 stated that there was a major involvement of the researchers on setting up the conditions for preservation, including assuming costs and acknowledging the benefits. Int#3 also commented that the academics were involved in the process for procurement of the service.

> Int#13 “So to that extent we've been doing a lot of work with our own research services department so the people cost for digital preservation, are aware of the need, of what they need to do, but also are aware of the benefits.”

Table 9 Summary of risks on procedural accountability and policy framework

Commercial/Cloud service	Centralized service	Partnership
Not meeting business objectives (e.g. digital objects requirements)	Collections policy and procedures diversity	Collections policy and procedures diversity
Lack of reporting / transparency	-	-

4.4.2.2. Contracts, licenses, and liabilities

Uncertainties on the cloud service providers' compliance with the Service Level Agreements (SLA) were also described by the interviewees in several occasions. As was already referred in section 2.2.6.4, SLA describes how the services are deployed and serve as the framework for the expectations and responsibilities related to them. As Kyriazis (2013) states, to ensure compliance there may be a certain need of verification of the agreed terms.

> Int#1 "we have an SLA with them as well, and they could potentially break it for whatever reason (…) we have to minimize the risk basically of corrupting or losing any data."

Non-compliance with SLAs implies the loss of service levels or availability, which were reported as sources of risk by the interviewees. There are different levels of severity considered, depending on the purpose of the service. For instance, in those institutions in which providing access is also the goal of the service, outages and network interruptions would be highly problematic. A different scenario would occur in the case that the assets stored were not copies for immediate access, but intended only for preservation. In that case, the urgency on the access might be lower, and therefore the level of risk would be consequently lower as well.

> Int#14 "with this cloud services quite often you can get a break on the service, they come down and they are vulnerable to attack so there is a concern that there might be not full access 24/7 that we are used to."

> Int#2 "in the one hand since we're a preservation service and not an access service often the timeliness is less important than the long-term guarantee, so sure we will have the occasional network interruption or power outages or something like that, normal things. But it is usually not the case that we have customers say 'I need to get my data right now' Tomorrow is ok, next week is ok. That's generally the kind of discussion that we have. And if one of our locations happens to be down we could just get it from a different one."

Ensuring data protection with cloud providers was expressed as a concern by some of the interviewees. In particular, Int#5 stated that the cloud provider in use cannot guarantee the levels of security that the content to be preserved requires. Int#3 argued that the protection cannot be taken for granted and that there is a need to understand how the security is actually implemented.

Int#1 also reported on the existence of sensitive data in their organization and the awareness of a potential risk on using cloud.

> Int#5 “A big issue for us is the security level that *Preservica* is able to guarantee. We need to store significant quantities of confidential or sensitive personal information in the digital repository. *Preservica Cloud* uses *Amazon* storage and is so far only accredited to store information with a Business Impact Level (BIL) 2 – some of our information requires BIL 3.”
>
> Int#3 “you have to actually check with the suppliers and say if we give you this data in confidence will you keep it private? and the companies as it is their business, they'll say yes but then we have to ask the next question: how can you keep it private?”
>
> Int#14 “we wouldn't at the moment advice anybody to put anything in the cloud that they are very scared might get lost or redistributed.”

The risks related to data protection are indeed one of the major concerns related to the use of cloud, and incompatibilities with the storage of sensitive data due to the restrictions on the data protection legislation have been reported by several authors such as ENISA (2009), Aitken et al. (2012), Dečman & Vintar (2013), Gellman (2009) or Convery (2010a).

Stakeholder concerns were reported in the case of disclosure of research data to a third-party commercial company, when the data has restrictions and its protection needs to be ensured. Those concerns seem to be well-funded under the point of view of Gellman (2009), who describes implications of disclosure to cloud vendors.

> Int#3 “I think where it comes more complicated is where it is mandated that you can't release to the public in two years or three years o however long, and then you still put it in a third-party company. I think this is an area where the academics are more concerned about.”
>
> Int#11 “many archives, national libraries have many legal restrictions for allowing them to put data outside of their own institution.”

Risk of unintentional disclosure of sensitive data due to attacks or malicious acts in cloud, including those related to activities of governments’ surveillance were also reported by Int#9. Moreover, disclosure can also be enforced due to regulations of certain countries such as the US (ENISA, 2009; Gellman, 2009).

> Int#9 “it might be a joke or even true that our National Security Agency is looking at the data.”

In the case of partnerships the potential risk of non-compliance with the legal requirements, such as data protection or copyright was also considered a potential challenge. Responsibility on legal issues is put on the staff of the institutions involved in the partnerships. Moreover, the individual institutions were also pointed out as responsible for the establishment of the right policies and the compliance with them.

> Int#4 “implies a number of challenges when it comes to both legal implications. Staff on both sides need to comply with regulatory and legal frameworks like personal data protection.”

> Int#16 “To make sure that their institutional policies are there and to make sure that they do not put any sensitive information or any copyrighted information in the PLN.”

Copyright issues related to the materials subject of digital preservation were also discussed with interviewees. In a general sense, processes related with digital preservation might introduce breaches of the regulations, in particular when they involve making copies or moving files from their original storage medium, for instance; nevertheless, regulations vary from country to country (D. Anderson, 2013).

In the case of distributed digital preservation, the interviewees commented about the responsibility and potential liabilities. The lack of agreements with data providers (Besek et al., 2008) that need to be made prior storing the digital assets in a third party service has been pointed out as a threat.

Service providers may limit the responsibility of liabilities through the contracts as stated by ENISA (2009) or Aitken et al. (2012). Service providers protect themselves through clauses in the contracts ensuring that the owners of the materials are giving them materials that they owned and that were not illegally obtained (Int#2). Moreover, Int#2 commented that being a service provider part of a University the limitations on financial resources were the reason to limit the liabilities.

> Int#2 “in the States is actually it's difficult to sue a university because we don't have money (…) so one of the things that we actually say in our contracts is that our legal responsibility is, if we do happen to take some things, some copyright material, we'll give it back, and will get rid of it (…) there is that kind extra layer of protection.”

Some interviewees also argued that the content stored in cloud or third parties is usually open content or out of copyright materials, whose aim is to be accessible as broadly as possible. And therefore, they didn’t express much concern about storing those materials within the third-party premises. Despite that, some measures for protection such as licensing or agreements will be commented when describing the mitigation strategies.

> Int#3 “So I think that once you put your data out, you have to accept that anyone can pick it anyone can do things with it anyone can store it. Big issues are that they do store in the same way as from origin.”

> Int#14 “the data that we hold is public data anyway. So you want to be exposed, really, that's what you do. You put it in the catalogue and expose it as much as possible. We are talking about public data, that is own by the public.”

A similar point was raised by Int#16 and Int#10 regarding their particular case within the partnerships. Open access content is the one that they are storing at the moment. But Int#16 also highlighted that the data they put in the PLN should not be restricted in any way.

> Int#16 “as we are only putting OJS content, and it is open, so nothing is copyrighted in OJS journals. We haven't dealt with any such issues.”
>
> Int#10 “those are open access repositories, what might happen?.”

Another regulation mentioned by a memory institution was the Freedom of Information Act. When public authorities are subject to this legislation, obtaining assurances from the third parties, cloud in the particular case of Int#1, on the compliance with its requirements is essential.

> Int#1 “In the UK we have the Freedom of Information Act, and we have to ensure that the cloud services that we use comply with it as well and that is probably the main requirement that we have.”

Service providers subcontracting third-parties can be a challenge for the final customers (NAA, 2014; Convery, 2010a); But also for the service providers themselves; such as the case of Int#15, who did not know whether the cloud provider they were subcontracting was using or not services from other companies.

> Int#15 “I don't know if Amazon uses third party services.”

Contractual relationships with partners can also entail difficulties. As reported by another service provider, Int#2, partnering with another organization is a challenge in the sense that, despite the roles and responsibilities might be clear, the day-to-day of the relationship between two different organizations are not necessarily so straightforward. Each of the organizations might have different levels of commitment or agendas, difficult to combine.

> Int#2 “It certainly makes things more complicated. Because now we have an additional contractual relationship that we have to maintain behind the scenes, and that is certainly important and a challenge.”

Informal agreements may be put in place in the case of close relationships with the third parties providing the service. Int#6 expressed that this lack of formal agreements could mean certain vulnerability for their institution.

> Int#6 “As not every single detail is laid down in contractual relations, this could be regarded as a risk.”

The jurisdiction on where the data is located was also mentioned as a factor to be taken into consideration by some of the interviewees, some of them using cloud and others partnering with others institutions abroad (Int#16, Int#10).

> Int#1 "cloud is geographically diverse so what we didn't want is be held in whatever fashion in a different jurisdiction, the United States for instance or anything like that; it has to be within the EU."
>
> Int#16 "Or that have some restrictions that cannot go beyond the borders so if there are any restrictions in any content then we won't send it over the US."

Similar arguments can be found in the research literature, and ENISA (2009) remarks the risks that assets in the cloud may have due to the storage in multiple locations in case there is lack of transparency and no information on the jurisdictions is provided.

Moreover, due to its diversity, cloud may imply certain loss of control in aspects like the location, and mitigation measures need to be in place (Int#1). Int#9 declared that their cloud providers might have "storage locations around the country, you don't know where exactly the data are in a given moment."

Table 10 Summary of risks on contracts, licenses and liabilities

Commercial/Cloud service	Centralized service	Partnership
Non-compliance with SLA and loss of service levels or availability	Non-compliance with SLA and loss of service levels or availability	Partnership members not fulfilling their duties
Non-compliance with data protection requirements	Non-compliance with data protection requirements	Non-compliance with data protection requirements
Disclosure of sensitive data / Attacks	-	-
Non-compliance with copyright laws	Non-compliance with copyright laws	Non-compliance with copyright laws
Lack of agreements with data providers	Lack of agreements with data providers	Lack of agreements with data providers
Service provider liabilities limitations	Service provider liabilities limitations	Partnership liabilities limitations
Service provider subcontracting services	Service provider subcontracting services	-
-	Informal agreements	Informal agreements
Jurisdiction – loss of control	-	Jurisdiction in international partnerships

4.5. Controls: risk avoidance and mitigation strategies

Affinity and alignment of mission and goals

Having a similar mission was considered relevant by different interviewees. It was especially valued as condition for trusting a third-party in the cases of centralized services such as those originated in the sphere of cultural heritage or academia by different interviewees.

> Int#9 “Duracloud is committed to the long term cultural memory. So you are dealing with the company but it is a company whose mission is explicitly aligned with our mission.”

From the point of view of memory institutions using those services, the fact that the institutions share similar principles was considered relevant in terms of having a better understanding of their needs; and also less likeliness of radical changes in the service provided, at least to the point of non-alignment with the institution’s needs.

> Int#12 “an institutional preservation cloud service like one run by your colleague universities it can change over time as well, but is less likely to change in such a way that is no longer feasible for you to use.”

Memory or educational institutions providing services to other fellow institutions pointed out that having a common interest is a critical factor, rather than exclusively making profit out of the situation.

> Int#13 “So it is never a purely financial transaction.”
>
> Int#2 “they often find that they're much more comfortable working with us because we are an educational institution and not a commercial company.”

Another interviewee (Int#4) whose institution is taking part of collaborative partnerships stated the need of having objectives in common to be able to run common projects. But also the pertinent organizational structures should be put in place to make the commitment operational. And Int#14, also involved in a partnership commented that they were working on the alignment of the activities of the different members to avoid overlaps and reduce potential tensions between them.

> Int#4 “You need to have common objectives to be able to steer common projects in the same direction.”

Establishment of partnership structures and policies

In relation to partnerships, some interviewees argued about the need of formal organizational structures to be able to work in the same direction, in a coordinated way. Int#4 remarked that for each project they set up a steering committee with members of the different organizations involved in the project.

> Int#4 “We normally do this by setting up a steering committee with management members from both organisations (in some projects more than two partners).”

Other interviewees pointed out that, as the common project is also more stable, they have more fixed structures, being most of them decentralized. This is the case of Int#16, whose institution is involved in a consortium that has a specific working group for digital preservation; Int#10, a consortium of universities themselves taking part of a cooperative for digital preservation; Int#14 using a consortium model and with an additional group acting as a community of practice around the project; or Int#9 whose institution is also built as a consortium with a council where the members are represented.

Most of them reported having regular meetings where aspects such as common policies or procedures are collectively discussed and decided.

> Int#9 “We have meeting every month, almost and main discussion have been around collection development.”

When talking about leadership in the partnership, different approaches have raised among the interviewees. For instance, in the case of Int#9, his institution acts on behalf of the council and takes operational decisions on a regular basis. In the case of Int#14, despite being the structure decentralized, the biggest library of the network acts as the *de facto* leader, as the structures were centralized in their premises.

> Int#14 “the library does lead because the project manager is based here and I'm based here and we have access to infrastructure and skills to a greater extent.”

Within partnerships, discussions need to be held among the members to find solutions and setting up common policies for the preservation of the collections when they are diverse, as Int#16 stated. Int#3 stated that in the case of partnering with other institutions and still outsourcing the storage of data as they are doing, it is important to keep the separate contractual relationships with the partners, even though they could be aligned in the same principles.

> Int#3 “it might well be that we encourage other institutions to work alongside on similar principles, but we should keep it l egally and contractually separate from anyone else.”

Anticipate future needs

Despite it might be seen as a general strategy to avoid losing opportunities in the future, forward thinking and re-evaluating decisions based on the potential future scenarios has been pointed out by Int#13 as an activity needed to be included in the design of the systems and planning. If something may inhibit or prevent a future potential development they need to re-evaluate the decision. Similar idea was expressed by Int#1, who stated that his way to look at the profession was to be “acting as a steward until the next generation of archivists come along (…) reacting and adapting situations within your timeframe.”

> Int#13 “we realized that the majority of users of our data don't exist yet, they haven't been born, so when we make a design decision we always evaluate and we say, ‘ok, we know what it all allow us to do’ but we also have to say what might stop us from doing.”

Similarly, Int#3, taking in consideration the uncertainties of dealing with the research data, commented that the option of outsourcing should be kept flexible. Alternatives might arise in the future, such as national or regional infrastructures or consortia with other institutions and therefore they need to make sure that it is possible to switch between the possibilities.

Having an estimation of the content to be ingested would anticipate the costs in which an institution may incur in the short to the long term. Int#1 noted that they have a detailed estimation of the increase of the content over the short term, but considered difficult to put figures in the medium to long term. Nevertheless, the short-term estimation was crucial for the decision to procure cloud storage and further review of the decision would be made if the balance between cost and benefit changes.

> Int#1 “we have a good estimate and a good idea in terms of what's the content that it is going to be ingested and preserved.”

To be able to keep an eye on the future needs, watching the sector through technology and community watches was mentioned by Int#1 as one of the activities they are planning to do in their organization. Int#7, an e-Infrastructure service provider, stated that they were monitoring the status of the digitization activities in their country to be aware of the needs and the situation in general.

Audit and self-assessment

Some audit and self-assessment frameworks were mentioned by the interviewees. Depending on whether they were service providers or memory institutions, there are different uses.

- Digital repositories: DRAMBORA, TRAC, Data Seal of Approval (DSA), ISO 16363 and nestor catalogue of criteria for trustworthiness. Five of the interviewees stated that their institution usually conducts self-assessment using one or more of those methodologies for the evaluation of their internal organization and processes, including Int#4 whose institution offers centralized services for digital preservation. Additionally, two of the memory institutions also went through the process of external audit and certification using TRAC or DSA, and one of the service providers certified using TRAC criteria.

- Quality (ISO 9001) and information security (ISO 27000) certification frameworks are commonly used by service providers to get accreditation on those aspects of their systems and services. It was reported by some of the service providers that they already were certified or were in the process of implementing the standards' requirements and also remarked by some of the memory institutions that their service providers were certified.
- Own catalogue of criteria about the services and guidelines were also reported as being used for self-assessment.

Accreditations, especially in terms of security, were considered relevant to trust the service by for some of the participants. Some of the interviewees expressed that using accredited providers could be an extra layer of guarantees.

> Int#5 "External accreditation, especially with regards to security, is also necessary."

Regarding criteria for trustworthy digital repositories it was pointed out as a factor helping to build up trust among the stakeholders, especially in the case of certification. Some interviewees expressed that it was considered relevant for the community of their repository, being those users or partners (Int#9, Int#13) or customers in the case of Int#2, even increasing the number of customers after the certification process.

> Int#9 "we have a very good TRAC record, a very long history of providing secure trusted preservation and no issues, so there is a level of built up trust among the community."

> Int#2 "more people talk to us, more new people and we have a couple different new big programs that we're working on and one of the reasons that people said is we did that certification"

Int#14 stated that the list of criteria of TRAC would be useful for her institution to assess the offer made by a third-party service, and therefore help them on the selection of service providers. In the case of assessing the cloud provision, she mentioned the guidelines published by relevant institutions, such as the National Archives or the Government. Int#6 and Int#5 commented that sources of criteria to get guarantees and measure the reliability, security and quality were the contracts and catalogue of services of their service providers.

Besides being an external proof towards stakeholders of the reliability of the organization, another reason pointed out to use standard criteria for assessment, specially done internally, was the usefulness for the validation of organizations' processes. Int#13 stated that their organization use DRAMBORA for that purpose, excluding technological aspects. And Int#14 mentioned self-assessment against ISO16363 for a similar purpose.

> Int#14 "We have been doing self-audit with ISO16363 as well of our own internal infrastructure and organisation."

Nevertheless, Int#13 stated that a framework for technological validation was still missing.

> Int#13 "You can come back and say that all your processes work fine and that there is good resilience there, but whether the basic technology, hardware and software and things like that actually works as advertised, is something you can really do other than testing."

Int#2 remarked that the processes related to certification and assessment requires an effort that, despite the benefits, it was not possible for them to do other audits besides being certified with TRAC. In this case, he reported that they were waiting until the ISO standard for repositories was fully released and international agreement was taken to invest the effort of obtaining certification according to its criteria.

> Int#2 "so it is an official metric that these repositories can meet and we are waiting for that to be completed because being honest, being certified is a lot of work and it takes a lot of time and a lot of effort and so we don't want to do it very often if we don't have to."

The UK G-Cloud framework for procurement of cloud services in the public sector was mentioned for some interviewees. Cloud providers are accredited by the framework and easier the task of the institutions on the procurement process, plus providing an extra-layer of trust (Int#1). Nevertheless, Int#3 stated that suppliers may be registered in different groups or bodies which may provide assurances on their reliability; in contrast, despite reporting to those bodies, no action could be taken by them to give support on improving the situation when something went wrong in his institution. Therefore, Int#3 didn't consider this aspect so relevant in the decision-making process.

> Int#3 All the certifications from external bodies, I think they are good, but only a small component of the decision making.

Risk assessment

Most of the institutions reported the use of risk assessment as a tool to support the choice of the most suitable implementation for the organization, supporting decisions such as the convenience of keeping the implementation in-house or using a third-party.

> Int#15 "It is a matter of assessing the risks, cost, benefits..."
>
> Int#1 "it is balancing the costs and the risks really, and I think the cost will simply be too prohibitive to do all this within the [institution], cloud storage has provided a really good option."

Risks management is also used to deal with the risks related to the chosen option, especially when integrating third-party services, due to the trust needed in the provider's performance inherent to those relationships. Institutions using those services reported assuming certain potential vulnerabilities, trying at the same time to meet the shortfalls identified through mitigation strategies.

> Int#9 "it is assessing your risk targets and also what you can do to prevent risks, so preparation, 'trust but verify'."
>
> Int#12 "So definitely there are very big issues and I don't think we've solve all the long term preservation and sustainability issues, but I think we are at least aware of the risks and what we need to do to mitigate those risks."
>
> Int#5 "During the pilot we hope to identify further risks and find ways to mitigate them."

Readiness assessment

Int#14 explained that conducting a survey among the members of the consortia was the first step to initiate the project of digital preservation, state the needs and requirements and decide over the suitable options considering the situation of the organizations.

> Int#14 "from that survey, we really identified huge areas of skill gaps, resourcing gaps but also repository gaps. There was nothing out there, no repositories, there was no infrastructure or architectures available to provide access to preserve digital information."

Benchmarking

Before engaging third-party services or putting a new service/system in place, two interviewees reported activities related with benchmarking. Int#13 reported using what he called "peer-review", consisting on exchanges with peer institutions, with which they work closely, to review each other developments.

> Int#13 "you take an architecture waiting to set up and we find other institutions, we work very close with Stanford for example, and we basically review the way each other have structured the repositories, how we look at delivery, how we look at storage and things like that."

Int#3 mentioned that, as part of the decision-making process for the procurement of a service provider for the storage of their data, they compared the practices of the different providers by talking to other customers of those providers. They do site visits to be able to approach the actual technicians dealing with the service in the institutions and interact face-to-face, instead of managers or other general staff. He mentioned that in the process they could not find customers that had actually left the service, but that would have been interesting to learn about how difficult that process was. Another aspect that Int#3 pointed out is that despite being their institution is a university library, contacts made were not necessarily in the academic sector, but from companies hosting big data with the service provider that already would have similar experiences related to the new needs that the library is facing regarding the storage of research data.

> Int#3 "make sense to speak to other customers, people that can provide what their experiences have been, how to improve"

Procurement process

Procurement is an essential part of the process of contracting services from third-parties for many organizations. Some of them, certainly the organizations in the public sector, follow well defined and structured processes to ensure transparency and accountability. Accordingly, Int#14 mentioned that the procurement rules have to be followed, especially when large investments are made.

> Int#14 "if we are going to invest considerably in something, we have to procure it according to our procurement rules."

In the case of the UK, the G-Cloud framework was mentioned by some of the interviewees (Int#1, Int#14, Int#5). The framework has established agreements with cloud providers that can offer services to the public sector organizations, making the process of procurement easier and giving certain levels of assurance on the reliability of the services, as agreed by all interviewees mentioning the framework. G-Cloud accredits that the service providers must fulfil certain parameters, such as those related with security and data protection, compliance with a set of service definitions, rules for tendering process or the public contracts regulations, plus other assurances in relation with the service management and commercial aspects.

> Int#1 “We use something call the G-Cloud framework. It is basically a big UK government initiative to encourage government and obviously [institution] to use the cloud services as much as possible.” (…) “the whole point of the G-cloud is that it provides accredited suppliers as well so the Government and Government organizations can actually use.”
>
> Int#5 “Procurement was straightforward through the government’s G-cloud procurement framework.”
>
> Int#14 “all the procurements can be done on that (…) it is much easier because it's been pre-approved and the procurement things are so tricky over here. That is a very useful thing to have.”

Additionally, Int#1 stated that, besides using the G-Cloud to support their decision in the procurement of services for digital preservation, his institution also follows its own processes, including additional financial checks and the viability of the companies to which they outsource services.

Int#3 described the process followed at his university for the procurement of the storage provision for the preservation of research data. After identifying the vendors suitable for their needs followed by a call for tender’s proposals, they organized two meetings with the suppliers, a panel with the main stakeholders (research office, the library, the IT department and academics) and a technical panel, from which a technical report was made.

> Int#3 “And they look each of them and critique them, in key issues like cost but also things like convenience of service, how easy will be to use for the academics and licensing” (…) “and then have a separate meeting just looking at the technical aspects of the solution.”

Additionally, with the objective of actually getting a first-hand impression through other customers using the services, site visits were also carried out by the institution’s staff as explained in Risk assessment.

Getting assurances on the service provider’s viability is one of the aspects that is usually examined during the procurement. It gains importance in the case of digital preservation, because the agreements are usually established with the aim of maintaining a long-term relationship. Nevertheless, despite the current situation can show good health on different aspects related with the viability of the companies, a component of trust is also in place.

> Int#3 “you need to have the trust that the company is going to be around in 10 years.”

Pilot phase and review of the selected option

After the process of procurement and once the preferred option was selected, some of the interviewees' institutions undertook the first period of time as a pilot phase of the project, subject of a first revision in a span of time from one to three years.

> Int#3 "Treat it as a pilot, see what issues there are because there might be things that we haven't considered and then correct those either by changing the agreement with the supplier or pulling out completely."

> Int#14 "we did a test installation of Archivematica in the cloud and we linked it to cloud storage, which was Microsoft Azure Cloud at that time for testing purposes." (…) "As I said, at the moment is very much with testing and we are re-evaluating."

Ongoing assessment

Some interviewees manifested their intention of continuously assess the option selected to fulfil the digital preservation needs of their institutions. For instance, Int#1 commented that in the medium to long term the rate of increase of the collection was uncertain, and therefore they need to re-evaluate whether the cloud storage is a suitable option or not for the long term.

> Int#1 "obviously we'll be looking at cloud storage and see how viable that is over the long term."

Int#3 also stated that they need to fulfil the requirements in relation with the storage and preservation of research data, but as it is likely that other options might arise in a few years, they will reconsider whether the outsourced one is the most optimum solution.

> Int#3 "I think that one of the vantages of outsourcing is the fact that we can review it."

In a general sense, Int#12 commented that he assumes that relationships with third-parties in digital preservation have to be understood as temporary. Organizational requirements may be fulfilled during a certain period of time, but there might be changes over time that make the institution feel the need of moving the data to other services.

> Int#12 "I know that is only temporary, so I know it is not going to be a forever type of concept. I have to move the data at some point in the future."

Cost analysis

Cost-benefit analysis was reported by Int#1 as a tool to support options appraisal in the decision-making process of selecting the cloud services option. One of the aspects considered was an estimation of the collection growth in the short term to anticipate the costs. Nevertheless, as the medium to long-term predictions were more difficult to perform, reviews on the analysis have to be done.

> Int#1 "we have tried to anticipate as best as possible the cost involved so we do have an estimate in terms of scale content, that we can pretty accurately put figures in the short term, but in the medium to long term it is more uncertain so we'll just going to have to see how it goes."

Moreover, they incorporated the risks identified into the analysis and even considering the additional costs incurred to mitigate those arising by using cloud for the storage of the assets, Int#1 commented that balancing the costs the option of outsourcing was still more beneficial for their institution.

Int#14 mentioned that they were looking at three different options and assessing which one would be more beneficial in their case, taking into account the costs in which they would incur. To support the analysis, the interviewee mentioned that they were doing pilot testing and re-evaluating each option. After the analysis phase, she commented that a business case was going to be elaborated to support the project.

> Int#14 "we chose *Archivematica* because it was open-source and we didn't have any money, but of course nothing is free, is it? (…) So we are at the moment looking at *Preservica*, (…) a third party and they produce everything in the cloud so we might go with them because the prices are coming down considerably or the library might put an instance and use that as a repository."

On the other hand, the idea of looking at the cost-effectiveness to assess the options was stated in several of the comments done by the interviewees, as was already discussed earlier in this chapter among the benefits perceived.

Despite a formal analysis of the cost-effectiveness was not described by the participants, it seems that its principles were implicit in their decision-making processes.

> Int#12 "I think that commercial cloud storage can still be not quite as cost effective as local data centre storage. But I think it is increasingly a viable option."
>
> Int#3 "We have to make sure that, if we do it in house, is it going to be as cost-effective as doing it outsourced."
>
> Int#10 "At a fairly low cost, but works well because we have tested it."

Business case

Using a business case as a strategy for getting funding for the project of digital preservation was mentioned by Int#14 and Int#1. Int#14 commented that the business case established a roadmap towards reaching their goals. Nevertheless, she mentioned that a new business case has to be elaborated to continue with the evolution of the project, once their choice of a system has been made. Similarly, Int#1 also mentioned that the business case kept evolving until they found the solution that they considered could fit better their institution.

> Int#14 "So we created a business case which was really directed to getting funding. So we did the business case with a roadmap with where we did want to go and one of the most important things was just selecting a repository architecture and finding the way of increasing skills."

> Int#1 "the business case to actually find investment and resources involved in this digital preservation project. So it's all been a continual evolution and a continual iterative process within [institution], and actually gaining support and gaining investment."

Having a clear value proposition in this context may be essential. For instance, Int#14, whose institution is outsourcing to cloud and may be looking into use a different business or funding models, mentioned that their actions were not be driven by profit, but to enable digital assets to be preserved in the future.

> Int#14 "The whole point of it is to enable digital assets to be preserved in the future. (…) it is a cultural thing, we are cultural institutions. Our aim is to preserve material of permanent interest for the nation, anything we do comes from that. We are not driven by profit and we wouldn't make money out of this anyway, would we?"

Raising awareness among stakeholders and funders

Raising the awareness on the need of digital preservation has been mentioned by different interviewees as a way of ensuring funding and sustainability of the activity, whichever the chosen option was. In this sense, the establishment of mandates for the deposit and preservation of research data has increased the demand of the services and the ability to establish a cost model for the in-house services, as described by Int#13. Collaboration among different units within his institution allowed the change of mind-set.

> Int#13 "we've been doing a lot of work with our own research services department so the people cost for digital preservation (…) are aware of the need, people are aware of what they need to do, but also are aware of the benefits."

Int#14 described the work in terms of advocacy that the consortium has been doing "to raise the profile of digital preservation", and the fact that they were trying to get it back in terms of funding.

> Int#14 "but we are very limited in terms of resources and as I said maybe the advocacy thing hasn't worked as effectively as it hopefully it will be."

Related with the mandates and legal obligations, there is some discussion on whether they actually lead to the achievement of digital preservation of the assets. In the case of Int#10, they consider that despite mandates on the deposit of publications and data are made, what is finally relevant for the institutions and will eventually allow progresses would be to find solutions that can match the requirements in an easy and effective way. On the other hand, Int#3 considers important that the mandates are actually enforced by high-level institutions, because most part of the success in this case is to persuade the researchers to actually perceive the need of the preservation of their research output. Nevertheless, Int#9 stated that the synergies generated from a bottom-up approach would be critical to gain awareness and ensure funding, and therefore, efforts have to be made to inform the community. As an example, informing stakeholders about the benefits of preserving and giving access to their research output and how it might increase their research impact (Int#13).

> Int#9 it is also important for us as a community to be a transparent and to keep them posted, if they are aware, what happen in the next years maybe the funding agencies or maybe even the PI start to care.

A greater involvement of relevant stakeholders in the activity has been suggested. From the procurement process, inviting representatives to take part on the panels with the service providers, gathering their opinions and lowering their concerns (Int#3); to a major participation on the process of curation of the resources, with the support of the library as a consultant and active participants from early stages of the information lifecycle (Int#13).

Business and funding models and the use of cost models

With the purpose of economic support for the needs of digital preservation, whether it needs to meet the expenses of outsourcing and partnering, or because there is an increase of the costs due to changes on the requirements, some of the interviewees commented on the potential or actual use of different business models and funding strategies.

Consortium represented by Int#10 mentioned that the costs of the partnership, either monetary or related to labour or infrastructures used were transferred to the universities members of the consortium. Int#16 commented that in their partnership they were looking at models to be able to combine the different needs of the members, for instance in the case of start storing research data in their PLN. Despite them not having a model at the time of the interview, the partnership was looking at examples from other institutions.

> Int#16 "if any institution comes up with data, they pay for it, (…) and then they can buy storage for all the members."

Also related with the new requirements of storing data Int#13 commented that to support the infrastructure, which is a private cloud in their case, they established a mix funding model. The reason argued is that preserving research data increase the need of scaling much quicker in terms of storage. In this case, they have a cost model for the research projects and charge them up-front to preserve their data. To calculate the costs, they consider that over time the cost of retaining data will drop as storage costs drop as well. Int#13 also reported that this funding stream is quite critical for the library, as the operational funds are not enough and there are still unfunded research projects whose data will be preserved but the costs might be absorbed by the library.

> Int#13 "So having all that cost model we can basically charge projects to say, we will archive your data, and say how much it will cost per terabyte."

The partnership represented by Int#14 was also considering a new business model to support the activity after the initial funding received by the public authorities. Despite the alternative did not seem to be very clear yet, they were considering services such as one to the general public to be able to use for their own personal data archiving, for instance.

> Int#14 "maybe we have to open at to not [consortium] members and running as a service. I don't see that happening at the moment, but it is obviously something that you have in the back of your mind."

Public funding was also pointed out as a source to support the development of specific projects. Int#6 commented that two projects were funded by a research council that allowed them to "develop the technical infrastructure and address the organizational matters" and an evaluation of the scalability and trustworthiness of the infrastructure supporting their project of digital preservation, in collaboration with their e-Infrastructure provider.

Sources of public funding at the national level usually provide initial funding to support the start of a project or a collaborative activity such as national or regional services, and then the institutions involved should be the ones funding beyond that (Int#3, Int#14). European projects provide funding for large-scale research projects through the cooperation of institutions, as reported by Int#13. Although he pointed out the difficulties of translate those projects into services.

> Int#13 "It is actually quite hard to construct a business model for a service like that, unless is funded centrally by an organization like the EU."

Contingency fund

Long-term funding of the activity was reported as one of the organizational areas of risk by several interviewees. Int#12 explained that his institution was maintaining some funds reserved to ensure that, if additional funds are needed, they will have access to them in the future. Despite it does not give assurances on funding the activity over the long term, at least it could mitigate the lack of funds in a particular moment of need.

> Int#12 "put money into a long-term bank account to help us ensure that money is available to maintain the content in the long-term."

Contracts and Service Level Agreements (SLA) guarantees

The relationships under study, especially those between service providers and customers, are usually regulated by contracts and service agreements. Therefore, a vast majority of interviewees referred to those agreements as the instruments where the conditions and guarantees were stated. The interviewees explicitly described some of the most relevant aspects for them contained on the formal agreements. Some common aspects were mentioned by different participants, regardless the type of relationship:

- Agreements in terms of data protection.

Agreement based on a legal report with additional guarantees described by Int#10, and described more in detail in the section dedicated to Legal compliance.

- Limitations on legal liabilities.

The service providers may limit their liabilities in cases such as copyright non-compliance by the customer depositing the data, and the contracts state the roles and responsibilities of each of them, as reported by Int#2. Similarly, in the partnership of Int#16, each participant is responsible for ensuring the legal compliance of their own data. More details in the section of Legal compliance.

> Int#2 "And if they do give us that and there are some legal problem arises it's their responsibility, it's not ours, because we basically say we don't know what the content is."

- Exit-strategy.

In cases such as the termination of the contract or the service provider going out of business that was earlier described as a risk (see 4.4.1.3), the conditions for data portability to export the data to another service or to the own datacentre, were mentioned by some interviewees. In the case of using service providers, Int#1 commented that the contract contained particular clauses addressing those aspects. In their case, the responsibility was placed on the side of the service provider. Int#15, a provider of services using cloud commented that using their service, the portability of the data was guaranteed and that there were not additional costs for the customers. Some interviewees (Int#1, Int#2) also explained that there are timeframes specified for advance notice in the case of going out of business or customers leaving the service.

> Int#1 “we try to ensure as best as possible through our contracts and SLAs that it will be possible to actually get our data back and the responsibility for that is actually placed on the commercial organization.”
>
> Int#2 “If we for some reason are going to be out of business we guarantee that we will give you at least a year notice and we also provide an exit-strategy to either give you your data back or put it somewhere else where you would like to put it.”

Another comment in relation with the termination of a contract of a service was done by Int#9 who stated that having multiples copies of the data as they do with different service providers, and taking in consideration the expenses that recovering the copies from a cloud provider might imply, they would ask the cloud provider to delete the copy or leave it in the cloud and terminate the contract, rather than export it.

> Int#9 “I don't believe we worry about giving the content back from them, I think that can be expensive so what we would do is essentially move on, just leave it or have them deleted, because that is not the only copy, it is just one copy.”

In the case of collaborative arrangements like the PLNs (Int#10, Int#16), having an exit-strategy did not seem to be perceived as much as a need as in the case of those using third-parties to provide the service. Nevertheless, some of the aspects reported that are needed to be considered were the case of no longer availability of the open source software, what to do with the content if the network ceases to exist (Int#16) or whether the data distributed among other members would be handed back in case of leaving the network (Int#10). As it was already mentioned, as a result of the need of a minimum number of members in those networks, Int#16 also showed concern about the protocols to follow if the network remains with fewer members than the minimum required.

Despite all stated above, two participants shared concerns about the exit-strategies. Int#14 whose consortia is outsourcing to cloud, stated that despite the exit-strategy was very important, other strategies such as keeping a local copy should be in place. Int#13 mentioned that exit-strategy has become a major source of risk and expense for their institution, and therefore one of the reasons to use their own private cloud with the purpose of storing the institution's data.

- Ownership of data.

Some participants expressed the need to make clear in the agreements, once the assets are transferred to other organizations.

> Int#16 "It will be clear because we will be signing agreements with them and they are just preserving our content."

Related to the issue of ownership, some participants also commented that there is a need for a different mind-set compared to the traditional datacentres and also related to the type of data (open, public) that is deposited in third party services.

> Int#1 "it is an interesting question the ownership of the data once is in the cloud and how you get it back and it just requires a different bit of the mind-set over traditional data centres (...)"

There are also some aspects of the agreements that are more specifically of interest for those outsourcing the service, either to a commercial company or to a centralized service:

- Levels of durability, integrity of data or security.

Those aspects are usually included in contracts and SLAs. Therefore if the service provider does not comply with the specifications stated in those agreed levels, penalties might be put in place.

> Int#5 "The conditions of our contract and external accreditation provide ways of assessing reliability, security and the level of service."
>
> Int#1 "the issues that have to be resolved around and the information security as well. That was another key requirement that had to be satisfied."
>
> Int#1 "we rely upon SLAs and contracts (...) so we don't actively checking once is in the cloud we rely on the commercial organizations to do that."

- Financial penalties or insurance.

In the case of failures or loss of service levels, the contracts or SLAs may contain different types of compensation measures. Nevertheless, for some interviewees (Int#14, Int#9), even if the penalties were seen as positive, they also commented that other mitigation strategies have to be considered, because if there is data loss, for instance, the credits or financial compensation will not be sufficient. But on the other hand, Int#3 commented that the fact that the service providers, whose aim is to make profit, promised financial compensation gave confidence to his institution.

> Int#14 "you have to make sure that you build in penalties in the contract and you need to build the insurance in the contract level. (…) so the trust is there but you also have to have something concrete underlying it."
>
> Int#9 "Even if they say 'I'll give you a hundred thousand dollars' that's useless if you lose your data."

- Third-parties subcontracted by the service provider.

It might not be explicitly specified in the contract. Nevertheless, Int#3 considered a requirement to acknowledge which companies work with the supplier and to be aware of what the procedures are and that terms and conditions are the same as agreed. Service providers such as Int#2 expressed their need as suppliers to have contracts with the third-parties they subcontract, so the conditions and expectations are clear.

> Int#3 "They have to tell us what other companies they use, which procedures they use and also any contract the university have with them, they have to make sure they have the same terms and conditions with the supplier."
>
> Int#2 "so we do have signed agreements and service agreements with them about what our expectations are for them and what their expectations are for us."

One of the conditions that institutions need to be aware of when suppliers subcontract services is whether there would be additional charges, as Int#3 explained.

> Int#3 "some of them, rely in other companies (…) you can store vast amounts of data but they charge you when downloading data."

Int#11 stated that in their case they do not use additional third-parties to provide their service. Their company and the other suppliers used for the activity of digital preservation, even having to have certain levels of relationship, should have to be kept as different contractual relationships done by the customer in their case. Roles and responsibilities become clearer this way under his opinion.

> Int#11 "defining who have to take care of each problem so they know who of us to call."

More specifically, some aspects of interest that should be considered on the agreements under the opinion of interviewees whose institution takes part in a partnership are described as follows.

- Terms and conditions of becoming part of a network.

Agreements on governance policies, membership conditions and other agreements were stated as relevant by Int#16.

- Independent contracts of partners with service providers.

This option was mentioned by Int#3 in particular in the case of collaborating with other institutions. Even working under the same principles, he commented that the institution should remain independent and able to make decisions such as pulling the data out of the service provider if they need to.

Negotiation of the contracts was reported as possible, being able for some of the institutions to introduce some of their own requirements besides the fixed general terms.

> Int#1 “there were specific parts of those SLAs that we had to specified.”

Int#1 also commented that it was possible to introduce specific requirements in the agreements with the cloud storage providers but some negotiation was needed. In his opinion, the reason behind that was that digital preservation requirements are new for those providers and they have not dealt with those types of needs before.

On the other hand, the service providers commented that they have standard contracts that have different levels of customization. In the case of Int#4 have been reviewed every year to adapt it to the customer needs. In the case of Int#2, the majority of the contract cannot be changed because is where the guarantees of the quality of the service are stated. Nevertheless on top of those conditions, specific needs could be added, for instance if they receive additional services, or a different structure for the payments.

> Int#4 “We did formalize contracts and service level agreements to match external customers’ expectations and this is a constant dialogue so we develop a new version of our standard contract at least once a year.”

> Int#2 “in a single contract, there is probably 80% to 90% that has to be the same for everybody, (…) but in terms of the guarantees that we offer and the reliability (…) will not change. That has pretty much to stay at the same that basic level of preservation.”

In addition to that Int#3 also reported the need of support by the university services or legal advice to review the contracts and ensure they are appropriate. Similarly, Int#10 also mentioned that they used legal support to guarantee that the agreements were adequate.

> Int#3 “decide what we need and get them to the University services, go through the contracts and ensure that they are appropriate.”

Customization of services and conditions

Standard services not always match with the needs of a particular institution, and some of the interviewees expressed that the services they were using or offering were adapted upon the customer requirements.

> Int#2 "we also try hard to meet their own special needs so that we work closely with them to try to define our own services so that it is sometimes much more customized for their individual needs, rather than everybody gets the same service."

Some of the customizations were reported to be done in the technological aspects:

> Int#1 "out-of-the-box solution for us and we heavily customized as well, and developed it further."

And others related with organizational issues, such as payment methods that could suit better the institution's requirements, plus new services in addition to the general ones, that would translate into the agreements.

> Int#4 "Apart from the standard contract, we might do special agreements with individual customers to match any specific needs and requirements."
>
> Int#2 "we would like to give you money in a different structured way and so we can put that into the contract."

Legal compliance

- Copyright and licensing.

The opinion of interviewees when commenting about the issue of copyright protection of the materials deposited within the network or third-parties was related with the own responsibility of the institutions owning the digital assets. In a general sense, the ideas expressed to mitigate risks related with copyright implied that the institution should establish its own policies on that regard, be responsible for clarifying the rights, reaching the agreements with the right-holders when necessary, associate the data with the appropriate licenses or to not store it within third-parties' premises if those aspects are not clear.

> Int#16 "for sensitive data or for copyright issues is the responsibility of the institution which is adding that content so it is not the responsibility of the PLN or other members."

Partnerships such as those in which Int#16 or Int#10 take part of do not check or monitor if the information stored in the PLN is non-compliant with any regulation. It is the individual institutions' responsibility. Similarly, service providers such as Int#2 stated that they do not exam the content, therefore they do not take responsibility for non-compliance or any potential liabilities, and the contracts with their customers are clear on that aspect.

> Int#2 “when they sign a contract they guarantee that they are not giving us copyrighted materials that they not own the copyright to or illegally obtained materials.”

In the case of Int#4, partnering with other institutions they consider the importance of the staff commitment to comply with legal frameworks that should be translated into the signature of agreements by both organizations expressing their compromise.

> Int#4 “staff on both sides need to comply with regulatory and legal frameworks like personal data protection which means there needs to be formal agreements (written) between organisations to comply with national laws.”

If the data is going to be made available through the service providers, Int#3 expressed the importance of taking care of which licenses are attached to the digital objects. In the case of one of the service providers they were assessing, the licenses they would attach to the data were exclusively Creative Commons with no need of attribution, which did not suit their needs. Nevertheless, negotiation with some service providers in the licenses was possible, and offered them alternate licenses.

> Int#3 “we want to use licenses that you have to attribute to [University] as the main source of the data and respective authors.”

- Data protection: sensitive data, confidentiality.

In the case of cloud customers, a common strategy reported was to keep sensitive and business critical information that needs to remain confidential in an in house data-centre, especially when the levels of security provided were not enough for levels of protection required.

> Int#1 “when it gets ingested in the repository is either open or closed if it is in the public domain or publicly available then its open and it gets stored in our service provider but if it is closed it gets stored entirely within [institution's] own data centre.”
>
> Int#5 “We are investigating the feasibility of using our own datacentre to store protected information, and the cloud to store open access information.”
>
> Int#9 “The most highly sensitive data never leaves our building, we don't make copies of it externally, we keep it here.”

Alternatively, using cloud services with higher level of security assurance could be an option. Nevertheless, a service provider Int#15 stated that despite there are possible options available in the market, the prices of those options would be higher than using services such as Amazon, which does not guarantee as much protection as needed in some occasions depending on the Business Impact Level of the information.

> Int#5 “Preservica are investigating whether it is possible to obtain a higher security accreditation.”

A common security strategy to protect the data in case of disclosure is the encryption of the digital objects (Int#9, Int#15, Int#3) in order to protect them when they are sent through the networks, as happens when the content is submitted to the service providers or when they are in the cloud.

> Int#9 “we also encrypt all our content or at least the confidential data when we send it to them, if anything is loss on the way (…) we know the content at least is protected against malicious acts.”

> Int#3 “they store the data encrypted so when they actually store it in Amazon they are storing the data [in a way] that anyone, including Amazon can read.”

Int#3 reported different options that were offered by the suppliers to deal with the content that for some reason have to be protected. For instance, use of different locations depending on whether the data is confidential or not; establishment of different schedules in case the data has to be made available or to keep it private; and there were also suppliers that stated that they would take some financial responsibility and that they have financial penalties.

> Int#3 “If they break an agreement, they give us money back.”

Legal assistance may be needed when some aspects are not clear. Int#10 commented that they needed a legal report to support the fact that they were partnering with institutions outside their jurisdictional area. That legal report was the base for an agreement with the partnership in which the cooperative gives assurances on the protection of the data entrusted to the network such as not making public the data and that no use other than that of preservation could be performed.

The copies for preservation in the partnerships of Int#10 and Int#16 are kept as dark copies, therefore, despite they are stored in other institutions, those institutions cannot be aware of what the content of those documents is. Therefore, if access is provided, it is done using a local copy in the institution responsible for the resources, and not from any other caches since their only aim is to keep additional copies of the content.

- Compliance with other regulations.

For instance, UK public sector institutions reported that they have to be compliant with the *Freedom of information Act*. Therefore, as stated by Int#1, his institution needs to make sure that the service provider, cloud in this particular case, comply as well. In this case the procurement process was the mechanism to get that assurance.

- Jurisdiction.

Control over the jurisdiction where the data is located has been pointed out as one of the requirements of the institutions storing data in cloud. For instance, Int#1 stated their requirement of keeping the data within the EU, and Int#9 commented that copies were not send outside the US, in their case. Accordingly, one of the service providers, Int#15, commented that they have customers both in the US and the EU, and instances of cloud in both geographical areas. Therefore, they let the customers chose their preferred option.

> Int#1 "we specified the cloud storage providers themselves have to be within EU jurisdiction (…) because the cloud is geographically diverse so what we didn't want is be held in whatever fashion in a different jurisdiction, the United States for instance or anything like that. It has to be within the EU."

Other service providers not using cloud but grid technologies, like Int#2 commented that they have always provide this guarantee, because the nodes of their network for distribution of copies are clearly located. He mentioned that this was a big difference with the offering of cloud providers some years ago, but the cloud providers have gained awareness on the need and offer more guarantees on the location of the data in the present.

> Int#2 "now they've realized people want to know that, that's an expectation."

In collaborative networks, jurisdiction may also introduce limitations for the exchange of sensitive data. Therefore, some data might have to be restricted and not sent abroad or to a different jurisdiction. Despite their potential interest on participating in a collaborative project in the US, Int#13 commented that they cannot join because of the legal restrictions due to the different levels of protection of the data compared to the EU, where the institution is located.

On the other hand, Int#16 mentioned a project that they were going to collaborate within the US, because the content was open and there were no restrictions. But similarly to Int#13, in case of legal restrictions, they will not send any data beyond borders.

- Disposal actions.

Disposal actions are especially relevant in the case of archives, due to the schedules for retention or destruction. Taking place within third-party's premises, especially cloud services offered by service providers Int#1 commented that they have to be enforced according to the organization's policies. Moreover, the interviewee stressed the idea that there was a need for negotiating requirements, using policies approved by high-level structures at the organization as a base.

> Int#1 "We have our digital preservation policy and strategy as well. We have that underline basis to actually negotiate with them so we just try and tie whatever we are going to outsource the best as possible into that, really. We got that policy in place and approved by high management (…) that we can use to enforce specific requirements."

Int#15, a service provider subcontracting public cloud for their services, was posed with the question of whether deleting data physically from the cloud is actually possible and effectively done, due to characteristics of the public clouds, such as loss of control on the copies or location of the data at a given moment, as stated by Int#9. He commented that it is not in the interest of the service providers to maintain the data, since it would be costly for them. They do offer an interface through which the customers can select the data that they need to delete.

Agreements with data providers

Deposit forms clarifying issues with the data depositors, such as the confirmation that they hold the copyright ownership of the data to be stored in the repository systems are commonly in place. Whether the data providers acknowledge and agree on whether the data is stored within third-parties was asked and Int#9 commented that it was not explicit in the agreement where the data was stored or any request of consent in that aspect. Nevertheless, for what the consent was asked was to process the data in any way to ensure preservation and access.

> Int#9 "they give us permission to re-disseminate, to describe, catalogue, validate and then to store, translate, copy or reformat the data collection in any way to ensure its future preservation and accessibility, so it is not explicit about where we store copies."

Nevertheless, despite explicit consent was not requested, the website of the institution contains detailed information about in which services the copies are kept and how the data is treated after deposit.

Tackling with the impact on staff of outsourcing or collaborating

Consortia have been pointed out as intermediaries absorbing some of the administrative and technical tasks of digital preservation using third-party service providers or partnering with other institutions. In the case of Int#10, the consortium is part of a cooperative. They are preserving the repository of theses on behalf of a number of universities that are benefitting from their services. For instance, the technical task of monitoring how the system is working and problem solving; but also legal advice, financial and budgetary activities; communication with the partnership and organization of outreach activities for the digital preservation community as part of the compromise with the cooperative; elaboration of common guidelines support.

Int#12 also reported an increasing number of consortia taking care of the support to the final users or the technical tasks involved in using an open-source solution, as described earlier. Int#9 also commented that their organization as a consortium also acts on behalf of a number of institutions, assuming the tasks of contracting the services with third-parties, technical activities related to the service, among others.

The need of the provision of new staff with dedication to the activities related with digital preservation was not equal in the different cases. Even though new staff was hired by some of the organizations, it could be related to a general increase of the needs just because the organizations were not dealing with digital preservation activities in the past.

In the case of Int#9, he commented that they had to hire and train staff members with specific knowledge on cloud technologies, for instance.

> Int#9 "We have to train or hire staff with experience in working with the cloud providers (…) to ensure that the content is stable over time and responsive."

Int#3 also commented the need to enhance the technical skills of the staff through training on the outsourced solution. And depending on the services they will be outsourcing, the training and skills gaining should be directed to different aspects.

> Int#3 "There is going have to be some training at the technical level depending on which solution is provided. If we take someone that is only going to host the data, then we have to look at how we develop the services."

In some cases, the skills needed were reported as not differing much compared to those that the staff already had because of the easiness at the technical level of the systems (Int#16). Nevertheless, the increase of tasks at the organizational level related with the digital preservation activities, required in some cases staffing the activity with new members to tackle with them. In the case of Int#14, the institution hired a full time project manager in charge of developing the project to start with the digital preservation activities, from the options appraisal to the development of the business case and some other tasks of coordination. In this case, hiring the new member was temporary and subject to the initial funding of the project.

In the case of Int#1, a full time staff dedicated specifically to the planning of the activity was hired, but as a result of the increase of the activities of digital preservation in the organization, independently on whether they are outsourced or not. In the same line, increasing the activities of custody and digital preservation in cases such as the research data management, have increased and modified the activities of some staff members, as reported by Int#3 in relationship with the IT staff, the library staff members and the academics. Nevertheless, despite the fact that the activities in these cases have been outsourced to cloud providers sure has an impact on the tasks, it has not been pointed out as the source of the major changes in the overall processes involved.

In general, the different level of expertise of staff in the institutions has been an underlying condition for the different approaches taken to tackle with the activity. Ranging from the expertise and research capabilities to generate in-house large developments as in the case of the private cloud implemented by Int#13's institution, to a lack of awareness reported by Int#2 on the basics of digital preservation among their customers. Therefore, the perceived need of more or less staff support to the activity, independently on what the option chosen is, seems also to be related with the level of awareness of its needs and the complexity it entails for each organization.

> Int#2 "oftentimes the people we work with are not preservation experts and so they are not necessarily aware of a lot of the challenges and the difficulties and the implications of some of this things."

Besides training, raising the level of skills through collaboration with other institutions has been pointed out by some of the interviewees involved in different types of partnerships. For instance, it is remarkable the case explained by Int#14 whose consortium created a community of practice for the pooling of expertise that the members could contribute with and as a way of distributing the resources and the tasks arising from the partnering activities for digital preservation. They detected that none of the institutions involved could contribute with a monetary input, but they could contribute with hours of dedication of staff members. The assessment of the readiness of the institution served as a basis for creating a roadmap to lead the consortium into the desired level of development including the staff skills development.

> Int#14 “But then we realized, we couldn't do alright, no one institution could do it, it had to be a collaborative thing, so you can share resources and build skills. So we established what is called [community] where all contribute to the skills gaining.”

Clearly defined roles and responsibilities are also important. Int#15 commented that outsourcing to their services allows a clear separation of the operational and functional roles, meaning that the decisions about the collections are kept entirely by the institutions. Therefore, the adequate skills and staff members should be in place, even using an outsourced solution.

Furthermore, it is important to be aware of the overall requirements of the activity clarifying roles and responsibilities of the different processes and acknowledging which ones are owned by the third- party.

> Int#15 “There is the functional role involved in deciding what you are going to ingest into the system, whether or not do a format migration, what your access roles are, cataloguing (…). That is all controlled by our customers. (…) as an IT system there is an operational aspect (…) by our service you don't have to worry about anymore.”

> Int#2 “we do not do any kind of file normalization or file updates or any kind, anything like that and whatever they give us, we will store and preserve and we will give it back. (…) it is up to them if they want to change it, to change it in their system and give it to us. (…) And that is something that we have to be very explicit about so they understand that.”

Communication and support

Int#10, Int#14 and Int#16 commented that they have a range of different ways of communicating in a constant basis with the rest of the members of the partnership. In particular, monthly meetings with the other members of the cooperative, exchange of emails or outreach events.

> Int#14 “we had a big *Archivematica* event (…) we've had around two-three meetings a year.”

The fact that processes related with the physical preservation of the digital objects are automated does not necessarily require a constant interaction in the case of using third-party service providers. From a service provider point of view, Int#2 commented that maintaining the levels of contact in long-term relationships is challenging, but necessary to not lose track with the right people in charge in the organization, because when operational decisions have to be made it is difficult for the provider to trace who to talk with.

A questionnaire was pointed out by Int#11 to understand better their customers’ needs and the levels of satisfaction with the service provided. On the other hand, they also provide a service helpdesk run by the developers directly, cutting intermediaries.

> Int#11 “we are a small company, we are very flexible and our help desk is done directly by the developers that are working in your solution so it cuts intermediaries and you talk directly with the person that is going to solve your problem.”

In addition to the ongoing communication, some service providers (Int#2 and Int#11) commented that they also assist their customers to understand the challenges of the activity and the type of procedures and policies that they need to put in place.

> Int#2 "oftentimes the people we work with are not preservation experts and so they are not necessarily aware of a lot of the challenges and the difficulties (…)."

Partnerships themselves seem to have an important role in terms of receiving support. In the case of Int#14, the philosophy of their partnership was that all partners have a contribution to make, despite the differences on expertise or levels of development. Therefore, this approach developed a climate of comfort and mutual support among the members.

> Int#14 "There is no conflict to the moment. Mutually supportive, I shall say."

Reporting and monitoring

In terms of monitoring the state of the collections, both Int#10 and Int#16 were using the same technology (PLN), therefore the system described did not differ much either. They use a private interface through which they can do the follow up of the reporting about the state of the collections (theirs or others in their machines) that are being checked through the automated processes.

> Int#16 "It is the responsibility of the LOCKSS software to monitor that and to fix any issues in the integrity and the bits."

Besides monitoring the performance of the system, other aspect that could be checked in the partnership model is whether the other members do not fulfil their responsibility of adding the content, as described by Int#16. In this case, they use software called SafeArchive[27] to monitor this aspect and if they detect any failures, they usually send a reminder to those partners that have to ingest data.

Int#4 also allow the monitoring of the performance of the data and the fulfilment of levels of service within the system for bit preservation that they offer as a centralized service. For that purpose, they use open APIs that let the users to connect with their own systems to obtain the full reporting of the performance.

[27] http://www.safearchive.org/

> Int#4 "The infrastructure we have been developing (…) is open source and there are open APIs that allow them to connect their own storage infrastructure to the central services offered by us to obtain a full overview of all their data including regular integrity checks etc."

In the case of commercial providers, it seems that the task of monitoring is not so accessible to the final customers. Int#1, whose institution is using cloud storage and a cloud implementation for the services for preservation, commented that they do the monitoring in their repository environment, but out of that they rely upon the SLAs and contracts guarantees provided by the commercial providers. Nevertheless, he stated that they were looking at other possibilities to monitor the assets in the cloud. Int#9 also commented that despite they received reports from the service providers, a similar level of trust as the one commented by Int#1 needs to be in place, since they don't have the guarantees on whether the reports are elaborated correctly, for instance.

> Int#1 "we don't actively checking once is in the cloud we rely on the commercial organizations to do that."

> Int#9 "but even when they send you those reports you still have to trust that they run their reports correctly."

Similarly, service provider Int#15 using cloud for implementation of their software and storage of their customers' assets, commented that they don't do any checking or monitoring over the conditions of the data stored in the subcontracted cloud services and rely on the cloud provider guarantees.

> Int#15 "they publish a figure for reliability, we don't measure it independently but they say they have never loss of our files. So they do have an ongoing checking program (…)."

To be able to accurately do the monitoring of the collection, there are some actions that can be done before transferring the data to the third-party services, to check in which state is the collection. Int#9 reported that they do checksums before sending the data to the service providers. They also do checks on their own environment.

> Int#9 "So we do a few things when we transfer content, one as we only have the checksums (…) on what the content is at that moment in time so we can tell if the content has changed."

Service provider Int#2 commented that despite they do provide complete reporting on the collections performance, they lack a friendly interface to easier the task of the customers on the monitoring and other interactions with the system.

The service provider Int#2 commented that they do a checking process when the collections arrive to their premises, gaining assurance through this process that the objects are not corrupted, for instance.

> Int#2 "So when we take data in from a data provider, we do a number of health and safety checks on the data. When we get it (…) we can prove that what we receive (…) it's okay it's not been corrupted, is fully functional."

Redundancy of data

Redundancy of data has been commented by the majority of interviewees as their main strategy considered for minimizing the risk of data loss. Three different approaches to achieve redundancy were discussed: organizational, geographical and systems based. The three different approaches are also interrelated and overlap, in most cases.

> Int#4 "Doing digital preservation together with other organisations gives a number of benefits: more secure preservation due to organisational and technological as well as geographical spread."

> Int#12 "the most important approach that we can take is multiple copies in multiple contexts, which has an expense."

> Inr#9 "diversifying the number of copies that we keep and also the institutions that run those copies were big factors [for using cloud]. (…) 6 to 7 copies distributed in different locations using different technology for each one (…)."

The use of distribution of copies as a strategy ranges from the minimum option, that could be the making a second copy of the data to more sophisticated options, in terms of the number of copies, type of stakeholder involved, geographical diversity or the variety of infrastructure as well as in terms of the tools or additional assurances provided.

In a general sense, keeping several copies of data will provide protection against loss of integrity, corruption of the data or bit loss. Also, against economic and organizational failures or natural disasters.

In the partnerships using LOCKSS systems as in the cases of Int#16 and Int#10, both reported that the copies are distributed among the different members of the partnerships and continuously checked by the software, replacing the copies that show differences with the other copies in alternative locations. In these cases, the copies are also distributed throughout different organizations and geographical areas, being located in the same country in the case of Int#16 and in different continents as in the case of Int#10.

> Int#16 "there is issue with data integrity of there are any bit loses, then it says this node has data that is different than 9 other nodes, so then automatically replaces data with the good copy."

The strategy of distributing data among different organizations has been pointed out to be a mitigation strategy in case of the cloud provider going out of business by Int#1, whose organization decided to procure a second cloud storage provider.

> Int#1 "We have weighted the risks and that's part of the reason that we use two cloud storage providers (…) have multiple copies of the same data, so if one cloud storage service provider go out of business we have the second storage provider to actually back on."

In some cases, when the assets are stored with a third-party service provider, copies are held also in-house, but it is not always the case and Int#14 was considering the option of keeping a local copy within the institution besides the one stored in cloud. Int#4 stated that most of the organizations to which they provide service usually keep a copy in their premises.

> Int#4 "Normally they would keep at least one copy of their digital holdings themselves."

The service provided by Int#2 and Int#7's organizations includes the distribution of copies using grid technologies, among different organizations and geographical locations in the same country.

> Int#2 "so each of this copies are very widely distributed and so it should be the case that no problem that affects one location also affects the other locations."

> Int#7 "including heterogeneous storage facilities and geographically dispersed storage (country-wide)."

Additionally, Int#2 commented that his organization, as service provider, was in the process of joining a larger project with more organizations, increasing therefore the diversity of the distribution of the data.

> Int#2 "we can have much stronger guarantees about if something happens to [organization] it doesn't matter because someone else also has the data."

Int#13 and Int#3 mentioned that one of the companies that they have been considering for the storage of their data use the system of escrow accounts, by which copies in tape of the data are held in an additional third party. In the case of the service provider ceased to exist, the copy of data and some of the money paid in advance could be recovered by the institution.

> Int#13 "When you give them money, part of that money is put in storage and they hold escrow copies of the data. So if anything happens to the company you can recover part of the money you have invested for long term storage and you recover the basic tapes with the data because they are held separately."

A strategy of redundancy of systems was exemplified by Int#13, since they were looking at different options to store copies of data in tape, and store them in other organizations with a tape archive.

> Int#13 "We are looking into outsourcing the tape copy of data, or the slow copy of data, that is feasible to move offsite."

Another example of systems diversity was explained by Int#9:

> Inr#9 "we also use different technology to back things up, so we have copies in LOCKSS, Duracloud, which is a cloud provider, Amazon (…) multiple copies locally and also in other locations."

Int#9 commented about the diversity of cloud itself, since a single provider uses simultaneously different datacentres geographically distributed and runs different copies at the same time. Although this feature would only give assurances as long as the cloud provider continues with the business activities or the service levels are kept. Additionally, they use the services of a cloud broker that diversifies even more storing the data in other cloud storage providers. Using this strategy they also addressed the shortcomings that could arise using a single organization, despite being as diverse as it can be in the cloud.

> Int#9 "So it is kind of like diversifying your cloud interest even further, so whereas with using Amazon you're using one cloud provider, that have multiple places, geographic locations, with Duracloud using one provider and then they use multiple other services."

Similarly, Int#12 argued that using organizations with different missions, such as commercial on one hand and non-commercial would be the strategy followed for his institution.

> Int#12 "So if I chose to use commercial cloud service like Amazon, then I'll minimize my risk by storing another copy of my content in this case, in this example in a National-University managed cloud service."

A completely different approach was stated by Int#12, whose institution opens their system for harvesting digitized out-of-copyright materials, in order to ensure that copies are held elsewhere.

Int#12 "I think it does help mitigate or reduce the risk when there are other copies of our content. The fact that they might or might not be in our control I think it is part of having some kind of long-term approach to preservation."

In-house data centre and private clouds

Allowing accessibility to the data when the copies are not exclusively for preservation, can be challenging with remote data-centres or using the cloud due to bandwidth restrictions. Int#13 explained that large volumes of data that need to be accessed, such as in the case of research data, were kept in house in the private cloud of his institution. He considered that there is scope for private companies to offer preservation services using commercial cloud; but the infrastructures are not suitable to allow immediate content delivery.

> Int#13 “I guess there are ways to deal with those issues, but they didn't solve really the need for a rapid access copy.”

For this same purpose, allowing rapid access to research data, Int#3 explained that one of the service providers proposed them to build a small datacentre on campus that would solve some of the technological issues, plus allow access to the institution’s users.

Furthermore, building a temporary own infrastructure might be required in terms of data portability needs when the service provider does not export data to other third-parties data centres (Int#3).

Use of open standards

The use of open standards and open-source software might increase the data portability guarantees and help to avoid lock-in of a vendor or system. Int#4 stated that the infrastructure built for centralized services of digital preservation has been built using open-source software and open APIs to allow the institutions using the service to connect with the central services, facilitating as well monitoring of the activity.

In the case of Int#12, the software developed was open-source and important features are the open standards file structure, well documented that will easier the migration process.

> Int#12 “So if in 5 or 10 years you need to move to another piece of software, the data that you put in is in a very well described well documented structure, so it should be very straight forward to migrate your content to another system.”

Support to open-source software

Implementing open-source initiatives can introduce an extra-burden for the organization in terms of the expertise needed in-house, but some interviewees pointed out that commercial companies were also offering support to the use and implementation in the institutions. For instance, Int#11 commented that their services consist on taking this risk out of the hands of the institutions. He also mentioned that lock-in would not be an issue since another commercial provider or the institution itself could take their place easily.

Similar ideas were expressed by Int#12, whose open-source initiative is also offered with support from different commercial providers. Moreover, he mentioned that not only companies are giving this support but also a number of consortia implementations of the software are supporting their members in the use of this software.

> Int#12 "So there is an increasing number of consortia who are providing the system for their members including support with getting data into their systems and all those kind of things."

Nevertheless, one of the main concerns in relation with open-source initiatives is the sustainability of the project itself. In the case of Int#12 the software developed has a large institutional implementation, the already mentioned support by commercial organizations and consortia, a Foundation as a non-commercial home base for the project, including a member of staff, financial support by the membership, an external community of developers and a number of formal processes around the project. The fact that the creator is not very involved at the current moment and the project continues evolving was pointed out as a good sign of sustainability as well.

4.6. Trust mechanisms used in the inter-organizational relationships

Established reputation and customer base

Some of the aspects that the interviewees valued in relationship with the reputation of the companies were the size of the company and level of infrastructure, the experiences of other customers, the type of customers, how widespread the company services are or the years of experience in the field.

Reputation among a large customer base was mentioned by Int#9 as a factor to trust a service provider. In the case of failure, it may be easier that other customers gain awareness about it.

> Int#9 I think the size of the company, you know Amazon, there are so many customers you hear from people if something is off.

In the case of Int#3 having discussion with other customers of the same company was considered one of the main determinants to trust the service provider and engage in a relationship with them. As already mentioned, visiting other customers with similar characteristics or needs and discuss the issues with technical staff was a determinant part of their procurement process. Besides the references provided by the company (Int#11), Int#3 commented that they also contacted other customers using the service of the service provider, especially those that are partners of the University or the Library. This last aspect was pointed out by Int#3 as important in the sense that without certain levels of closeness in the relationship, it might be difficult that the company that has an agreement in place with the supplier gives a negative impression from it.

> Int#3 "it is useful to get to speak to the people doing actual work and not a general person in the organization."

Another factor related with reputation was stated by Int#5, who mentioned that a factor for trust was the company's "already established reputation working with larger, national institutions." Therefore, she pointed out that the type of organizations that are customers of the company, such as large memory institutions in this case, was a major factor of trust for her institution. Int#1 also adds that their choice is both widespread in their immediate geographical context, but also in other countries. Moreover, he commented that keeping relationship with the customer base, at least in the region was a key factor.

> Int#1 "And I think a number of organizations across the UK are taking the same decision (…) use the same type of technologies (…) and the user base of Preservica itself is quite widespread throughout the world."

Characteristics of the company such as the size, the viability, the level of development in terms of infrastructure or the expertise were also pointed out by Int#9.

> Int#9 "a large company like Amazon that is well established, has very smart people, good infrastructure, you arguably can trust that are doing their job."

> Int#3 "it is probably easier to trust if there is a cooperation out there with experts in one thing, and they say 'we'll held your data securely, we'll make it for a hundred other customers, here is the insurance we offer, here are the bodies to show that.'"

Previous experiences using third-party services

The fact that some components of the service for digital preservation in the archive was already outsourced, was mentioned by Int#1 as a factor of confidence for them. The last procurement process was a new set of services with the same company using cloud for their open and public content.

> Int#1 "it is not the first time we outsource services within digital preservation, because our digital repository for instance is provided by a commercial company."

Having experience in the institution with outsourcing other services including those that store sensitive data, such as in the case of the emails was stated by Int#3 as a factor for trust third-party services. Moreover, as academic institution they are responsible for the security of the data of a variety of stakeholders, and therefore counting with their agreement and making them trust the services is highly relevant as well. Therefore, he explained that the outsourcing of the emails started five years ago with the students' accounts and after a period of revision and identification of requirements they were going to outsource the staff and academics emails in a second project. As the first one was successful, stakeholders did not show concerns about it. In this sense, Int#9 commented on the importance of keeping transparency with the community in the decisions made.

> Int#3 "because we have experience now with external services, we are pretty confident. (...) We are convinced certainly that the right company can provide our needs."

Int#9 commented that their institution is using more cloud services to outsource other activities, after the successful experiences with preservation. And Int#5 that other departments of her institution were already using third-party services before their unit started using cloud services for digital preservation.

> Int#9 "Internally we started using more cloud providers for other areas of the organization in addition to preservation."

Risk management and levels of risk acceptance

The levels of acceptance of risk that an organization or its managers are able to accept might also condition the levels of trust in other organizations.

Identifying the risks that using third parties involves for the activity and therefore keeping them under control would help to trust the third-party providing the service as mentioned by Int#14 and Int#9.

> Int#9 "There is a certain level of trust in these companies and so it is assessing your risk targets and also what you can do to prevent risks, so preparation, 'trust but verify'(...) trusting someone but also doing your best to verify that their trust is well deserved."

Nevertheless, despite controls can be set up, Int#3 stated that "you check on them you have to make sure they are worthy to trust, but at the end of the day, you can't know how everything works, so you have to rely on them."

Another aspect commented was the idea of balancing risks of trusting in a third-party and the staff and procedures of the own organization.

> Int#9 “using someone you don't know that you can't see and is not there locally has these issues. But even having someone locally, even someone of my own company there is also going to be trust issues.”

Longevity of the organization

Int#13 considered the length of time an institution has been operating as a critical factor for trust in long term preservation activities. He referred at the frameworks to assess trustworthiness of repositories because they mostly look at processes, but not at characteristics such as this one.

> Int#13 “what they don't look at is at the longevity of the organization and in practice that is probably the most critical factor when you are looking for a preservation store.”

Length of time with experience in the field was stated by the service provider Int#7 as a factor considered by the customers to trust their services.

Contracts and agreements

Formal agreements where the definitions and conditions of the service are detailed and agreed are the instruments to back up inter-organizational relationships. The third-parties or partners assume certain compromises through them that allow verification.

> Int#3 “confidentiality agreements, the data will be stored in Europe, not in the US (…)”

Before entering into conversations or contracts with third-parties, having access to models of agreements (e.g. SLA) has been pointed out as something that customers consider relevant.

> Int#2 “one of the very first things people always ask is to see an example or to see a service agreement and that's one of the reasons that we made a sample of that available on our website.”

Guarantees are also build in the contract, and on this regard, Int#3 mentioned that having those guarantees increase their level of trust compared to developing the service in-house. And Int#14 also comment that the contract should build upon some assurances to maintain trust.

> Int#14 “you have to make sure that you build in penalties clues in the contract and you need to build the insurance in the contract level.”

Going through the process of procurement and the actual conditions in the contracts was stated by Int#1 giving his institution the confidence on the service they were contracting.

> Int#1 “you have to be confident in the fact we've done as much as possible to actually mitigate the risks involved in using cloud storage providers through the g-cloud framework and the contracts that we've actually signed.”

Recognition of inter-organizational relationships as temporary

Int#12 argued that the relations established with service providers have to be assumed as temporary. He considered that in the case of long-term preservation, as any organization can guarantee that they will be able to keep their compromise without time limits. Having this approach, he stated that he feels comfortable trusting other organizations. Assuming that conditions might change or that the organization might cease to exist makes his institution be aware and mitigate the potential risks.

> Int#12 "I recognize that none of the organizations I work with will be able to guarantee the safety of my digital data forever. (…) So I think that oddly enough, understanding that is temporary makes it easier to trust the service in the near term or long term (…)."

Int#3 also mentioned that the option chosen might vary in the near future, since they might change their requirements or other opportunities that fit better their purpose might arise. Similar arguments were done by other interviewees such as Int#1, Int#9 or Int#10.

Certification, audit and accreditation

Int#13 mentioned that at their institution they use DRAMBORA self-assessments as a way to support the validation of some of their processes, and he commented that "if people are willing to accept those process validation and they are well defined they might work with us."

Being able to use accredited service providers through a procurement framework such as the already mentioned G-Cloud was considered an aid to establish an additional layer of trust on the commercial providers (Int#1, Int#14, Int#5).

Using TRAC as a framework for certification was also mentioned for a couple of organizations to be recognized by the stakeholders as a sign of trust (Int#9, Int#3). Int#14 mentioned that they might use the TRAC framework to assess against them the offer by a third-party and see if they fulfil their criteria.

Nevertheless, Int#3 commented that despite the certifications give some guidance about the service provider and how their performance is, they were not the most important requirement they have to trust third-parties.

> Int#3 "All the certifications from external bodies, I think they are good, but only a small component of the decision making."

Alignment on goals and mission

Working more comfortably with academic and research institutions compared with doing it with commercial organizations was also mentioned by a couple of interviewees. Int#9, commented that working with organizations with a long-term commitment with cultural memory and therefore explicitly aligned with their mission was considered positive. Same perception was stated by Int#2 when talking about their customers:

> Int#2 "they often find that they're much more comfortable working with us because we are an educational institution and not a commercial company."

Int#10 mentioned that their trust in the cooperative for digital preservation was built upon a shared vision. Moreover, the participation as equals in the decision-making is another factor for trust once involved in the partnership.

Taking own responsibility in the partnerships to follow common policies is one of the factors mentioned for trust in cooperative arrangements (Int#10, Int#16).

Cooperation and personal relations

Having ongoing cooperation in joint projects with the third-party service provider as detailed by Int#6 can be another source of trust. In their case, some projects were carried out to create the infrastructure at the organizational and technological level to support the collaboration. In this case due to the levels of trust they reported that some agreements were not based on contractual relations.

Int#2 also commented that the partnerships that they have with other organizations are not based on contractual relationships, but on informal arrangements as a result of years of collaboration in projects and sharing the same vision. He described it as a very successful model.

> Int#2 "A lot of the trust that we have is built up over years of working with people, knowing people or having shared goals and objectives." (…) "But really the trust is trust in the old-fashion sense of 'I trust you because I know you or because I work with you', rather than 'I trust you because we signed a piece of paper'."

Communication and personal contact with the IT staff of their cloud service providers was regarded as a factor to build trust for Int#15.

Individuals' trust preferences

Managers or staff members may have their own preferences and trust different options based on their own experience, and those might be also translated into the organizations' choices.

For instance, Int#9 stated that starting using cloud technologies, when they were still not very spread out was because their "director and staff were pretty open from the very beginning of using the cloud services."

Int#14 exemplified the situation with the case of the cloud provider they were thinking on contracting the services. The case she used was Amazon and she described that the years of relationship as a personal customer, the reputation and brand were the aspects that made her trust the company also to provide the cloud service for the institution.

> Int#14 "you can do that as much as you like but in the end it is a gut reaction, a feeling on whether you can trust them or not."

Table 11 Summary of trust mechanism in inter-organizational relationships

Reputation		
Large customer base	Level of expertise	Levels of infrastructure development
Type of customers	Geographical spread	Size of the company
Trusted institutions as customers	Viability of the company	Years of experience

Individual's trust preferences	Previous experiences with 3rd parties	Risk management
Managers or staff members choices	Service already outsourced	Levels of acceptance of risk
Customer at the personal level	Other services outsourced in the institution	Risk assessment
"Gut reaction"	Transparency with stakeholders	Balance with local risk

Longevity of the organization	Contracts and agreements	Temporary relationships
Institutions with long-term commitment on digital preservation	Access to models of agreements, Procurement process	Assuming that long-term compromise is not possible for third-parties
Years of experience of third-party	Guarantees build in the contracts	Ongoing assessment

Certification, audit, accreditation	Alignment on goals and mission	Cooperation and personal relations
Self-assessments	Cultural heritage / academia institutions	Joint projects
Accredited providers	Shared vision with partners	Informal agreements
Certification	Common policies in partnership	Communication and personal contact

4.7. Collaborative trends in the scope of digital preservation

The interviewees commented current developments and their perceptions about how the collaborative trends that are emerging will shape the digital preservation landscape. They commented on different initiatives that their institutions are part of or considering to joining in the near future.

Development of services and infrastructures for preservation

Initiatives for the development of infrastructures for digital preservation are increasingly becoming the most suitable option for some institutions. Int#13 commented that in his institution they were looking for partners in other institutions in their area or within the UK in general to build collectively an infrastructure that could meet their requirements of preserving and giving speed access to research data.

> Int#13 "a number of the EU initiatives haven't really make the same progress that we would like to have so another thing we are investigating (…) is a link with other institutions in the area or within the UK where we can build something that actually meets our requirements."

The Digital Preservation Network was mentioned by Int#13 and Int#2. This US initiative will be officially launched in 2015 and Int#13 showed interest in the technological characteristics of the network, such as the use of Internet2, to the point of start doing enquiries to relevant stakeholders and institutions on whether replicating that type of infrastructure could be possible in the UK or in a European scale. A similar initiative could meet their requirements in relationship with the storage and delivery of research data.

Nevertheless, what Int#13 considered a factor of major relevance was that the institutions involved in the network are universities and national libraries, which on his opinion are better suited for long term preservation than commercial organizations, because of their long history.

> Int#13 "The big issue there is that is generally populated by institutions, by Universities and National Libraries which have history of being around for a long period of time. And that makes them better candidates for holding preservation archives than commercial organizations."

Int#2 commented that they are part of the network and described it as a big change in terms of being able to offer better quality on the service, reliability and trustworthiness, besides more guarantees to the users and lower the prices, making use of the economies of scale. But internally, for their organization, partnering with other institutions was described a way of pooling experience and evolving on the services for digital preservation.

> Inr#2 "And is helping us to understand what kind of new services we want to offer and because it is a whole bunch of really smart people, and so they gives us more people to talk to and more people to kind of work through all the different problems that we face."

Int#12 introduced an emerging regional storage cloud service which is another big initiative that is taking form in Canada through a partnership of universities. He commented that this community cloud would be considered for his institution as primarily method of cloud storage and preservation due to its relevance for long term sustainability.

> Int#12 "In the meetings that I have today, part of the discussion will define that regional cloud storage service and who provide what services and how we fund it and all that kind of stuff."

In the case of the institution of Int#6 the development of the infrastructure for digital preservation was set-up through collaboration with an e-Infrastructure, but *ad hoc* and for their institution exclusively. Int#6 stated that the collaborative strategy is the way to follow in digital preservation and that this initiative is an example of their commitment.

> Int#6 "In our digitisation and preservation activities it soon became clear, that these projects only make sense in the context of cooperation and cannot be conducted on our own."

Development of standards and open-source projects

Another aspect of digital preservation where collaboration seems to be the norm is on the development of standards and open-source projects.

Int#13 explained that regarding the development of standards, software or techniques his institution choice was to collaborate, and not to develop all them by themselves. In particular, he mentioned that they are actively collaborating with initiatives in the US such as the development of Fedora Commons or Duraspace. Relating to this Int#12 remarked the commitment of large institutions supporting Fedora.

> Int#13 "standardizing software, processes, (…) digital object (…) preserved and copied to move to other places, standardization will help to make that happen easily.

The open source project related to digital preservation described by Int#12 has formal structures such a foundation to store the code base and as non-commercial home base for the project and commercial suppliers providing support services. Nevertheless, he remarked that the last release was completely managed by the external community and it also has a membership that costs the model.

> Int#12 "so there was a volunteer release manager and component managers and testers and it was the first time that was announced and released pretty much on time, and it was a completely volunteer effort."

Collaborative initiatives for the preservation of journal articles

Three interviewees from the academic sector made comments related to different initiatives for preserving content in a centralized way. Int#13 mentioned that his institution is participant of LOCKSS, CLOCKSS and Portico, for the purposes of preserving journal articles. These three initiatives have similar target markets and he commented that they are not align to each other and might be some tension between the three projects. Int#10 also commented that despite there are initiatives like, for instance, Portico they did not considered joining because the cost involved, especially in times of economic constrains, was not balanced with the need of their institution to preserve published journal articles.

Int#16 commented that his institution had joined Hathi Trust to preserve and give access to a specific collection of Canadian books.

> Int#16 "So they preserve ebooks and we have Canadian books collection and we are in the process of preserving that content in Hathi Trust."

Exchange of experiences

Exchange of experiences and knowledge with other institutions working in the field of digital preservation has been pointed out as vital or critical for the activity. Int#1 commented that his institution keeps a close relationship with other institutions in their geographical area using the same technologies as they do.

> Int#1 "I think that is really key with any kind of digital preservation activity, collaboration and knowledge exchange and all that kind of stuff is vital."

PASIG conferences and the membership of the Digital Preservation Coalition (DPC) or the Digital Curation Centre (DCC) were mentioned by Int#13. He commented that professionals in those types of forums such as PASIG conferences were interested in cooperate and not competing by any means, including the service providers involved. He described those forums as not being a place to commercialize services or products, but to learn and evolve in the field.

Another interesting point he made about these conferences was that they were increasingly populated by other type of organizations not related with the academic or cultural heritage sectors, such as commercial companies seeking for solutions for the preservation of their digital assets.

The type of organizations involved in the conferences or enrolled as members in DPC or DCC was also starting to be more heterogeneous, in Int#13's opinion. For instance, commercial organizations such as those from the pharmaceutical, banking or engineering sectors are increasingly populating those forums. The fact that their requirements are the similar for long term preservation of materials and long term accessibility of the materials, are starting to create synergies between the different sectors. Moreover, Int#13 commented that collaborations between the memory institutions and those other sectors might start arising as well, due to the recognition of the expertise in the field of memory institutions and also as neutral partners.

> Int#13 "they are starting to see us as potential sources of expertise in that area, but also as a potential source of collaborators. If you are a pharmaceutical company, it is easier to collaborate with a library or a museum than it is to collaborate with another pharmaceutical company, it is a competitor."

Int#14 and Int#10 commented that they have been organizers of activities such as conferences in the scope of their cooperative activities. Organizing outreach events for the community with their partners and stakeholders was also considered very relevant to strengthen the relationship within their collaborative activities.

Int#2 commented that they maintain a close relationship with other similar projects of distributed digital preservation, such as MetaArchive, people from institutions part of the LOCKSS Networks. There seems to be also a connection in the fact that those services were developed in the sphere of the projects funded by the Library of Congress. Through informal collaboration, those groups develop ideas and common strategies, informing or influencing high-level policies as well.

> Int#2 "Again, maybe it is informal like we don't have projects that we are doing all the time but we are always meeting and talking about these things."

Regional partnerships, consortia

Some of the activities under study follow these schemas for collaboration in terms of being organized as a consortium or partnerships at a regional level. Some of the interviewees commented their ideas about the prospect spread of these models as long as the pressure for preservation increases.

Int#10 stated that centralization of the operations may be a way to add value to the activity. Preservation was described as a fragmented activity, difficult to centralize into a National Library, for instance. Therefore, collective organizations may be the best suited to have a role to provide preservation services in his opinion. Single university investing in digital preservation might find it difficult if the activity does not add value to them and only costs. Int#10 also commented that preservation makes sense being centralized by leveraging economies of scale.

Int#13 also commented that a model emerging would be composed by various institutions working together, possibly under the leadership of National libraries or EU initiatives which might have the role of forming the groups. Nevertheless, the groupings envisioned would operate with a small number of organizations with more direct control. He mentioned that they looking for potential partners and consortia that they could join to either create new or use already existing joint facilities.

> Int#13 "Certainly we are looking at a number of certain consortia that could provide different facilities within the UK and possible further."

Similarly, Int#3 also mentioned that in the framework of his university activities for digital preservation, they were also negotiating with other institutions in the region a potential collaboration on setting up joint infrastructure and that negotiation at the national level were also under discussion. In this context, public funding was mentioned as critical to kick off those types of collaborative projects. More discussion related to public funding will be presented under the section Funded projects and beyond.

> Int#3 "we've certainly spoken with some other neighbouring institutions and there is the possibility of doing a regional collaboration. (…) There are also possibly national negotiations happening."

Moreover, Int#3 stated that regional cooperation is already part of the dynamics of the stakeholders involved in fulfilling the needs of the activities of research data preservation. He remarked the high rates of cooperation between their academic library with other regional academic libraries and the research offices across the different institutions in the region. As possible outcomes of the regional cooperation he pointed out the possibility of either build data centres in each institution and enable collaboration, but also potential cost reductions or service improvements for a potential regional consortium using a third-party service.

> Int#3 "So I think you have to keep it flexible and at certain point if there is a national, regional or some institutions that we can align among ourselves."

The Int#14's consortium is an example of memory institutions working together on digital preservation on these types of collaborative initiatives. Developed with funding and under the umbrella of the regional council for libraries and archives was formed to pool together expertise and infrastructure.

Funded projects and beyond

Public funding was pointed out as one of the major sources of critical funding for collaborative initiatives on digital preservation or to support the launch of joint infrastructures. Some of the institutions that the interviewees mention were the European Union, the Library of Congress, and other national or regional institutions.

Int#13 mentioned some initiatives in Europe around digital preservation in the long term access to data such as SCAPE or Open Planets that they are following up. Int#13 added that as they are looking into similar areas, they keep track of their advances, reusing what is useful for them and contributing with their expertise and information when it is relevant. One of the shortcomings he pointed out about this type of projects is that, despite they have receive funding from the EU in the phase of development, ongoing funding for the transition to the deployment of services is not available.

> Int#13 "that is where the transition for all these projects become quite difficult; how do you take a big project and then becomes a national or international service?"

On the other hand, in the case of various institutions funding an ongoing digital preservation platform, other problems might arise due to economic differences. There are inequalities on the resources of the institutions, and even if an institution might have a particular need, it might be not possible to afford to contribute very much, as Int#3 argued. Moreover, he also stated that constructing a business model for a service like the ones required for digital preservation are not so easy to develop, unless it they are funded centrally by organizations like the EU.

The Library of Congress is a very active source of funding through is programs for digital preservation. Int#2 whose organization developed the service of digital preservation through initial funding of the Library of Congress commented that they worked for a long time with them. Nowadays, that they have developed their own business model and don't receive funding from that institution in terms of paid services or shared development, they still maintain an informal relationship and work together in the development of recommendations or plans.

> Int#2 "it is a professional working relationship but we don't have any kind of official capacity there."

Int#11 argue that being partners on European funded projects help them to "stand by the state of the art of research" in the development of their services. He mentioned that they were partnering in outstanding projects in the field such as SCAPE or 4C, for instance, with European National Libraries or Archives.

At the regional level Int#14 explained that they received support for their collaborative activities from a funding body responsible for Archives, Libraries and Museums in their region. The body support the sector through the distribution of advice and assistance, but also critically funding.

Int#3 also commented that they were requesting financial support from JISC for the development of a project with other institutions in the region. As a funding body in the UK, JISC looks at the development of national services such as national repositories, and provides oftentimes initial funding for the projects, expecting institutions to fund beyond that.

> Int#3 "At the moment we are looking at, as a region, to try to get some money from a funding body to investigate the idea of institutions working together." (…) "But in a national body negotiate take an awful long time to do things. So we are not probably expecting to hear anything from them in a year or so."

Chapter 5. Conclusions

5.1. Introduction

This chapter presents the most relevant findings based on the data analysis performed on the previous chapter, and relating them with the research questions of the study.

5.2. Research questions

1. What are the benefits perceived using distributed digital preservation?

The benefits perceived by memory institutions outsourcing their services were mostly related with the costs involved in adopting a solution for digital preservation. This fact evidences the idea that one of the main issues driving the need of engaging in distributed options seems to be the economical aspect. Costs can be high and there are difficulties in making the decision of which investments have to be done to deploy the most suitable infrastructure for digital preservation.

Cost-effectiveness ranked as one of the most positive benefits perceived by the interviewees overall, matching therefore with the approaches stated by Lavoie & Dempsey (2004) in the case of using third-party services or Dhar (2012), in relation to cloud services in particular; but also with the approaches expressed by Lindlar et al. (2013) or Trehub & Halbert (2012), linked to the benefits of collaboration.

Another general benefit of distributed digital preservation stated was the mitigation of the risks of keeping the complete infrastructure for digital preservation locally. Therefore, to avoid management and disaster threats or a set of different vulnerabilities specially related with data loss (Barateiro et al., 2010), the institutions chose either partnering or the use of third party services depending on the identified needs.

The institutions using services provided by third-parties made a comparison between the costs involved in using an in-house solution and the use of outsourced options, and infrastructure and labour were pointed out as the most important factors to assess in relation to cost.

The major benefits commented in relation with the different types of services outsourced were the use of cloud as a utility, avoiding the need of the purchase or maintenance of infrastructure in-house; expected decreases on the costs (e.g. cloud storage, use of economies of scale); easiness on the identification of the costs outsourcing the service compared to keeping it in-house; need of staff provisioning or expertise in the organization lowered by outsourcing; lower barrier to entry and shorter time to set up.

Overall, the benefits of outsourcing digital preservation functions stated by the interviewees confirm what Dečman (2007) states in relationship with outsourcing to a service provider, in particular, ideas such as eliminating up-front capital expenditure, a predictable cost structure, expertise and a quick set-up of the service. Nevertheless, aspects such as costs-effectiveness of using cloud over the long-term for digital preservation have been discussed by Rosenthal & Vargas (2013), who do not consider cloud the most suitable option.

In the case of partnerships between memory institutions, the major benefits pointed out were the lower costs of the solution adopted and the use of their own resources in terms of infrastructure and expertise, which agrees with the stated by Jordan et al. (2008). Moreover, partnerships have been described as a more sustainable option, in the light of the needs of long-term preservation. Achievement of sustainability through the collaboration of memory institutions has been described as a successful strategy by Downs & Chen (2010) and Giaretta (2008).

From the above discussion, it can be concluded that the options considered by the study in the field of distributed digital preservation are perceived by practitioners as cost-effective options and a suitable to minimize the risks of keeping the activity exclusively in-house. Overall, the benefits perceived by those institutions outsourcing to third parties, especially to cloud, reside in the capacity of those services to provide a quick start, without the need of investing either in infrastructure or development of expertise. Moreover, future cost reductions were also expected due to the economies of scale. On the other hand, partnerships between memory institutions, despite implying a greater commitment from the organizations, are perceived as a more suitable option for digital preservation over the long-term.

2. **What are the major organizational and regulatory risks for memory institutions with long-term preservation responsibilities using distributed digital preservation models?**

A greater variety of risks related to the financial management and sustainability aspects were described in the case of outsourced services, especially in relation with the use of cloud. The risks perceived were the inadequacy of the ongoing payments required by the cloud services to the budget structures of the institutions; possible additional costs due to the lock-in practices in the case of data portability needs.

Pricing models in the case of outsourcing to cloud and that they may imply additional charges due to the increase of data stored, the size of the files or the actions performed was pointed out as a risk. In opposition, other types of third-party services may charge greater up-front costs but not increase the expenses for those described actions. In this context, using cloud services, the inability to estimate the growth of the collections could also be a vulnerability.

Due to the lack of security on the long-term funding for the activities of digital preservation, facing the expenses related with third-party services might entail difficulties. In the same line, the ongoing payments to the partnerships and the expenses incurred locally to perform actions in the scope of the partnership might constitute a threat for the organizations.

In relationship with the organizational structure and staffing, a common risk in relationship with the use of third parties was the mismatches and lack of coordination that can arise between the different services. Another common risk pointed out was related with the need of adaptation to new workflows and procedures, especially to avoid misunderstandings in the processes conducted in the third-party premises. Furthermore, new capabilities and knowledge were also mentioned as a need that has to be addressed in some of the cases of outsourcing to third parties.

In the case of partnerships, the lack of skilled staff or expertise or the lack of capacity and resources to contribute to the partnership would be risks that would not allow the viability of the collaboration.

In terms of governance and organizational viability, common risks to any of the options for distributed digital preservation may be related with changes on the strategy of the organizations involved. In the case of service providers, the unilateral changes on conditions or the overall strategy were stated as a threat for their customers; in the case of the partnerships, diversity of mission, goals and strategy can be a challenge.

The service providers ceasing business has been pointed out as a risk in any kind of relationship established with third party services. Whereas in the case of the partnerships, requirements such as minimum number of members to effectively run the collaboration, can be a threat when members leave the network.

In the case of cloud, loss of reputation was described as a potential risk, but only in worst-case scenarios. A more common source of risk could be the loss of governance due to the vendor lock-in practices. Loss of governance could be also argued in the case of partnerships, in cases such as the leadership being taken *de facto* by the larger organization in the network.

In the context of the procedural accountability and policy framework some aspects reported could be a threat for the institutions. Using cloud, organizations may face a lack capacity to meet their business objectives, for instance, not being able to ensure that the requirements of preservation of the digital objects have been effectively met. This may be exacerbated due to a lack of adequate reporting allowing the institutions to check whether the levels of service are being actually addressed or not by the service provider. In this sense, lack of transparency in some of the procedures of the cloud providers has been reported.

In the case of partnerships or centralized services for a network provided by memory institutions, the diversity of the collections policies and procedures related to them in the different institutions were reported as risks.

The aspects related with contracts, licenses and liabilities, are the ones showing a major amount of common risks in any of the options used for distributed digital preservation throughout this study. Risk of non-compliance with the regulations on copyright or data protection and the lack of agreements with data providers and clarification of rights of the assets, may be possible in any of the choices.

The non-compliance with the SLA, and loss of service levels or availability could be possible in any type of third-party services. Parallelism can be found in the case of partnership members not fulfilling their duties. Liabilities seem to be typically limited either in the side of the service providers or the other members of the partnership, as the responsibility for the clarification of the legal status of the assets deposited has been reported to be posed in the organization owning the resource. The jurisdiction on which the digital assets are located might be as well a risk, especially to those stored in the cloud, due to its diversity; but also was reported in the case of international collaboration with partners in other jurisdictional areas. Nevertheless, the difference is that, in the case of cloud, the organization may lose control over the location, whereas the partnerships have clear defined boundaries.

Subcontracting third party services has been understood as a potential risk in the cases of third-parties. It could be a threat in cases where there is lack of transparency on the agreements.

Informal agreements with close collaborators were reported as a potential risk, in particular in the cases of centralized services provided by memory institutions and in the cases of partnerships.

3. What are the controls or mitigation strategies for the identified risks?

Relationships with other organizations are typically regulated by the contracts and agreements (Jordan et al., 2008). Therefore the conditions and guarantees for the service, that would serve as mitigation strategies for the risks identified by the organizations, are described in those instruments. Common aspects mentioned in the contracts that would regulate any of the relations in the scope of distributed digital preservation would be the agreements in terms of data protection; the limitations on legal liabilities; the exit-strategy; the ownership of the data; levels of durability, integrity of data or security; the financial penalties or insurance; the third parties subcontracted by the service provider. In the case of partnerships, specific agreements are those related with the terms or conditions of becoming part of a network; and independent contracts of the individual partners with service providers. Ability to negotiate the contracts and legal advice have been reported as useful, due to the difficulties in cases such as cloud services (although not exclusively) to introduce own requirements.

To ensure legal compliance in copyright aspects, organizations were reported to be the ones responsible for the establishment of clear policies and clarify the rights, when necessary. And the right licenses should be attached to the data, in case it is made available through distributed services. Additionally, in the case of partnerships an especial emphasis has been made on the compromise of the staff to comply with the regulations. Data subject of protection does not usually leave the premises of the memory institutions, and it is kept in most cases in a datacentre owned by the institution. Otherwise, the levels of security have to be adequately guaranteed by the service provider, especially in the case of cloud. Likewise, when institutions have the mandate to comply with other regulations, and the service providers must provide guarantees of compliance. Control over the jurisdiction needs to be ensured to ensure that data is not subject of different regulations or levels of protection, especially in the case of cloud.

In the economic aspect, using cost-analysis, especially looking at aspects such as cost-benefit and cost-effectiveness were pointed out as methods needed to support decision making and the appraisal of the options to avoid financial risks. To obtain funding support for the projects the elaboration of compelling business cases with a clear value proposition were recommended.

Raising awareness among stakeholders and funders either as a way to ensure guarantees on the levels of funding for the activity or to make easier the integration of a cost model for the self-sustainability of the activity, without raising stakeholders concerns. High-level mandates, major implication on the processes and transparency have been pointed out as critical.

The need to seek for alternative business models or funding strategies were mentioned as a way to ensure sustainability of the activity. Mix-funding models, using a cost model to charge for services; introducing additional charges for members of the partnerships with special needs; or offering services to the general public are some of the possibilities. Relying on public funding for special projects and maintaining a contingency fund for emergencies were strategies mentioned in this context.

Gaining awareness on the needs, risks and having ability to anticipate future needs are relevant aspects for an organization to be able to avoid or mitigate risks. Conducting a readiness assessment survey, benchmarking, conducting risk assessment or audit and self-assessment were stated as relevant methods both to evaluate the own organization and, when appropriate, the service providers or partnerships.

In relation with the staff, the new relationships established may require the development of new skills through the adequate training or provisioning of new staff members. In collaborative approaches, organisms such as those running the consortia seem to be useful to centralize administrative and technical tasks. With a more decentralized approach, the creation of a community of practice to distribute the new tasks and enhance skills and expertise was also explained.

The procurement process has been considered especially relevant for institutions entering in contractual relationships with third party services, to ensure transparency and accountability. Furthermore, the processes and checking to be performed to select the right service provider have to be carefully chosen. Getting assurances of service provider's viability, involving stakeholders of the own organization in the panels and site visits to other customers were pointed out. Additionally, frameworks such as the G-Cloud in the UK were relevant in the sense that accredited service providers are made available.

Pilot phase and ongoing assessment of the chosen option when entering into agreements with a service provider, were also stated.

In the case of partnerships, the alignment of mission and the establishment of common objectives were considered highly relevant. In the other hand, the establishment of common structures and policies to be able to work in a coordinated manner are also essential.

Communication, support and reporting / monitoring strategies and protocols should be in place to allow the necessary coordination between the actors involved, and also the checking of the right functioning of the systems. In the case of commercial providers, namely cloud, the reporting and monitoring have been reported as challenging in several occasions.

Redundancy of data, whether through organizational redundancy, geographical or systems-based, is broadly accepted as one of the major mitigation strategies to avoid the loss of data. In relationship with organizational aspects in the scope of distributed digital preservation, the risk of losing data because of the service provider goes out of business would be avoided. Diversifying the geographical areas would avoid not only the risk of, for instance, disasters, but also potential economic effects in a region. Organizational diversity through the use of commercial and non-commercial organizations was also recommended. The use of in-house data centre or private clouds are also possible.

4. **Which are the mechanisms of trust in inter-organizational relationships in the scope of distributed digital preservation?**

Some of the mechanisms to build trust in the inter-organizational relationships arising through the distribution of responsibilities of digital preservation among different actors, could be considered as trust built over time (Day, 2008). The established reputation and customer base was a relevant factor for many of the participants on this study, considering as relevant that the service provider had a large customer base; certain typologies of customers, such as institutions considered trusted by them (e.g. national libraries, archives); the level of expertise, geographical spread, the viability of the company, levels of infrastructure development, the size of the company or the years of experience. Moreover, the longevity of the organization was also pointed out as a critical factor for trust.

Similarly, having previous experiences using third-party services was also considered relevant, especially in case of the service being outsourced had been successfully outsourced previously; other services in the institution outsourced; and transparency with stakeholders granted.

Walters & McDonald (2008) considered risk management as a mechanism to establish trust, and so did some of the participants of this study. Some of the conditions may be the levels of acceptance of risk, the assessment of the risks and the balance with the risk of keeping the set up for digital preservation in the own premises. In the light of potential risks, it was also suggested that inter-organizational relationships should be considered temporary in the case of long-term digital preservation; those relationships will vary over the long-term and mitigation strategies should be in place.

Contracts, evidence-based practice (Walters & McDonald, 2008) and adherence to standards (Day, 2008) are also factors relevant to support trust in inter-organizational relationships; among the findings of this study, the establishment of contracts and agreements and certification, audit and accreditation were considered important to trust the other parties.

Sharing values or culture (Day, 2008) are relevant for organizations to trust each other. In this sense, alignment on goals and mission and the personal relations established through cooperation have been pointed out during the interviews. Moreover, individual's trust preferences may be determinant in the process of trusting another organization.

5. Which are the major collaborative trends in the field of digital preservation?

The major collaborative trends pointed out by the participants are focused on the development of services and infrastructures, the exchange of experiences, distributed initiatives for preservation or cooperation in funded projects.

The development of services and infrastructures has been pointed out at different levels, such as regional, national or even wider geographical areas. Creation of regional cloud services, collaboration with e-Infrastructures or network of universities for the distribution of data were some of the examples pointed out. Cooperative development of standards and open-source is also one of the major collaborative activities for the development of services. Additionally, the configuration of consortia or regional partnerships to provide preservation services for a network was also considered relevant to either establish the service collectively in their own premises or to contract outsourced services as a consortium.

Exchange of experiences through professional events and informal networks was pointed out as critical for the evolution of the field. Moreover, finding potential partners from the own or other sectors was possible through these gatherings. In relation with this aspect, joint research projects are an active form of collaboration, usually receiving public funding for the early stages of the project but leaving the transition from the prototype to actual deployment of the services or infrastructures unfunded.

Lastly, the examples of distributed initiatives for digital preservation mentioned during the interviews were, on one hand initiatives for the preservation of journal articles such as LOCKSS, CLOCKSS or Portico. Nevertheless, other initiatives were mentioned such as cooperatives for digital preservation or other services such as Hathi Trust.

References

Aitken, B., McCann, P., McHugh, A., & Miller, K. (2012). Digital Curation and the Cloud Final Report. Presented at the JISC's Curation in the Cloud Workshop.

ALA. (n.d.). *Outsourcing and Privatization. Professional Tools*. Retrieved June 23, 2014, from http://www.ala.org/tools/outsourcing

Altman, M., Adams, M., Crabtree, J., Donakowski, D., Maynard, M., Pienta, A., & Young, C. (2009). Digital Preservation Through Archival Collaboration: The Data Preservation Alliance for the Social Sciences. *The American Archivist*, *72*(1), 169–182.

Anderson, D. (2013). Preserving Europe's Digital Cultural Heritage: A Legal Perspective. *New Review of Information Networking*, *18*(1), 16–39.

Anderson, M. (2008). Evolving a network of networks: the experience of partnerships in the National Digital Information Infrastructure and Preservation Program. *International Journal of Digital Curation*, *3*(1), 4–14.

Askhoj, J., Sugimoto, S., & Nagamori, M. (2011). Preserving records in the cloud. *Records Management Journal*, *21*(3), 175–187.

Atkins, D. (2003). *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. National Science Foundation.

Barateiro, J., Antunes, G., Cabral, M., Borbinha, J., & Rodrigues, R. (2008). Using a Grid for digital preservation (pp. 225–235). Presented at the 11th International Conference on Asian Digital Libraries: Universal and Ubiquitous Access to Information, Kuta, Indonesia.

Barateiro, J., Antunes, G., Freitas, F., & Borbinha, J. (2010). Designing digital preservation solutions: A risk management-based approach. *International Journal of Digital Curation*, *5*(1), 4–17.

Barbour, R. (2008). *Introducing qualitative research*. London, England: SAGE Publications.

Beagrie, N. (2006). Digital curation for science, digital libraries, and individuals. *International Journal of Digital Curation*, *1*(1), 3–16.

Beagrie, N., Charlesworth, A., & Miller, P. (2014). *Guidance on Cloud Storage and Digital Preservation: How Cloud Storage can address the needs of public archives in the UK*. The National Archives. Retrieved from http://www.nationalarchives.gov.uk/documents/archives/cloud-storage-guidance.pdf

Beagrie, N., & Jones, M. (2008). *Digital Preservation Handbook*. London, England: Digital Preservation Coalition. Retrieved from http://www.dpconline.org/advice/preservationhandbook

Becker, C., Barateiro, J., Antunes, G., Vieira, R., & Borbinha, J. (2011). On the relevance of Enterprise Architecture and IT Governance for Digital Preservation. Presented at the IFIP Tenth conference on electronic government (EGOV 2011), Delft, Netherlands.

Becker, C., Kulovits, H., Guttenbrunner, M., Strodl, S., Rauber, A., & Hofman, H. (2009). Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. *International Journal on Digital Libraries*, *10*(4), 133–157.

Besek, J. M., LeFurgy, W. G., Rasenberger, M., Weston, C. D., Muir, A., Atkinson, B., … Mossink, W. (2008). *International study on the impact of copyright law on digital preservation*. Washington, D.C.: The Library of Congress National Digital Information Infrastructure and Preservation Program, The Joint Information Systems Committee, The Open Access to Knowledge (OAK) Law Project, The SURFfoundation.

Bloor, M., & Wood, F. (2006). Sampling. In *Keywords in Qualitative Methods* (pp. 154–158). London, England: SAGE Publications.

Blue Ribbon Task Force. (2008). *Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation*. Blue Ribbon Task Force on Sustainable Digital Preservation and Access.

Blue Ribbon Task Force. (2010). *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*. Blue Ribbon Task Force on Sustainable Digital Preservation and Access.

Bradshaw, S., Millard, C., & Walden, I. (2010). Contracts for Clouds: Comparison and Analysis of the Terms and Conditions of Cloud Computing Services. *Queen Mary University of London, School of Law*.

Burda, D., & Teuteberg, F. (2013). Sustaining accessibility of information through digital preservation: A literature review. *Journal of Information Science, 39*(4), 442–458.

Caplan, P. (2008). Digital preservation: standing the test of time. Presented at the 33rd, International Association of Aquatic and Marine Science Libraries and Information Centers.

Castelfranchi, C., & Falcone, R. (2000). Does control reduce or increase trust? A complex relationship. In *Proc. of Deception, Fraud and Trust in Agent Societies workshop, Autonomous Agent* (pp. 49–60).

Chowdhury, G. (2013). Sustainability of digital information services. *Journal of Documentation*, *69*(5), 602–622.

Convery, N. (2010a). *Cloud Computing Toolkit*. Aberystwyth University and Archives and Records Association UK and Ireland. Retrieved from http://www.drnicoleavena.com/storage/articles/Cloud_Computing_Toolkit-2.pdf

Convery, N. (2010b). *Storing Information in the Cloud: Project report*. Archives & Records Association UK & Ireland.

Creswell, W. (2007). *Qualitative inquiry & research design : choosing among five approaches* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

CRL. (2012). *Certification Report on Chronopolis*. Center for Research Libraries.

Day, M. (2008). Toward distributed infrastructures for digital preservation: The roles of collaboration and trust. *International Journal of Digital Curation*, *3*(1), 15–28.

DCC, & DPE. (2007). *Digital Repository Audit Method Based on Risk Assessment*. Retrieved from http://www.repositoryaudit.eu/

Dečman, M. (2007). Long-term Digital Archiving-Outsourcing or Doing it. *Electronic Journal of e-Government*, *5*(2).

Dečman, M., & Vintar, M. (2013). A possible solution for digital preservation of e-government: A centralised repository within a cloud computing framework. *Aslib Proceedings*, *65*(4), 406–424.

Dhar, S. (2012). From outsourcing to Cloud computing: evolution of IT services. *Management Research Review, 35*(8), 664–675.

Dobratz, S., & Schoger, A. (2007). Trustworthy digital long-term repositories: The Nestor approach in the context of international developments. In *Research and advanced technology for digital libraries* (pp. 210–222). Springer.

Downs, R. R., & Chen, R. S. (2010). Self-assessment of a long-term archive for interdisciplinary scientific data as a trustworthy digital repository. *Journal of Digital Information*, *11*(1).

ENISA. (2009). *Cloud Computing. Benefits, risks and recommendations for information security*. Retrieved from http://www.bulentkutlu.com/reports/Enisa_Cloud_Computing_Security_Risk_Assessment.pdf

ERPANET. (2003). *Erpa-Tool - Risk Communication Tool*. Electronic Resource Preservation and Access Network. Retrieved from http://www.erpanet.org/guidance/docs/ERPANETRiskTool.pdf

European Commission. (2012). *Unleashing the Potential of Cloud Computing in Europe* (No. COM(2012) 529 final). Brussels. Retrieved from http://ec.europa.eu/digital-agenda/en/european-cloud-computing-strategy

Evens, T., & Hauttekeete, L. (2011). Challenges of digital preservation for cultural heritage institutions. *Journal of Librarianship and Information Science*, *43*(3), 157–165.

Gellman, R. (2009). *Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing*. World Privacy Forum.

Giaretta, D. (2008). The CASPAR approach to digital preservation. *International Journal of Digital Curation*, *2*(1), 112–121.

Gillham, B. (2009). *Case study research methods*. London ; New York: Continuum.

Halbert, M. (2009). Comparison of strategies and policies for building distributed digital preservation infrastructure: initial findings from the MetaArchive Cooperative. *International Journal of Digital Curation*, *4*(2), 43–59.

Harmsen, H., Keitel, C., Schmidt, C., Schoger, A., Schrimpf, S., Stürzlinger, M., & Wolf, S. (2013). *Explanatory notes on the nestor Seal for Trustworthy Digital Archives*. nestor – Network of Expertise in Long-Term Storage. Retrieved from http://files.d-nb.de/nestor/materialien/nestor_mat_17_eng.pdf

Hilton, J., Cramer, T., & Minor, D. (2013). The Case for Building a Digital Preservation Network. *EDUCAUSE Review*, *48*(4).

Innocenti, P., Ross, S., Maceviciute, E., Wilson, T., Ludwig, J., & Pempe, W. (2009). Assessing digital preservation frameworks: the approach of the SHAMAN project. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems* (pp. 412–416). France: ACM.

ISO. (2009). *ISO/FDIS 31000. Risk management — Principles and guidelines*.

Jordan, C., Kozbial, D., Minor, D., & McDonald, R. (2008). Encouraging Cyberinfrastructure Collaboration for Digital Preservation. Presented at the International Conference on Preservation of Digital Objects (iPRES 2008), London, England.

Jøsang, A., Fritsch, L., & Mahler, T. (2010). Privacy policy referencing. In *Trust, Privacy and Security in Digital Business* (pp. 129–140). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-15152-1_12

Kaur, K., Darby, R., Herterich, P., Schmitt, K., Schrimpf, S., Tjalsma, H., & Lambert, S. (2013). *Report on Testing of Cost Models and Further Analysis of Cost Parameters*. APARSEN.

Kaur, K., Herterich, P., Dallmeier-Tiessen, S., Schmitt, K., Schrimpf, S., Tjalsma, H., … McMeekin, S. (2013). *Report on Cost Parameters for Digital Repositories* (No. D 32 . 1). APARSEN.

Kejser, U. B., Johansen, K., Thirifays, A., Nielsen, A. B., David, W., Strodl, S., … Tjalsma, H. (2014). *Evaluation of Cost Models and Needs & Gaps Analysis* (No. D3.1). 4C - Collaboration to Clarify the Cost of Curation. Retrieved from http://www.4cproject.eu/component/docman/doc_download/34-d3-1-evaluation-of-cost-models-and-needs-gaps-analysis?Itemid=

Kyriazis, D. (2013). *Cloud Computing Service Level Agreements. Exploitation of Research Results*. Brussels: European Commission.

Lambert, S., Hein, S., Bazzanella, B., Proell, S., & Strodl, S. (2014). *Overview of Preservation Services*. APARSEN.

Lapan, S. D., Quartaroli, M. T., & Riemer (Eds.). (2012). *Qualitative Research: An Introduction to Methods and Designs*. San Francisco, CA: Jossey-Bass.

Lavoie, B., & Dempsey, L. (2004). Thirteen Ways of Looking at...Digital Preservation. *D-Lib Magazine*, *10*(7/8). Retrieved from http://www.dlib.org/dlib/july04/lavoie/07lavoie.html

Lindlar, M., Friese, Y., Müller, E., Bähr, T., & von Trosdorf, A. (2013). Benefits of geographical, organizational and collection factors in digital preservation cooperations: The experience of the Goportis consortium. Presented at the iPres2013 - 10th International Conference on Preservation of Digital Objects, Lisbon.

Lindström, J. (2011). *Areas and problems to consider within information security and digital preservation during procurement and use of cloud services*. Cloud Sweden.

Mason, J. (2012). *Qualitative researching* (2nd ed.). Los Angeles, California: SAGE.

McHugh, A. (2012). A model for digital preservation repository risk relationships. Presented at the World Library and Information Congress: 78th IFLA General Conference and Assembly, 11-17 Aug 2012, Helsinki, Finland. Retrieved from http://conference.ifla.org/past-wlic/2012/216-mchugh-en.pdf

McHugh, A., Ross, S., Innocenti, P., Ruusalepp, R., & Hofman, H. (2008). Bringing self assessment home: repository profiling and key lines of enquiry within DRAMBORA. In *Archiving Conference* (Vol. 2008, pp. 13–19). Society for Imaging Science and Technology.

Minor, D., Sutton, D., Kozbial, A., Westbrook, B., Burek, M., & Smorul, M. (2010). Chronopolis digital preservation network. *International Journal of Digital Curation*, *5*(1), 119–133.

Moore, R. (2004). Evolution of data grid concepts. In *Global Grid Forum Data Area Workshop*. Retrieved from http://www.nesc.ac.uk/events/GGF10-DA/programme/papers/06-Moore-Grid-evolution.pdf

Moore, R. (2008). Towards a theory of digital preservation. *International Journal of Digital Curation*, *3*(1), 63–75.

NAA. (2014). *Outsourcing digital data storage*. Retrieved February 10, 2014, from http://www.naa.gov.au/records-management/agency/secure-and-store/outsourcing-digital-data/index.aspx

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: SAGE Publications.

Peterson, M. (2011). *Digital Preservation in the Cloud*. *Long-term Digital Preservation Reference Model*. Retrieved June 23, 2014, from http://www.ltdprm.org/reference-model/preservation-in-the-cloud

Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013). The NDSA Levels of Digital Preservation: Explanation and Uses. In *Archiving Conference* (Vol. 2013, pp. 216–222). Society for Imaging Science and Technology.

Pickard, A. J. (2007). *Research methods in information*. London, England: Facet.

Reich, V., & Rosenthal, D. (2009). Distributed Digital Preservation: Private LOCKSS Networks as Business, Social, and Technical Frameworks. *Library Trends*, *57*(3), 461–475.

RLG, & OCLC. (2002). *Trusted Digital Repositories: Attributes and Responsibilities*. Mountain View, CA: Research Libraries Group.

RLG-NARA Task Force. (2007). *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)*. CRL, OCLC, NARA. Retrieved from http://www.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/trac.pdf

Robson, C. (2002). *Real world research : a resource for social scientists and practitioner-researchers* (2nd ed.). Oxford: Blackwell.

Rosenthal, D., Rosenthal, D. C., Miller, E. L., Adams, I. F., Storer, M. W., & Zadok, E. (2012). The economics of long-term digital storage. *Memory of the World in the Digital Age, Vancouver, BC*.

Rosenthal, D., & Vargas, D. L. (2013). Distributed Digital Preservation in the Cloud. *International Journal of Digital Curation*, *8*(1), 107–119.

Ross, S., & McHugh, A. (2006). The Role of Evidence in Establishing Trust in Repositories. *D-Lib Magazine*, *12*(7/8).

Rubin, H. J., & Rubin, I. S. (2005). *Qualitative interviewing : the art of hearing data* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Ruusalepp, R., Justrell, B., & Florio, L. (2014). *Deliverable D4.1 Trust Building Report*. Digital Cultural Heritage Roadmap for Preservation - Open Science Infrastructure for DCH in 2020.

Sanett, S. (2013). Archival Digital Preservation Programs: Staffing, Costs, and Policy. *Preservation, Digital Technology & Culture*, *42*(3), 137–149.

Schmidt, R., King, R., Steeg, F., Melms, P., Jackson, A., & Wilson, C. (2009). A framework for distributed preservation workflows.

Schultz, M., & Skinner, K. (2014). *Comparative Analysis of Distributed Digital Preservation (DDP) Systems*. Educopia Institute.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, *22*(2), 63–75.

Sinclair, P., Billenness, C., Duckworth, J., Farquhar, A., Humphreys, J., Jardine, L., … Sharpe, R. (2011). Are you ready? Assessing whether organisations are prepared for digital preservation. *The International Journal of Digital Curation*, *1*(6), 268–281.

Skinner, K., & Halbert, M. (2009). The MetaArchive Cooperative: A Collaborative Approach to Distributed Digital Preservation. *Library Trends*, *57*(3), 371–392.

Smith, M., & Moore, R. W. (2007). Digital archive policies and trusted digital repositories. *International Journal of Digital Curation*, *2*(1), 92–101.

Stewart, C. A., Simms, S., Plale, B., Link, M., Hancock, D. Y., & Fox, G. C. (2010). What is cyberinfrastructure. In *Proceedings of the 38th annual ACM SIGUCCS fall conference* (pp. 37–44). Retrieved from http://dl.acm.org/citation.cfm?id=1878347

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research : grounded theory procedures and techniques*. Newbury Park, CA: SAGE Publications.

Toussaint, M. G., & Rounds, S. (2013). *Report on Digital Preservation and Cloud Services*. Minnesota Historical Society. Retrieved from http://www.mnhs.org/preserve/records/docs_pdfs/Instrumental_MHSReportFinal_Public_v2.pdf

Trehub, A., & Halbert, M. (2012). Safety in Numbers: Distributed Digital Preservation Networks. Retrieved from http://conference.ifla.org/conference/past/ifla78/216-trehub-en.pdf

Vermaaten, S., Lavoie, B., & Caplan, P. (2012). Identifying Threats to Successful Digital Preservation: the SPOT Model for Risk Assessment. *D-Lib Magazine*, *18*(9/10).

Vincent, M., Hart, N., & Morton, K. (2011). *Cloud Computing Contracts White Paper. A Survey of Terms and Conditions*. Sydney: Truman Hoyle Lawyers.

Walters, T. O., & McDonald, R. H. (2008). Creating Trust Relationships for Distributed Digital Preservation Federations. In *Proceedings of the 5th International Conference on Preservation of Digital Objects (iPRES)*. Retrieved from http://smartech.gatech.edu/handle/1853/38979

Walters, T. O., & Skinner, K. (2010). Economics, sustainability, and the cooperative model in digital preservation. *Library Hi Tech*, *28*(2), 259–272.

Waters, D., & Garrett, J. (1996). *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information.* Commission on Preservation and Access and the Research Libraries Group Task Force on Digital Archiving. Retrieved from http://files.eric.ed.gov/fulltext/ED395602.pdf

Wittek, P., & Darányi, S. (2012). Digital Preservation in Grids and Clouds: A Middleware Approach. *Journal of Grid Computing*, *10*(1), 133–149.

Yin, R. (2003). *Case Study Research: Design and Methods* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Zierau, E., Kejser, U. B., & Kulovits, H. (2010). Evaluation of bit preservation strategies. In *Proceedings of the Seventh International Conference on Preservation of Digital Objects (iPRES2010)* (pp. 161–169). Retrieved from http://www.academia.edu/download/30839470/PubDat_191968.pdf#page=161

Zierau, E., & Schultz, M. (2013). Creating a Framework for Applying OAIS to Distributed Digital Preservation. Presented at the iPRES 2013 - 10th International Conference on Preservation of Digital Objects, Lisbon.

Appendices

Appendix 1: Interview structures

Interview structure for memory institutions using third-party services

Why did your organization decide to use this particular type of third-party service? Overall, what has been your experience outsourcing to them?

Organizational and regulatory risks:

Can you tell me about threat/risks to organizational and regulatory aspects of digital preservation that you perceive using a third-party provider? If there are no risks identified, do you foresee that they may occur?

How does it modify the previous situation? Can you provide some details or specific examples?

Which controls does your organization deal or have planned to deal with those risks?

Trusting the service provider:

How do you assess or control aspects such as trustworthiness, reliability, security or quality of your service provider?

How do you control those aspects in your organization?

In your opinion, are the service agreements and contracts enough to have guarantees and comply with the requirements of your own stakeholders?

Changes in the organization:

What happened when you engaged the third-party service, did the contractual relationship modify organization’s aspects such as strategy or policies?

Can you summarize the benefits of the relationship with the third-party service for your organization and the activity of digital preservation more in particular?

Access to relevant documents (will be used in confidence and not quoted):

Would it be possible to have access to (anonymised) documents of your organization such as: risk register, SLA, contract, policies or strategies relevant to digital preservation1, or any other relevant document related to the third party service provision?

Interview structure for service providers

Can you briefly explain the major benefits for the institutions using your service?

Organizational and regulatory risks:

Can you tell me about the risks (organizational, regulatory) for the institutions using your services?

What about your own organization, which are the organizational and regulatory risks that could affect the service provided?

Which controls does your organization use or have planned to deal with those risks, either for those that might impact on your customers or your own organization?

Trusting the service provider:

Has your organization been asked to prove the trustworthiness of its preservation service? If yes, how did you demonstrate the trustworthiness? Are there any controls in place to demonstrate trust in the service (e.g. assessment, certification…)?

Can you talk about the service agreements and contracts and how they provide guarantees for the institutions using your services?

Changes for memory institutions using third-party services:

In your experience, how does the use of your service introduce changes on policies and strategy of the memory institutions?

What happens before and after the signature of the contract, can you tell me about the relationship and perceptions of institutions using your service?

Can you summarize the major improvements that you think your services introduce to the final outcome of the preservation service?

Access to relevant documents (will be used in confidence and not quoted):

Would it be possible to have access to anonymised documents or templates for: risk register, models of contracts or SLA or any others considered relevant?

Interview structure for memory institutions in partnerships

Can you briefly explain which the major benefits for your institution to engage in the partnership?

Organizational and regulatory risks:

Can you tell me about potential organizational and regulatory threats/risks to digital preservation that you have identified partnering with other institutions (e.g. related to sustainability, staff, institutional and legal framework, accomplishment of objectives, etc.)? Can you give some details or examples on those more relevant for your organization?

In case you provide services to third-parties not involved in the partnership, how do the risks affect to the service provided to them?

Which control mechanisms is your organization or the partnership using or has planned to use to deal with those risks?

Trusting the partnership:

How does your organization assess and prove aspects such as trustworthiness, reliability, security or quality of the service? Are there any controls in place to ensure trust within the partnership?

Can you talk about the service agreements and contracts and how they provide guarantees for the institutions using the shared services?

Changes for the institutions in the partnership:

What happened when you engaged the partnership, how did this new architecture introduce changes in your organization's policies and strategy?

What happened before and after the formalization of the agreements, can you tell me about the relationship with and perceptions about partnering institutions and any other stakeholders of relevance?

Can you summarize the major improvements to the outcome of digital preservation for the institutions collaborating or using the service?

Access to relevant documents (will be used in confidence and not quoted):

Would it be possible to have access to anonymised documents or templates for: risk register, models of contracts or SLA, policies and strategies related to digital preservation or any others considered relevant?

Appendix 2: Informed consent form

INFORMED CONSENT

This M.A. study aim is to identify risks and other implications at the organizational and regulatory level for memory institutions using distributed systems for digital preservation, including third-party services, such as those in the cloud.

The participants for the data collection exercise were selected using a set of pre-defined criteria. The characteristics taken in consideration:

- Memory institutions that are using third-party services for digital preservation.
- Providers of services for digital preservation.
- Members of partnerships for digital preservation.

Semi-structured interviews with same or similar questions will be asked in each group that share a similar role. The interviews will be completed using Skype or email. The interviews will be recorded for the purpose of note-taking only.

The data collected will be treated as anonymous to ensure confidentiality, unless the opposite is expressed by email or at the beginning of the interview by the participants, in terms of including personal or organizations' names. Including those names will only be possible if agreement is reached with all participants.

The researcher guarantees that the data collected will be used for academic purposes only and in the context of the master's thesis of the Digital Library Learning (DILL) program, and insights gathered by you and other participants will be used in writing a qualitative research report. Though direct quotes from you may be used in the paper, your name and other identifying information will be kept anonymous (see stated above).

Participants have the right to withdraw from the study at any time. In the event you choose to withdraw from the study all information you provide will be omitted from the final report.

The interviews will take ca. 40 min. In case an interview using Skype was not schedule, responses should be sent to evamontenegro@gmail.com. Period for data collection: May 12th - 23rd, 2014.

Should you have any questions, please do not hesitate to contact me at evamontenegro@gmail.com / Skype: evamontenegro

By writing "I AGREE" bellow in this consent form I __________________ confirm my agreement to the above specified terms.

(Participant´s name)

______________________ ____________________

(Expression of agreement) (Date)

Appendix 3: Presentation letter

Good morning,

My name is Eva Montenegro and I am a master student on a program on digital libraries (http://dill.hioa.no/). I am writing you seeking for some advice and perhaps collaboration for a small-scale study that I am conducting for my master thesis. When it comes to advice, any comments on the ideas that I express below will be welcome. The contribution I would be asking will be in the form of a brief interview or filling in a questionnaire in your role as an expert in the field I am addressing, depending on your availability and willingness to participate. In case of acceptance, we can discuss further details later on.

The aim of the study is to investigate risks and implications of using distributed or collaborative IT systems or organizational infrastructures for digital preservation in public institutions. Within this framework, I am interested in exploring the outsourcing of services, for example using cloud technologies, and to draw a comparison with other types of distributed options currently in use in digital preservation.

In this context, I would like to shed light on a few these questions in my study:

- What are the major organizational, policy and legal risks for institutions with long-term preservation mandates and to their digital assets in distributed digital preservation systems?
- What are the controls or mitigation strategies for these identified risks?
- What are the implications of those risks to trust in digital preservation?
- How does outsourcing IT infrastructure or the preservation tasks introduce changes to the final outcome of the service?
- What are the changes that risks associated with outsourcing introduce to the organization's information policies and strategy?

I would appreciate any comments that could help me to better frame and focus the study, and also to know if you or other possible participants from your institution would agree to participate as experts for my data collection exercise.

Looking forward to hearing from you soon. Best regards,

--
Eva Montenegro
Tallinn University / National Library of Estonia
linkedin.com/in/evamontenegro/en

Appendix 4: Codes used for the data analysis

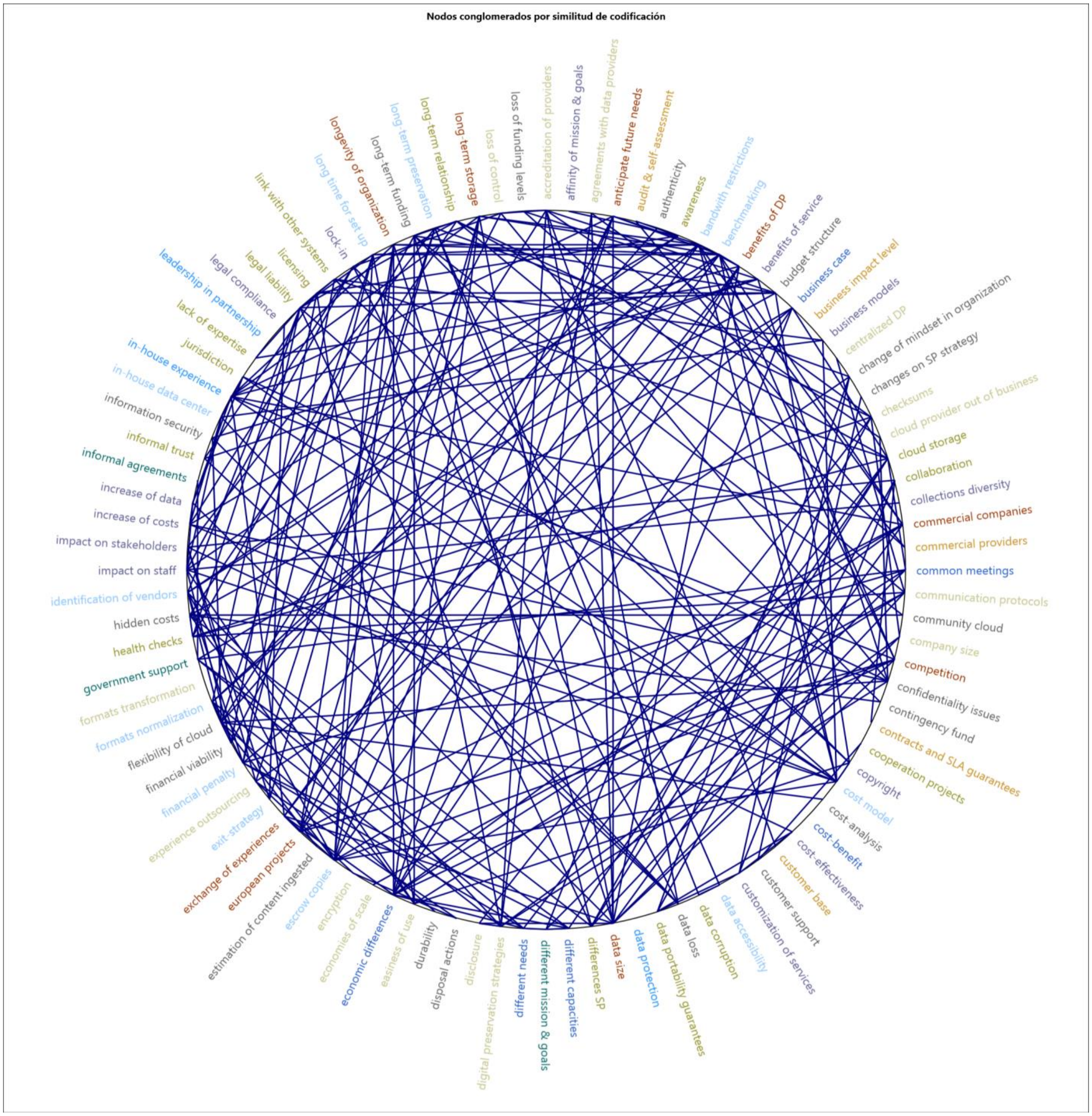


Figure 5 Nodes used for the codification, grouped by similitudes on the codification patterns