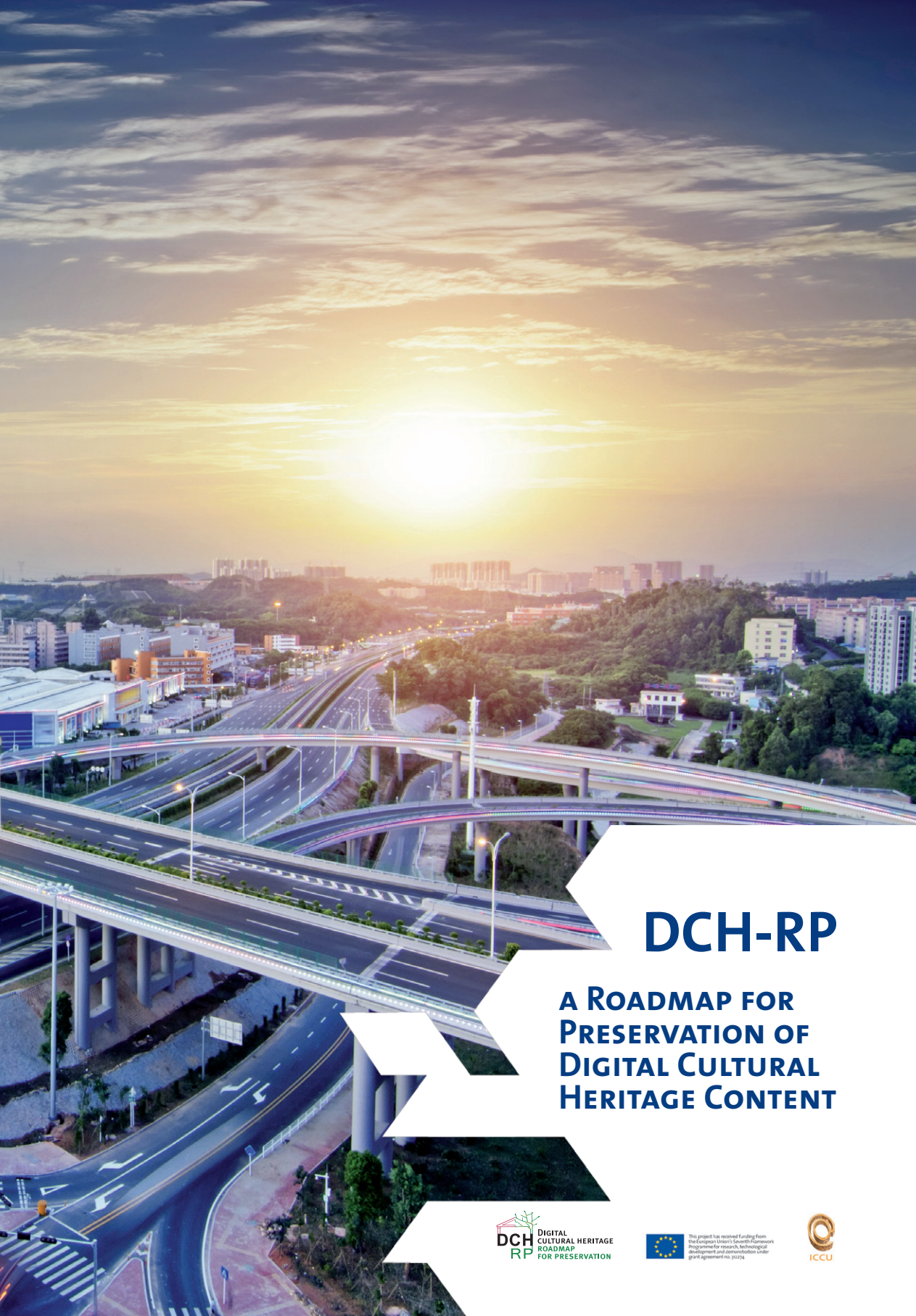


DCH-RP

A Roadmap for Preservation of Digital Cultural Heritage Content

DCH RP Digital Cultural Heritage Roadmap for Preservation

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 312274

ICCU

DCH-RP

A Roadmap for Preservation of Digital Cultural Heritage Content

DIGITAL
CULTURAL HERITAGE
ROADMAP
FOR PRESERVATION

DCH-RP Project
c/o ICCU
viale Castro Pretorio, 105
00185 Roma, Italy
Tel. + 39 06 49210427
info@dch-rp.eu
www.dch-rp.eu/

General Coordinator
Rossella Caffo

Technical Coordinator
Antonella Fresa

Coordinator of WP3
"Preservation Roadmap"
Börje Justrell

TABLE OF CONTENTS

Acknowledgements

Writing a handbook on the preservation of digital cultural heritage content is in itself a challenge. The fact that the focus of the handbook is on a roadmap for synergies between digital preservation and e-Infrastructure, i.e. distributed services, makes the challenge even more difficult and ambitious.

Knowledge in this particular field is just beginning to be explored, and good examples are rare and not easy to find. However, through the DCH-RP project's well-deployed network of common interests, representatives of stakeholder groups like DCH organisations, e-Infrastuctures, researchers, and commercial publishers have shared their expertise to provide us with invaluable inputs to this handbook. Many thanks for that.

Moreover, the project is in the favourable position of having gathered together highly skilled professionals from all over Europe. First of all we would like to mention Raivo Ruusalepp, who wrote a report on preservation services in the former DC-NET project, which sketched the basic concept for the DCH-RP project. Raivo has also had the lead in the work on a suitable trust model for DCH organisations to use e-Infrastructure services.

Special thanks go also to Tim Devenport from the international standards coordinating group EDItEUR, partner in the DCH-RP project, for helping us to shape the English language. Remaining peculiarities in its use must be blamed only on us.

We would like to express our gratitude to all the partners in the DCH-RP project, who have not only shared their great knowledge with us, but actually produced most of the basic information for the roadmap and, in consequence of that, this handbook. It has been a real teamwork.

Antonella Fresa and Börje Justrell

Foreword

This handbook has been developed as a part of the project Digital Cultural Heritage Roadmap for Preservation (DCH-RP), supported by the European Commission within the framework of the Seventh Framework Programme for Research and Technological Development.

The DCH-RP project has lasted for two years, from October 2012 until September 2014, with the participation and factual contributions of thirteen partners from eight European countries. Furthermore, a wide group of other institutions have been involved: as associate partners, as contributors of data in proofs of concepts conducted by the project, as participants in the project's events, as responders to surveys and questionnaires, or by accepting being interviewed and attending dissemination events where DCH-RP has been presented. The DCH-RP project has also established cooperation agreements with several other projects, in the first place EU financed ones, exploring opportunities of synergies and engaging in discussions about future initiatives. All these activities constitute a real network of common interest representing an important legacy of the project. This network will continue to exist after the DCH-RP project has formally finished, and will give concrete life to the actions foreseen in the present roadmap. The project has also created a dedicated web-space where the most recent version of the roadmap can be found together with other relevant material, such as information about services linked to the roadmap or supplementing it. This web-space is hosted within Digital meets Culture (http://www.digitalmeetsculture/net) and will be a practical instrument for keeping the network together.

The premise of DCH-RP is a shared vision among the partners of the need to implement a federated infrastructure dedicated to supporting the application of open science in arts and humanities.

Actions need to start now with a longer-term time horizon of not more than two decades. Even if this may appear to be a long period, this timeframe is an estimation deriving from the fact that such infrastructure implies the achievement of many complementary factors and steps. Naturally, the availability of technology, services and resources are basic conditions, and we can state today for sure that they are already mostly available. There are also on-going projects which are expected to deliver the necessary middleware and application services. Among them, and very important for the Digital Cultural Heritage (DCH) sector, is the work on preservation formats in digital archives. This research and development activity is based on the application of international standards whose outputs need in turn to be checked when implemented. A new Pre-Commercial Procurement project named PREFORMA[1] is expected to provide this area with a suite of open source instruments for conformance checking and reporting.

In addition to the technological area, two other domains are of extreme importance for the achievement of the DCH-RP vision: (1) innovations around the internal workflows of the organisations operating in the DCH sector and (2) policy support for the implementation of federated infrastructures.

Internal workflows currently encountered among DCH players imply that a number of actions need to be taken by many memory institutions that are engaged in digital archiving, in order to make their digital resources accessible and usable in the long-term. Firstly, roles inside the organisation have to be re-organised to guarantee that preservation is a mandatory step in the process of handling their digital resources. Secondly, in order to create new skills and competences, cultural heritage practitioners have to be trained in both understanding and the handling of the new conditions

[1] See www.preforma-project.eu

associated with digital preservation, i.e. the changing forms of artefacts and metadata, the changing methods of work, and the rapid changes in technology itself. Furthermore, decisions have to be taken about the procurement of services related to storage and computing resources, not just for storing the 'raw' data, but also to take into account preservation requirements. All these actions require time to be performed and financially resourced. Advocacy of the need for digital preservation is, therefore, another important action in order to create the conditions required for understanding, acceptance, and endorsement by decision makers.

Beside the challenge of taking care of a rapidly growing amount of digital resources, the DCH sector also has the challenge of handling the complexity of the information itself. Common procedures and workflows, shared internationally, would reduce costs in terms of both time and money to be allocated to this task and would contribute to the general interoperability and openness of scientific DCH data. The so-called 'hard sciences' are already demonstrating that research capabilities can be enhanced by the use of e-Infrastructures offering high-speed connections, shared computing and storage resources, sophisticated authentication and authorisation mechanisms etc. A basic assumption for the work of the DCH-RP project is, therefore, that existing e-Infrastructures for research and academia (including NRENs, NGIs and other data infrastructures) could also be efficient channels for the delivery of advanced services that can be used by the DCH sector in the field of digital preservation. One effect will be that DCH organisations will have to relinquish full control of where their data are physically stored. Another foundation for the DCH-RP project is the assumption that it will be possible to establish common policies, processes and protocols which will allow DCH organisations to access e-Infrastructures, despite the

fact that NRENs and NGIs are national entities, often with different policies and procedures for access and usage. The DCH-RP project's roadmap for preservation, therefore, concentrates on creating synergies between digital preservation in the DCH sector and existing and evolving e-Infrastructures.

Policy support is another fundamental component for the successful implementation of the federated infrastructure. Policy makers should be encouraged to endorse the roadmap in order to define policy and financial support to the federated infrastructure.

The varying interests of different stakeholders should be taken into account and harmonised. On one hand, e-Infrastructure providers should convincingly acknowledge the potential represented by the emerging DCH sector as an area for investment, in the same way as for other scientific domains. On the other hand, the cultural heritage sector must find methods for establishing and communicating trust in distributed digital preservation services provided by the e-Infrastructures so they can be commonly regarded as safe, secure and trustworthy. All this requires changes in professional approaches that need to be supported by awareness raising activities, and eventually also reflected in operating procedures, regulations and legislation. It is also important to liaise with the important actions that the DARIAH[2] initiative is carrying out.

This handbook aims to provide an overview of the challenges connected with distributed digital preservation of cultural heritage data and to illustrate which are the main missing components that are needed to enable the DCH sector to address these issues.

[2] www.dariah.eu

1 > INTRODUCTION

1.1 Document overview

This handbook is an outcome of the project DCH-RP (Digital Cultural Heritage Roadmap for Preservation), a coordination action supported by the EC FP7 e-Infrastructures Programme. It presents the project's main product, which is the DCH-RP roadmap to implement a distributed preservation infrastructure for digital cultural heritage.

The DCH-RP roadmap aims to provide a practical instrument to decision makers, offering an overview of the principal problems and challenges that digital preservation poses, a range of references to existing solutions, and a critical synthesis of the steps that memory institutions and policy makers should be ready to take.

The roadmap is based on the principle of a synergic liaison between digital preservation and e-Infrastructure. It is built on two basic assumptions: firstly, that existing e-Infrastructures for research and academia can also be efficient channels for the delivery of advanced services to be used by the digital cultural heritage sector for distributed digital preservation, and secondly, that it will be possible to establish common policies, processes and protocols which will allow digital DCH organisations to access e-Infrastructures, despite the fact that NRENs and NGIs are national entities, sometimes with different policies and procedures for access and usage.

The handbook starts by reviewing different aspects of the DCH-RP project and introduces aspects and challenges associated with digital preservation, especially when planning for distributed preservation systems and processes.

The memory institutions that participated in the DCH-RP project identified a number of services to address and requirements to set as priorities when a DCH organisation is planning for

distributed digital preservation. These services and requirements, aiming to cover the whole preservation process, are presented in a separate chapter.

The roadmap itself consists of four parts:

- A short review of what the roadmap stands for;
- An overview of the digital preservation landscape, describing the current situation as the DCH-RP project has captured it and some basic issues to consider when addressing distributed digital preservation services;
- The main components of the roadmap (vision, timeframe, appraisal and selection of digital resources to preserve, major areas, sustainability)
- Condensed versions of the roadmap for the short, medium and long term, strictly focusing on what to do and when.

The complementary Action Plan propose how to establish a value chain but has its main focus on the most important actions to take in major areas of the roadmap, for example to identify critical system requirements, to look into tools like the DCH-RP Registry of Services and Tools, and to take decisions about authentication and authorisation, a trust model, a governance model and a business model.

The DCH-RP project decided to create a dedicated web-space where it is possible to download the most recent version of the roadmap and the supporting documentation. The web-space can also be used for providing feedback and to give comments, i.e. it represents a kind of forum dedicated to the use of e-Infrastructure services and facilities for the long-term preservation of digital cultural content.

1.2 Document structure

Apart from the Introduction, this document consists of five chapters.

Background – This chapter reviews the common aspects of the DCH-RP project, such as aims and key stakeholders, the added value of the project, and the work carried out so far in the project; the aim is to give the reader a context within which this document can be considered.

Setting the Scene – This chapter provides an introduction to the many aspects associated with digital preservation and the main challenges to be faced while planning the implementation of a distributed digital preservation system.

Services to Address – This chapter lists the services that the roadmap should take into consideration as priorities for memory institutions addressing the preservation of their digital objects.

A Roadmap for Digital Preservation – This chapter presents the different parts of the roadmap, including condensed versions relating to the short, medium and long term, focusing on what to do and when.

An Action Plan – This chapter propose how to establish a value chain but points also at the most important actions to take in major areas of the roadmap.

2 > BACKGROUND

2.1 Aims and key stakeholders

The DCH-RP roadmap primarily targets two main communities in order to help them plan ahead:

- Policy makers on different levels, and
- Owners of digital preservation programmes at cultural heritage institutions.

The aim is also to assist managerial teams of cultural heritage institutions in taking decisions related to digital preservation.

2.2 Added value

The added value of this roadmap is the provision of specific recommendations for its various target groups. These recommendations are summarised in the following paragraphs:

Policy makers

- Policy makers should support and buy into awareness and communication programmes, which will help shape the requirements to be served by the actual policies.
- Links with the private sector will be necessary in order to sustain investment in digitisation, access and preservation of cultural heritage resources. Examples of relevant private partners include the publishing and tourism sectors. **Public-private partnership** needs to establish clear rules that respect the interest of both parties in a balanced way.

DCH institutions

- Preservation and access are closely related and rely on the availability of digitised content. When DCH institutions embark on digitisation programmes, they should include preservation and access components in these programmes.
- A range of **selection criteria** need to be agreed, also at European level, and with a multidisciplinary approach. These

criteria will help to identify priority areas for digital resources that need to be preserved and also to establish approaches on 'what' cultural heritage resources are (a content approach) and 'when' they become cultural (a process approach).

- **Training** of staff involved in the implementation of digitisation, access and preservation projects is fundamental. The quality of human resources has a crucial role in the success of the new initiatives.

E-Infrastructure providers

- E-Infrastructure providers need to develop a service-oriented approach. The DCH sector typically has low technical skills and needs to be supported in understanding how best to use infrastructure facilities.
- The DCH sector has its own very elaborated standards that must be recognised and understood by e-Infrastructure providers who want to serve this sector.
- Access to infrastructure services is a key aspect. This means that access should be open, but also controlled in order to guarantee the safety and protection of cultural assets.

Furthermore, social and cultural factors are major drivers; for these aspects the DCH-RP roadmap has considered four complementary dimensions:

- *Political dimension*
  Fragmentation of programmes between regions and Member States in Europe represents a barrier to be overcome.

- *Legal dimension*
  Diversity of regulations could represent an obstacle to common approaches in Europe. Awareness and information actions are very important to contribute to eventual harmonisation.

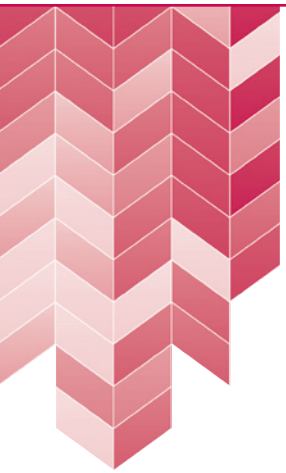
- *Financial dimension*
  'Who will pay' is a major question. Different national and regional funding systems need to be understood and common elements identified to support joint programming actions. The sustainability model needs to be chosen early: a **market/ business approach** where DCH institutions are considered 'customers' and therefore need to pay for the services they receive **versus** a universal **public service** where DCH institutions are treated as 'users' receiving public grants. In some countries **some kind of hybrid** of these two approaches is likely to be involved.

- *Organisational dimension*
  There is a strong need for raising awareness (incl. training) on digital preservation issues in DCH organisations.

Memory institutions should be able to exploit the most advanced results of the research without suffering the effects of using incomplete prototypes. In this light the Pre-Commercial Procurement mechanism put in place by the European Commission seems to be an interesting opportunity (the project PREFORMA[3] is a challenging example to be followed).

[3] www.preforma-project.eu

The time horizon for problems of digital preservation is often an order of magnitude longer than electoral cycles, and thus often too far from the mind of many decision makers. For this reason, a risk management approach allows the need for 'digital preservation' to be explained as insurance on a considerable investment, thus making the issue more understandable.



Alliances, closer collaboration and synergies should be established or strengthened between DCH and the 'research world' or its representatives, namely:

- Ministries responsible for Culture, Education and Research
- Research e-Infrastructures

Win-win cooperations should be created between DCH and e-Infrastructure actors:

- DCH can reinforce the research in the social sciences and humanities contributing to multi-, cross- and interdisciplinary research topics. This could be relevant also in the context of addressing societal and environmental global challenges.
- DCH is generating larger and larger amounts of 'research data'. In this light, the DCH sector is facing problems and challenges similar to those encountered by the hard sciences and solutions developed in other areas of the 'research world' should be exploited by DCH too.
- E-Infrastructure providers engaging with DCH can thereby serve another large community of users, reinforcing the social relevance of e-Infrastructures and strengthening advocacy for its governmental political and financial support.
- Other research sectors can find in the DCH domain another area for the exploitation of their results.

DCH has a potential role to play also in programmes aiming at e-government. This is particularly true today with regard to the archival sector and the creation of what will in the future become their digital content[4]. A similar approach has been successfully taken by the MICHAEL and MICHAEL Plus projects[5] whose results influenced and provided guidance to the initial design of Europeana[6].

Finally, the future of the roadmap process itself needs to be taken into account. The roadmap is only a start; it should continue

[4] Most e-government initiatives are on a national level. An example of coordination on the European level in the archival sector is the Archives Portal Europe with the ambition to get "……the archival landscape in shape for the quickly advancing digital future of our society." See http://www.apex-project.eu.

[5] www.michael-culture.eu MICHAEL and MICHAEL Plus were supported by the eTEN Programme as part of the e-government theme as initial deployment projects, involving several European Member States, with a total investment of more than 90 million euros, from 2005 until 2009

[6] www.europeana.eu the European Digital Library Flagship Initiative of the Digital Agenda for Europe

as a living document, owned by the community that DCH-RP has established through its Network of Common Interest and engaged in a broad and ongoing consultation process.

2.3 Work carried out in the project

The DCH-RP project lasted two years, from October 2012 until September 2014. This rather short time, however, allowed for a very rich programme of activities and interesting results.
The main concept of DCH-RP has been to understand the needs of the DCH sector concerning digital preservation, to study what is available – in terms of services, technologies, policies, programmes, skills and networks, and to identify the resources that the e-Infrastructures can make available for serving the DCH sector. Then, on the basis of these results, the roadmap described in this handbook has been derived and shared among the partners and the larger network of common interest built during the project. In particular, the scope of the project has been to explore the following aspects:

1 ► How to harmonise data storage and preservation **policies** in the DCH sector at European and international level;
2 ► How to progress with the **dialogue** among DCH institutions, e-Infrastructures, research and private organisations;
3 ► How to establish the conditions for these sectors to **integrate** their efforts into a common approach;
4 ► Which are likely to be the most suitable **models for the governance, maintenance and sustainability** of such an integrated infrastructure.

DCH-RP has taken a practical approach, based on identification and study of best practices and implementation of a large portfolio of Proofs of Concept.

In order to keep a very pragmatic focus, the memory institutions involved both as partners and as members of the DCH-RP Network of Common Interest defined a set of very concrete **scenarios** related to DCH preservation and grouped around three themes, as follows:

Theme 1 – "Organisational challenges"
- Use specialised digital preservation tools on in-house data
- Integrate a new tool into existing infrastructure
- Select an existing digital preservation solution at an institution with an advanced information technology support
- Preservation from a consortium of collections on the cloud
- Preserve a 3D visualisation
- Retrieve archived data

Theme 2 – "End user concerns"
- Researcher discovers a historical database
- Research and select a tool serving a specific purpose
- Access digitised content from schools
- Gain access to archived websites

Theme 3 – "New services & infrastructure integration"
- Proof of authenticity in distributed archiving
- Defining new services
- Integrating new services into existing infrastructure

Each partner selected which scenarios were achievable using available local and national resources, and fostering possible trans-national collaborations. The chosen scenarios became the basis of the proofs of concept[7], and yielded a broad spectrum of best practices.

Best practices have also been surveyed and studied by the DCH project in three specific areas:

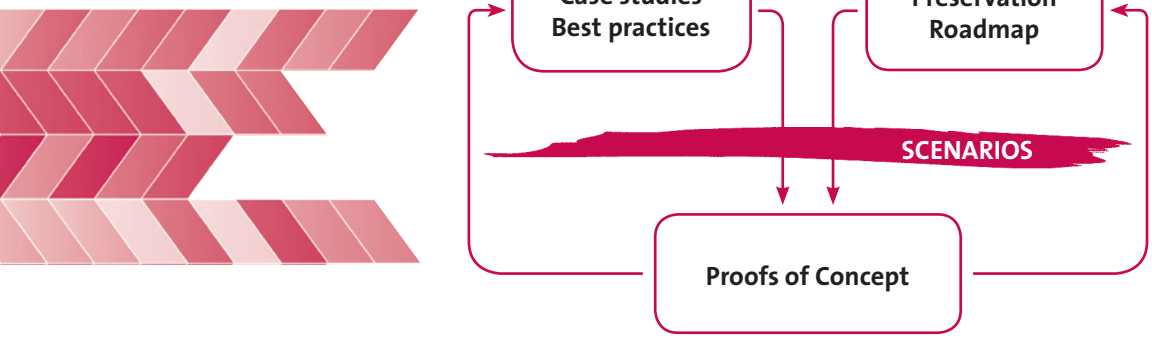
- Trust building, as a pre-requisite for any change in the existing organization of data storage of memory institutions

[7] Full documentation on the results of the DCH-RP Proofs of Concept are available in project deliverables D5.3 and D5.4

- Authentication and authorization, as tools contributing to the trust building process
- Engagement with the private sector, as a way to attract additional economic resources for the sustainability of the infrastructure

For each of these areas, the project has produced focused reports that entered into the process of building the DCH-RP roadmap. Proofs of Concept, best practices and the roadmap itself became the objects of an agile and iterative process of refinement and improvement, as illustrated in the following figure.

Figure 1: Building blocks of the DCH-RP workplan



While developing the roadmap, an important exercise has been to identify services that already exist, to map them against the most frequently used standards in the DCH sector and to develop a Registry of Services an Tools[8] to be offered to the DCH community as a practical resource. The following figure represents a screen-shot of the Registry of Services that is conceived as a living resource, where users are encouraged to contribute with comments, remarks and suggestions for improvement.

[8] The DCH-RP Registry of Services and Tools is available online at http://www.digitalmeetsculture.net/heritage-showcases/dch-rp/new-registry-of-services-and-tools/

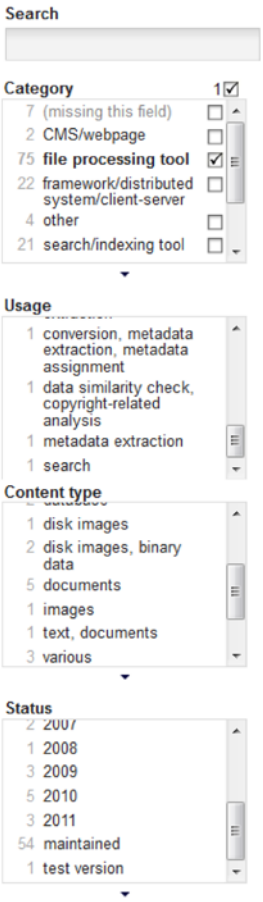


TABLE • LIST

75 filtered from 139 originally (Reset All Filters)

Name	OAIS stage	Functionality	Web
AccessToSiard	data mgmt	A collection of scripts to automatically convert MS Access files to the SIARD format.	http://sourceforge.net/projects/accesstosiard/
Acrobat XI Pro	ingest, data mgmt, conversion	Create, modify, convert from/to PDF documents. Convert a Web page or an entire Web site to a single PDF. offline viewing.	http://www.adobe.com/products/acrobatpro.html
AFF	ingest, acquisition	Tools for the creation of disk images, used in conjunction with the AFF open and extensible file format to store disk images and associated metadata	http://afflib.org/
Apache Commons Imaging	ingest, data mgmt	library that reads and writes a variety of image formats, including fast parsing of image info (size, color space, ICC profile, etc.) and metadata	http://commons.apache.org/proper/commons-imaging/
Apache POI	ingest, data mgmt	Java API for Microsoft Documents	http://poi.apache.org/
Apache Tika	ingest, data mgmt	toolkit detects and extracts metadata and structured text, using existing parser libraries	http://tika.apache.org/
AVS Document Converter	ingest, data mgmt	Transfer regular text formats to e-Pub format and create e-books. Open and convert such e-book formats as DjVu and FB2 to all key formats supported by AVS Document Converter.	http://www.avs4you.com/index.aspx
AVS Image Converter	ingest, data mgmt	Convert images between JPEG, PDF, RAW, TIFF, GIF, PNG, RAS, PSD, PCX, CR2, DNG, APNG, etc. Resize, rotate, apply effects, watermark pictures.	http://www.avs4you.com/index.aspx
b2x Translator	ingest, data mgmt	Tools for batch convertion from binary to XML MS Office formats.	http://b2xtranslator.sourceforge.net/
Bagit Library and Tools	??	Creation, manipulation and validation of data files according to BagIt specification	http://sourceforge.net/projects/loc-xferutils/
BitBlocker	data mgmt?	scan file sharing sites and BitTorrent trackers for copyrighted material, based on pHash	http://www.phash.org/

Figure 2: Interface of the on-line version of DCH-RP registry

3 > SETTING THE SCENE

3.1 Preserving Digital Objects

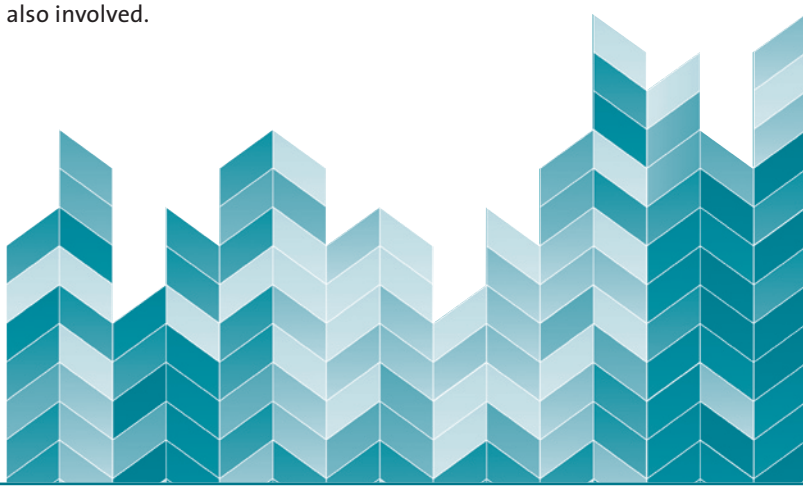
3.1.1 Definitions and strategies

The importance of preserving digital objects is well understood in today's society. Hardware and media obsolescence, lack of support for older computer formats, human error as well as malicious software can all lead to loss of digital objects. If several of these factors are involved, then the probability of loss is higher. Preservation, however, is not concerned only with sustaining single digital objects. To be used meaningfully in the future, digital objects should be preserved in contexts which make them understandable to future users. Digital preservation is defined by the DigitalPreservationEurope project as "a set of activities required to make sure digital objects can be located, rendered, used and understood in the future".[9] A more comprehensive term 'digital curation' is often used in parallel with digital preservation. It has a wider meaning and involves "maintaining, preserving and adding value to digital data throughout its life-cycle".[10]

The key challenge in preserving the usability of digital objects over time is to overcome technology obsolescence, but a set of other issues around managing collections of digital objects is also involved.

[9] http://www.digitalpreservationeurope.eu/what-is-digital-preservation/

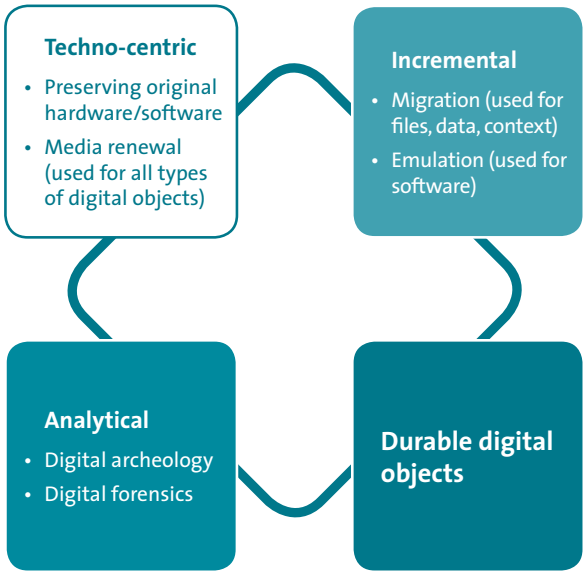
[10] http://www.dcc.ac.uk/digital-curation/what-digital-curation



During the past two to three decades, focus has moved from finding the 'ideal' long-term storage media to weighing the advantages and risks of different digital preservation strategies, and to define practical solutions based on standards that may use a number of strategies concurrently. Today, there are several strategies available for sustaining the use of digital objects in the future. The main ones are shown here:

Figure 3: Strategies for sustaining the use of digital objects

Source: *Digital Preservation Services: State of the Art Analysis* (Raivo Ruusalepp and Milena Dobreva)



The *techno-centric strategy* aims to preserve original hardware and software in a usable state in the future. It involves regular storage media renewal to make sure that the physical digital objects are not corrupted.

Incremental change relies on either migration of digital objects into new formats or preserving the formats of the digital objects and using emulation to be able to use them. The migration

strategy normally uses standardised file formats which are repeatedly converted to keep up with current technical generation. The emulation strategy preserves the original file formats and uses emulation at alternative levels. During technical generation changes either to the original software, to the original operating system or to the original technical platform are emulated into the new technical environment, in the latter cases combined with preserved original software.

Analytical strategies are currently based on techniques used in computer forensics. The underlying logic for this strategy is to apply specialised methods for recovery of objects which are in demand in the future instead of 'mass preservation' which does not seem realistic, bearing in mind the volume of digital information involved.[11] This is basically a strategy for selecting digital objects to be stored long term and methods most suitable for preserving them.

[11] The pioneering work in this domain was called *digital archaeology*

Yet another strategy seeks for methods of changing the formats of the digital objects in a way which allows the objects themselves to invoke preservation actions. Such objects are sometimes called *durable digital objects*.

The first three strategies require rigorous organisation of processes in organisations; the fourth one is still under development. All these strategies outline the principles of preservation; in practice they are implemented within archival lifecycles that integrate various tools and/or services. These lifecycles can be specific to organisations, depending on organisational mandate, the types of object they hold, and their target users.

Of the strategies mentioned here, the migration strategy has for a long time been the dominant one. Combined with the OAIS model – see below – it is used by most institutions working with digital preservation. Standardised file formats are normally used for the

digital objects to be preserved. To avoid technical obsolescence the digital objects are converted to new standardised file formats at the point of technical generation changes. These conversions are expected to be carried out without information loss. In the foreseeable future the migration strategy will probably continue to be the most used one, at least for in-house preservation. In a longer perspective, increased use of distributed preservation services like e-Infrastructures may change this situation.

Regardless of the chosen strategy or combination of strategies, cultural heritage institutions often make a distinction between the *master version of digital data* and at least one *surrogate delivery version*. The master version should contain as much intellectual, visual or audio content as possible, be saved in a standard (non-proprietary) file format, and preferably be duplicated across multiple locations. Delivery versions of data may be re-sized, compressed, and saved in whichever format is suitable for delivery to the user. Delivery versions are typically of lower quality (more compressed) than their original master files.

3.1.2 The OAIS model and the analysis of preservation layers

The diversity of both digital objects and types of cultural heritage institutions that are responsible for the preservation of digital resources creates variations in the level of tools used in practice, but the underlying process could be described as universal.
The pivotal standard in the domain, *ISO 14721 Space data and information transfer systems – Open archival information system – Reference model*, widely known as the OAIS model, is a functional framework that presents the main components and the basic data flows within a digital preservation system. It defines six functional entities that synthesis the most essential activities within a digital archive: ingest, preservation planning, archival storage, data management, administration, and access. Recently, some major European libraries have proposed to combine these six stages into a smaller number of use-cases that preservation systems address.[12]

[12] A report of four major national libraries in Europe looks at three core functions – ingest, retention, and access (*Long-Term Preservation Services*. A description of LTP services in a Digital Library environment. BL, KB, DNB, NB, 2010; see also http://public.ccsds.org/publications/archive/650x0m2.pdf)

The OAIS model looks at data stored in the digital archive as a fluid object that can (co-)exist as three types of information packages:

- Submission (SIP) is used to transfer data from the producer to the archive,
- Archival (AIP) is used for the archival storage and preservation,
- Dissemination (DIP) is used within the access function when consumers request archived materials.

As a reference model, the OAIS standard does not imply a specific design or formal method of implementation. Instead, it is left to users to develop their own implementation by analysing existing business processes and matching them to OAIS functions.

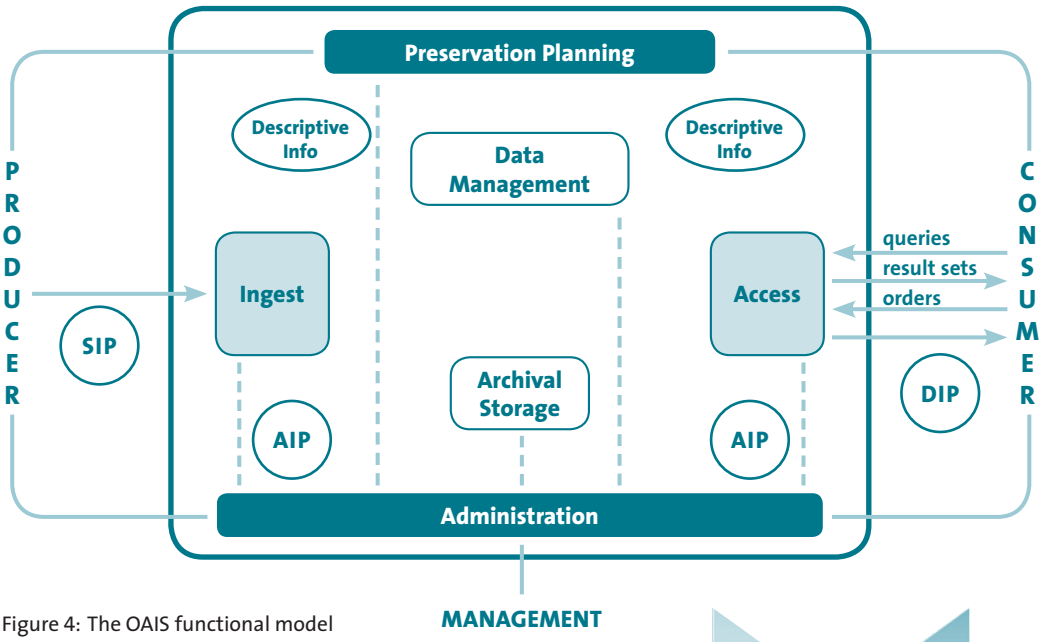


Figure 4: The OAIS functional model

In maintaining the accessibility and usability of digital objects over time, an often used method for analyzing them is built on the presumption that every digital object consists of three layers: a physical, a logical and a conceptual. All three layers and their relations have to be considered and understood in order to decide on appropriate preservation actions. These actions are often identified and referred to as "bit preservation", "logical preservation", and "semantic preservation".

Bit presentation is seen as a number of basic actions ensuring the integrity of the 0s and 1s (the sequence code) over time and serves as the foundation for any other preservation actions.

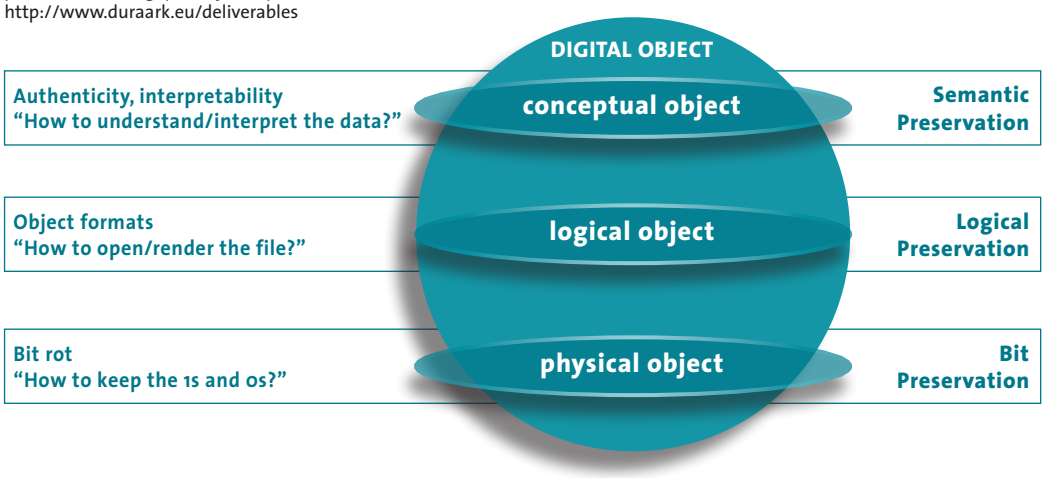
Logical preservation focuses on the representation of the digital object and activities in this field have the aim of ensuring the quality of being able to reproduce the object and maintain accessibility over time. File format sustainability is of course one major issue here. Much effort has been invested over the years on setting up requirements and recommendations for file format sustainability.

Semantic preservation includes activities focusing on understanding the content long-term but also on capturing contextual information about the domain/ environment in which the digital object was created.[13]

[13] See the EU project DURAARK, deliverable D6.6.1 Current state of 3D object digital preservation and gap-analysis report, (http://www.duraark.eu/deliverables) and references therein; concerning research in file formats, see for example the InterPARES project (E. Peters McLellan, General study 11: Selection digital file format for long-term preservation. Online, March 2007, http://www.interpares.org/ip2/ip2_general_studies.cfm)

Figure 5: The layers of a digital object

Source: EU project DURAARK, deliverable D6.6.1 Current state of 3D object digital preservation and gap-analysis report, see http://www.duraark.eu/deliverables



The core challenges addressed by DCH-RP are in the first place targeted towards the OAIS preservation functions, but they are interconnected with a number of other functions that together form the digital archive.

3.1.3 Digital preservation and roadmaps in a European context

Member States of the EU have taken the position that the preservation task should be their responsibility. Therefore, each Member State is developing and implementing its national preservation strategy, which includes the preservation of digital master copies that takes place at national memory institutions or

at other public institutions which are the direct responsibility of governments. National frameworks that regulate this area, like rules on legal deposit and the handling of public records, exist and the publishing sector is also involved, especially with regard to born-digital material.

However, there are many commonalities that exist among the national preservation strategies which have to be addressed in common and in a coordinated manner among memory institutions, the Member States of the EU and more generally internationally in order to share solutions and to contribute to interoperability and openness. Common procedures and workflows, shared internationally, would reduce the cost both in terms of time and money to be allocated to digital preservation and would contribute to the general interoperability and openness of scientific data (including research data from the DCH sector) which is stated as the priority for the global knowledge society.

The importance of long-term preservation and its complementarities with digitisation efforts was highlighted in the report of the Comité des Sages (Reflection group on bringing Europe's cultural heritage online) that clearly stated the digital preservation mandate of memory institutions.[14] Also important is the EC Recommendation on digitisation and online accessibility of cultural material and digital preservation[15] published by the EC on 28/10/2011.

The attention and commitment of the EC to research and development in the domain of digital preservation was highlighted at the Commission's expert workshop *The Future of the Past*, held in Luxembourg in May 2011.[16] This workshop discussed previous research agendas in the domain of digital preservation and formulated a number of potential research topics of high relevance to the future development of the domain,

[14] The New Renaissance, 2011: 6

[15] Full text of the recommendation is available online at: http://ec.europa.eu/information_society/activities/digital_libraries/doc/recommendation/recom28nov_all_versions/en.pdf

[16] Billenness, C. (2011) The Future of the Past, Report on the Proceedings of the Workshop, European Commission, Luxembourg, 4 – 5 May 2011. Available: http://cordis.europa.eu/fp7/ict/telearn-digicult/future-of-the-past_en.pdf

among them digital preservation infrastructure – a key area of focus for the DCH-RP roadmap.

Roadmaps are useful instruments both for presenting the scope and coverage of an e-Infrastructure and to highlight the steps to be taken by the actors involved at different levels (technical, managerial, political, financial, societal, etc.) for implementing it. Roadmaps are also frequently used within projects and institutions in the digital preservation domain. Some roadmaps can be very detailed, as for example the roadmap developed for the UK Parliamentary archives (2008)[17] which presents environmental, policy, preservation, presentation, standards, skills, and communication developments over time. The Open Planets Foundation developed a *Tools and Services Roadmap*[18] to outline their software development plans. The APARSEN project roadmap[19] presents research topics and larger themes; preservation services are a research topic under the theme of sustainability. Some projects use roadmaps to present various formats, e.g. the PrestoSpace[20] project presents formats for audio-visual material. The 4C project is in the process of developing a roadmap for tackling issues around the cost of digital curation.[21] There are also a number of national roadmaps, especially in the area of research infrastructures that address arts and humanities.[22]

However, there is no existing roadmap which encompasses the use of the e-Infrastructures for digital preservation that the DCH-RP project could build on or progress further. The project has had to develop its own roadmap for the specific domain and task that it is addressing. This roadmap will be supplemented by practical tools which will help on one hand the monitoring of activities and thus would be of benefit in a political context, but will also offer knowledge instruments to stakeholders from the DCH domain (see chapter 2.1 above) to make informed decisions on digital preservation.

[17] http://www.parliament.uk/documents/upload/strategy-road-map-final-public.pdf presents the roadmap diagram and http://www.parliament.uk/documents/upload/digital-preservation-strategy-final-public-version.pdf - the justification.

[18] http://www.openplanetsfoundation.org/community/tools-and-services-roadmap

[19] http://www.alliancepermanentaccess.org/index.php/current-projects/aparsen/aparsen-roadmap/

[20] http://wiki.prestospace.org/pmwiki.php?n=Main.roadmap

[21] http://4cproject.eu/roadmap

[22] See for example the Danish roadmap for RI http://en.fi.dk/publications/2011/danish-roadmap-for-research-infrastructure-2011/uk-roadmap.pdf; Large research (Czech roadmap, 2010) http://www.infrafrontier.eu/docs/national_roadmaps/Roadmap_CR.pdf; Australian humanities infrastructure http://www.paradisec.org.au/blog/2011/03/australian-humanities-research-infrastructure-funding/

3.2 Main challenges

The main challenges that the DCH-RP project has to meet have, of course, their roots in difficulties with preserving accessibility and usability of digital objects over time, but they are also connected to questions like the benefit of using e-Infrastructures for preservation, what to preserve, various sustainability issues and how to raise awareness about the roles of different actors in the implementation of a distributed digital preservation infrastructure.

3.2.1 Making current and future digital information accessible and usable over time

The general challenges that the cultural heritage sector faces in making current and future digital information accessible and usable over time can be summarised as follows:

- Solutions for preservation must have a high level of automation and self-reliance to be able to handle the rapidly growing amount of DCH information;
- The tremendous rapidity in the development of new technology requires preservation solutions adaptable and flexible enough to really solve permanence and longevity issues;
- The infrastructure and organisational models must be highly scalable and adaptable to the various levels of input, storage and access required now and in the future.

Keywords when addressing these challenges by using a distributed preservation infrastructure include these:

- Distinct functional and technical requirements;
- Solid models for handling business issues, governance and trust;

- A service architecture that altogether can guarantee the authenticity of the digital resources over time, physically and technically preserve them over time, and verify that they are accessible and usable over time.

The challenges in making digital information accessible and usable in the long-term are also closely related to a number of noteworthy differences between digital and paper-based materials. But it is not only the changing form of cultural heritage objects that is new. The changing ways of work that follow the introduction of objects in digital form, force cultural heritage institutions to integrate new concepts, methods and tools for digital preservation to be carried out in parallel with traditional paper-based preservation.

The Digital Preservation Coalition has pointed out six main differences between digital and paper-based materials: [23]

[23] http://www.dpconline.org/advice/preservationhandbook/digital-preservation/strategic-overviewn

Machine dependency - digital materials require specific hardware and software in order to access them.

Technology obsolescence - the speed of changes in technology means that the timeframe during which action must be taken is very much shorter than for paper. It is measured in a few years compared to decades or even centuries when preserving traditional materials. Technology obsolescence is, therefore, generally regarded as the greatest technical threat to ensuring continued access to and use of digital resources.

Fragility of the media - the media digital materials are stored on are inherently unstable and can break down very quickly without suitable storage conditions and management, even though they may not appear to be damaged externally.

Loss of integrity - the ease with which changes can be made and the need to make some changes in order to manage the

material means that there are challenges associated with ensuring the continued integrity, authenticity, and history of digital materials.

Doing nothing is not an option - the implications of allocating priorities are much more severe than for paper. A digital resource which is not selected for active preservation treatment at an early stage will very likely be lost or become unusable in the near future.

Preservation prior to creation - the nature of the technology requires a life-cycle management approach to be taken to the maintenance of digital resources. A continual programme of active management is needed from the design and creation stage of a computer system and onwards, if preservation of that system is to be successful.

All of these differences are interconnected, and together they clearly indicate that a radically different approach is required in managing digital objects compared with paper-based ones. It is also important to have in mind that the greatest asset of digital information, the ease with which it can be copied or transferred, is paralleled by the ease with which the information can be corrupted or deleted.

3.2.2 Showing the benefit of using e-Infrastructure for preservation

Results from the work on the DCH-RP project's roadmap shows that the two basic assumptions on which the DCH-RP roadmap is built are achievable:
Firstly, existing e-Infrastructures for research and academia are also efficient channels for the digital cultural heritage sector to use them for distributed digital preservation.
Secondly, it is possible to establish common policies, processes and protocols to allow digital DCH organisations to access e-Infrastructures, despite the fact that NRENs and NGIs are

national entities, sometimes with different policies and procedures for access and usage.

However, a ground breaking part of the concept is the possibility of customising the services provided by e-Infrastructures, i.e. tailoring the service portfolio and characteristics to the actual preservation tasks and requirements. Even if the e-Infrastructure resources already today seem to be allocated in ways that could support preservation functions and sub-functions quite well, the general conclusion must be that the market for distributed digital preservation services is still in its infancy.

From the public assessment of an intermediate version of the roadmap it was clear that no convincing proofs have yet been presented that cooperation between the DCH sector and e-Infrastructures for the delivery of advanced digital preservation services gives better value for money than other solutions. The value chain that DCH and e-Infrastructures can create together is still to be defined!

E-Infrastructures are not for free, and there are not many commercial distributed preservation systems in place today. E-Infrastructures are normally not especially skilled in preservation but have great knowledge about data management. So, if e-Infrastructures have to develop preservation systems it will be costly, but e-Infrastructure providers can on the other hand benefit from economies of scale.

The DCH-RP project has looked into other domains, to see if there are experiences in the digital preservation field that are transferrable to the DCH domain. In the e-journal preservation community, much has been achieved in terms of evolving mechanisms and organisations to look after digital preservation. The technical, organisational and financial challenges have been proved to be solvable, given strong commitment from the

communities involved. The key issue seems to be the ways in which these communities have organised themselves to bring about long-term agreements and infrastructures for preservation.

Another important issue is the level of maturity in the DCH sector to handle distributed digital preservation solutions. E-Infrastructures can reach their maximum potential in serving the DCH preservation community only if the DCH sector is prepared to exploit the opportunities the e-Infrastructure provides. This is obviously not always the case today. Representatives of both e-Infrastructure providers and DCH institutions express feelings of dissatisfaction, the latter also reporting difficulties in utilising the facilities and tools offered.

3.2.3 Models for what to preserve

Which digital objects need to be preserved, and which can actually be preserved - and how should the choice be made? The question of a selection process arises sharply given the huge amount of digital resources that are produced and waiting to be preserved. Memory institutions may have a public commitment to preserve (including legal deposits and public archive duties), but one cannot talk about digital culture preservation without talking about value. What is important to some users (contemporary and future) is not necessarily important to others.

Criteria for disposal of cultural heritage information and objects are in place but differ between domains and professions. For example, the archival community has for a long time used methods for classification of archival records with respect to their value, mainly based on evidential and informational aspects. Handling this appraisal process is a fundamental part of an archivist's professional duties.

Different types of data and digital object also require different types of preservation methods and activities. This has already,

within all sectors in the cultural heritage domain, had an impact on expected areas of expertise and resulted in a growing number of professionals with hybrid competences (library science/archival science/museology and ICT).

Another aspect of what to preserve is that different countries have different rules about preserving digital data. It is for example not always possible to preserve valuable data outside a particular country or on a server owned or controlled by a commercial service provider.

3.2.4 Sustainability issues

There are different dimensions of sustainability that the roadmap has to tackle: the sustainability of the roadmap itself and the sustainability (political, financial, organisational, and technological) of the preservation infrastructure to which the roadmap leads.

For the sustainability of the roadmap and thus to ensure a clear direction, the roadmap needs an end point and should be maintained as long as this end point has not been reached.

For the sustainability of the preservation infrastructure, funding and opting for a pan-European solution is regarded by many as a must. Broadly recognised and accepted standards are a basic part of the concept. There seems to be a widespread opinion that digital preservation cannot be realized without funding at a national level (storage, software, etc.) and to some extent at a European level. The way to a digital preservation arrangement/service that "runs itself" is too long and too winding. Actions are needed already in the short term.

To get a sustainable preservation infrastructure there must also be willingness in the DCH community to cooperate at national as well as pan-European level. Otherwise the e-Infrastructure will be

presented with a too diffuse and fragmented customer market for their distributed preservation services, and the roadmap will run the risk of remaining an abstract document.

In the e-journal domain, the governance arrangements developed around e-journal preservation have generally been carefully designed to involve and keep on board a variety of potentially competing interests – in order to achieve a shared common good that can be sustainable into the future.

3.2.5 Awareness raising

When using distributed preservation services from e-Infrastructure providers, raising awareness is an important element:

Firstly, the owners of digital collections have to understand the importance of preserving their content; training and learning resources should, therefore, be made available for this purpose.

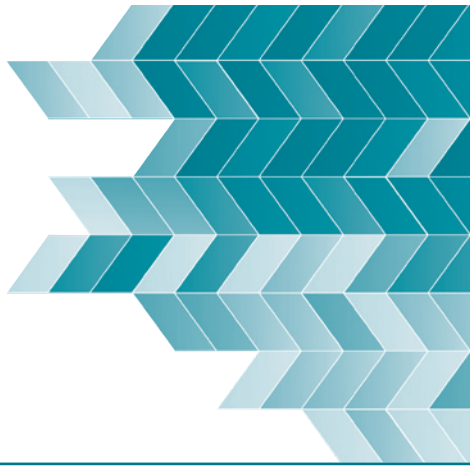
Secondly, cultural managers (museums, libraries, archive directors) have to support the owners in designing the correct workflow and understanding the digital preservation phase as an integral part of the digitisation process. The dramatic speed of technological change has led to many organisations not being able to fully articulate their needs in this field, much less employ or develop staff with appropriate skills.

Traditional training in the cultural heritage sector does not always provide knowledge of skills and tools necessary to deal effectively with emerging information technologies. Neither does individual self-improvement by staff members, which smaller institutions sometimes rely on. The step from in-house digital preservation to distributed solutions makes it also crucial to understand the concepts and procedures used in information system applications from a preservation perspective. Today's

situation has resulted in a growing number of professionals with hybrid competences (library science/archival science/museum science and IT) within all sectors in the cultural heritage domain. However, these professionals are too few in number and not well enough equipped to fully embrace the preservation implications of digital objects created and embedded in different kinds of information system applications.

Thirdly, policy makers have to understand the need for a plan for investments in preservation as long as they plan for investments in digital assets. Member States have recognised their responsibility in digital preservation, and national preservation strategies are implemented all over Europe. But these strategies have to be followed by concrete efforts in terms of resources, human as well economic. To create digital objects, born digital or converted from analogue originals, without taking into account the preservation of these objects is a very risky approach.

Fourthly, the users of the digital resources have to be aware of the various contributions they can provide to the digital preservation process (annotations, inputs for content selection, etc.). User involvement is still at a basic stage, but concepts like "crowd-sourcing" and "user-interaction" are rapidly growing in the cultural heritage domain.



4 > SERVICES TO ADDRESS

The following services to address and requirements to set up have been identified by the memory institutions that participated in the DCH-RP project as priorities when planning for distributed digital preservation.

Naturally, this list is a high-level summary of the areas of services and requirements to be considered and each institution will need to explore the details, with regard to the individual work-flows and specific vocation of the organisation concerned.

The services and requirements discussed in this chapter aim to cover the whole preservation process and are also related to the OAIS model as illustrated in the following figure:

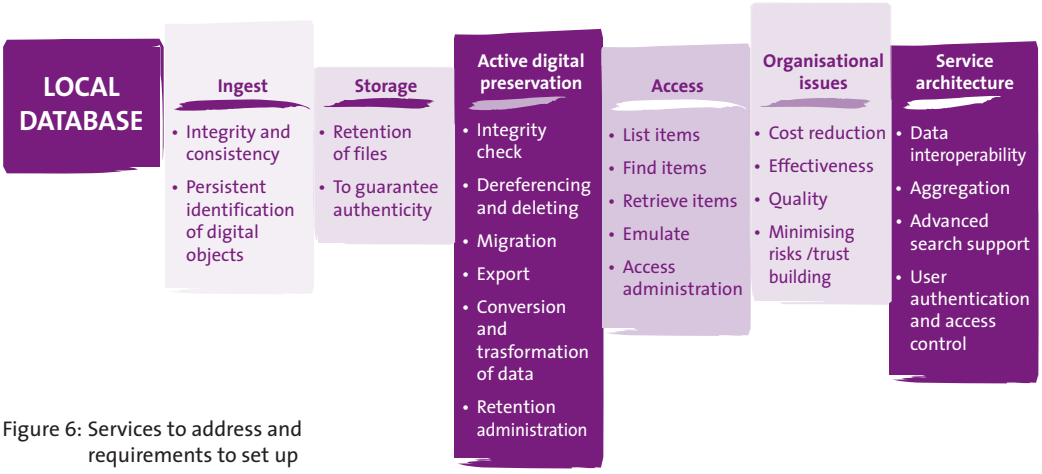


Figure 6: Services to address and requirements to set up

4.1 Ingest

To ingest different record types to an e-Infrastructure-based preservation system, all files:

- Need to be checked for integrity and consistency with standards using automated routines that document the outcomes of checks;
- Need fixity information to be associated with them, including persistent identifiers that will allow for identification and to check file integrity at any point in time.

Meeting these requirements makes it possible for the cultural heritage institutions to evaluate:

- To what extent tools for the required ingest processes are in place;
- How well they are running;
- What are the time and effort required.

Check points: Tools run without failures - Processes run fast - The integrity of all files can be checked after the ingest process - The level of automation of the entire process is high - Time and effort required is manageable.

4.2 Storage

An e-Infrastructure-based preservation system has to store the files in such a way that they can be retained with full accessibility and usability. The authenticity of the files should also be guaranteed. Strategies for replacing obsolete technology with new technology have to be in place.

Meeting these requirements makes it possible for the cultural heritage institutions to evaluate:

- To what extent the requirements on storage are met;
- What are the time and effort required.

Check points: Requirements on formats and standards for raw data are fulfilled - Appropriate metadata standards are in place as well as a trustworthy strategy for replacing obsolete technology - Time and effort required is manageable.

4.3 Active digital preservation

An e-Infrastructure-based preservation system has to have a number of complementary curation services including:

- Schedule-based integrity checking;
- De-referencing and deleting;
- Migration of (and possibilities to actually move) preserved files to new versions of software and/or hardware;
- Possibilities to export data;
- Conversion and transformation of data;
- Administering retention.

Meeting these requirements makes it possible for the cultural heritage institutions to evaluate:

- To what extent an e-Infrastructure is mature enough for implementing active digital preservation;
- What additional capacity it needs to develop in case there are any deficiencies.

Checkpoints: Tools run without failures - Curation services run quickly and meet the requirements - Level of transparency is acceptable - The level of automation of the entire process is high - Time and effort required is manageable.

4.4 Access

Services needed are to:

- List items;
- Find items;
- Retrieve items;
- Emulate;
- Administer access.

Meeting these requirements makes it possible for the cultural heritage institutions to evaluate how they can select services meeting their needs for access, and how to select from available offers.

Checkpoints: Tools run without failure - To what extent services for access are in place and are running well - Time and effort required is manageable - Matrix of metrics and minimum requirements for quality are in place.

4.5 Organisational issues

There have to be clear agreements on outsourcing in place covering aspects like:

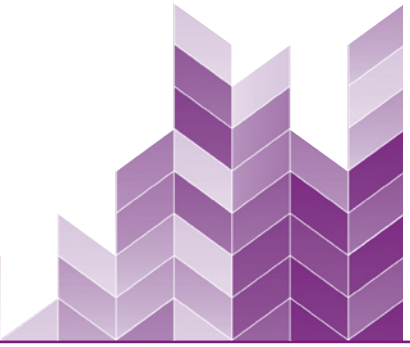
- Cost reduction;
- Increased effectiveness;
- Increased quality;
- Acceptable level of resources (technical and human);
- Minimising risks/trust building.

Policies for outsourcing have to be decided by the cultural heritage institutions. The level of technical and human resources should be taken under control in order to keep it at an acceptable level, as fixed by the archival owner.

Meeting these requirements makes it possible for the cultural heritage institutions to evaluate how e-Infrastructures are able to handle distributed digital preservation.

Check points: Draft text of agreement that both the cultural heritage institutions and the service providers have judged to be right and praiseworthy.

4.6 Service architecture

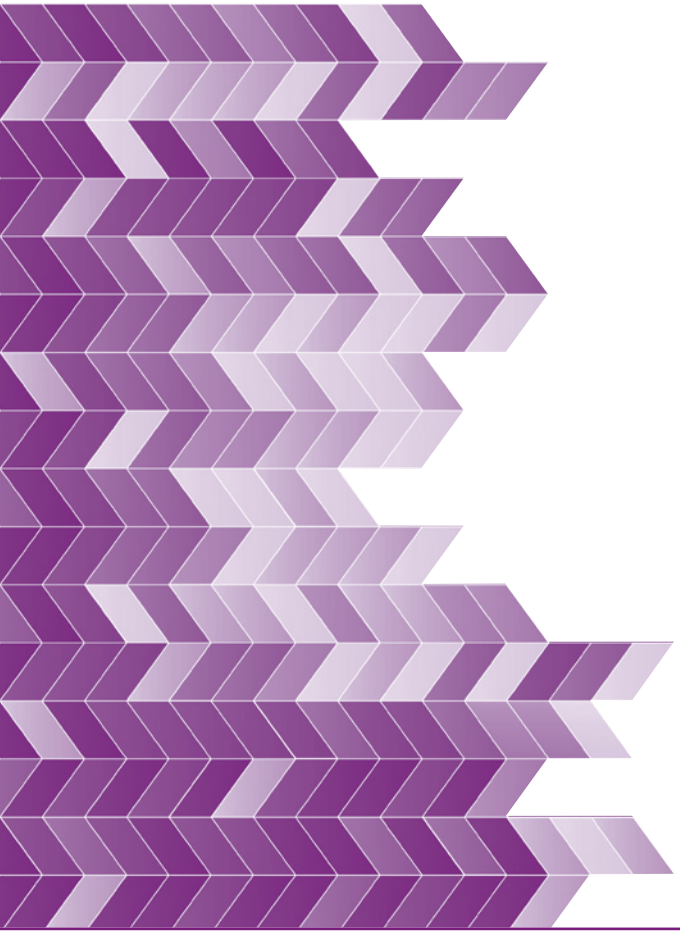


Agreements on standards have to be in place that cover services like:

- Data resource setup interoperability;
- Aggregation;
- Advanced search support;
- Persistent identifiers;
- User authentication and access control.

Meeting these requirements makes it possible for the cultural heritage institutions to evaluate to what extent an e-Infrastructure has the capacity to offer the service architecture needed.

Check points: Draft text of agreement that both the cultural heritage institutions and the service providers have judged to be right and praiseworthy.



5 > A ROADMAP FOR DIGITAL PRESERVATION

5.1 The roadmap as an instrument

The "map" in the roadmap draws the landscape of digital preservation for the DCH sector based on the current situation, but needs also to take into account how the situation will change in the future. Much depends on the maturity of both the preservation process in the DCH sector and the preservation services available from the e-Infrastructures. Preservation and access need to have a dynamic approach. It is important that the preservation process does not remain only a post-production task.

The digital preservation landscape is also changing at different levels; technical, political and legal. Distributed solutions like government clouds are becoming increasingly prevalent and some DCH institutions may be forced to make use of them. Data infrastructures with a portfolio of services, including different levels of storage and preservation, are being built. Societal changes have also to be taken into consideration.

The "road" in the roadmap points to an action plan, and actions are needed in a number of areas: tools, services, authentication, trust, governance models, user requirements, funding models and business models, skills / training / awareness etc. It goes without saying that many of these areas are relevant not only for digital preservation but exist also in other domains. It should not be forgotten that DCH data is research oriented, and many commonly identified challenges are shared with the handling of traditional research data.

The DCH-RP roadmap integrates three domains of necessary intervention (business change, policy framework and better tools) with the major PEST factors (political, economic, scientific, and technological). The compilation of the roadmap also needs integration of a multitude of viewpoints and aspects, both those foreseen in the planning of the project and new ones discovered during the project's lifetime.

Most cultural heritage institutions have in-house solutions for storing and processing their digital collection and holdings. When comparing in-house digital preservation with distributed e-Infrastructure services, it is inevitable that some discrepancies will appear, such as incompatibility of purposes or scope, lack of technical or semantic interoperability, reliance on different standards, and jurisdictional and legal barriers, etc. Therefore, the DCH-RP roadmap has a strong focus on what to do and on the usability of services and technologies.

In order to achieve this, the DCH- project has adapted the following structure for its roadmap:

- Firstly, it gives a description of the digital preservation landscape ("the map"), concentrating on how to meet stakeholder needs and the selection of the main components of the roadmap.

- Secondly, it identifies an action plan ("the road") with challenges and potential benefits to target, practical actions to take up, and services to address.

5.2 Describing the digital preservation landscape – to meet stakeholders' needs

5.2.1 A snapshot of the current situation

When the DCH-RP project has met digital cultural institutions to get their view on distributed digital preservation – in workshops, proofs of concept, questionnaires etc. – the outcomes have indicated a very fragmented picture, which can be summarised as follows.

It is important that long-term preservation issues are already taken into account during the creation of digital information; examples given on such issues are format conversion and storage. On the other hand 'storage' of digital objects is often defined

as purely technical storage (on bit level), and 'preservation' as securing the stored objects in such a way that future users can reach, access, and understand them.

There seems to be fairly broad consensus about the time spans of preservation solutions:

- Short-term preservation – solutions that are used for a short time, 5 years maximum.
- Medium-term preservation – solutions that are used during a system's lifetime, 10 years maximum.
- Long-term preservation – solutions that are used after the originating system's lifetime, number of years unspecified.

That fits well with an opinion expressed by many of the cultural heritage institutions that digitised objects and "born-digital" objects are to be considered differently, mostly because of the link that exists between a digitised object and its "original" physical form (i.e. a digital object may just be considered as a digital copy of a certain physical object).

It seems to be a common opinion that preservation solutions proposed by DCH-RP should be tailored towards domain specific requirements, but on the other hand many cultural heritage institutions say, when asked, that the project should propose "hybrid" solutions including both generic elements and others specific to the cultural heritage domain.

There is a general concern in the cultural heritage domain about the continually increased amounts of digital heritage content, which will lead to higher costs both for its management (including storage) and for preservation. For those cultural heritage institutions that already have digital objects, but as yet have no process or specific systems for dealing with them long-term, the situation sometimes appears close to being desperate and not manageable. Therefore, when asked

whether the DCH-RP project should offer any advice on how digital materials are selected for preservation, the answer is in most cases a clear YES (in capital letters!).

The collections and holdings at cultural heritage institutions are sometimes described in a (so called) "traditional" way, and international standards for metadata are, in these cases, seldom implemented fully. Thus, metadata structures differ greatly between the institutions. However, within the library community, stable and widely accepted standards are frequently used, both for metadata and raw data. There are often also systems for long-term preservation available.

Archive information (both metadata and raw data) is described and preserved according to established principles and international standards. As is the case for libraries, there are often systems for long-term preservation in place.

Today, many cultural heritage institutions often experience a lack of support in setting priorities for digital preservation, and for developing routines and strategies for long-term preservation. Inside the institutions, a clarification of internal roles necessary for digital preservation (how responsibilities should be divided between the administrators and managers of information and their IT systems counterparts) is often asked for.

Cultural heritage institutions seem fairly often to be in favour of centralised solutions for storage/preservation and centralised support functions like handbooks and training – but are equally vocal that these should be located within the cultural heritage sector!
Most of them say, when asked, that they think it is vital, very important or at least desirable that the DCH-RP project should address outsourcing issues in the context of using shared digital preservation infrastructure and services, but only a third clearly state that they have considered outsourcing any of their preservation-related responsibilities to a shared service.

Most cultural heritage institutions appear not to have links with e-Infrastructures and about one half of them say that they might

be prepared to consider private sector solutions or partnerships as components in their digital preservation programmes. The other half do not know. Just a small number of institutions say no, when asked this question.

The lack of training is often reported as an issue in discussions about digital preservation. Most Member States seem not to have any establish organisation or focal point for professional cultural heritage training in this area. Surprisingly many institutions declare that they are nevertheless prepared to act as "centres of excellence" to spread best practice on DCH preservation to other institutions

5.2.2 Distributed digital preservation services

Different parts of the DCH domain have clearly different needs, depending on whether they are small or large, the kind of digital objects for which they are responsible, etc. The conditions (e.g. resources) for managing digital preservation differ also quite widely. Services for distributed digital preservation, therefore, have not only to be flexible, but they must also be easy to adapt and utilise, and address several areas. Some basic issues when addressing distributed digital preservation services include these:

- Functional requirements;
- Service types and objects to address;
- Type of service architecture;
- Level of maturity;
- Licensing conditions.

I. FUNCTIONAL REQUIREMENTS

Getting the right set of functional requirements in place is crucial, but equally important is that these requirements are defined in a way that makes it possible to measure how they are met. In

the previous Chapter 4 a number of services to address and their functional requirements were defined, following the OAIS model and/or the preservation process.

II. SERVICE TYPES TO CHOOSE

There are two main levels of services for distributed digital preservation, which can be considered as basic for the DCH community:

- Level 1: those already available or that could easily be made available by e-Infrastructures to support digital preservation activities conducted by cultural heritage institutions. This 'kiosk-model' could contain supplementary services like federated authentication, audit and certification, and persistent identifier distribution, which are typical network services that would make life easier for institutions or networks of institutions that manage digital preservation 'on their own'.
- Level 2: those cloud or grid based 'turn-key' services that can offer the entire process covering all the phases and functions of the OAIS model, with a particular focus on storage, curation services and other organisational aspects like trust.

The advantages of such two-level service architecture would be:

- It would allow a gradual approach to digital preservation services, paid or payable on the cloud or grid-based, by cultural

heritage institutions that have digital objects but difficulties in managing them; an institutions can initially use the services of level 1 and later upgrade to level 2;

- The different levels of services for digital preservation would be associated with different patterns of costs and, therefore, highly flexible when it comes to decisions about what is reasonable, taking into account the financial resources at hand.

Close to the 'kiosk-model' is an approach called 'microservices' presented just a few years ago. It represents a step away from integrated digital archive systems and is, therefore, under discussion in the DCH community. The key idea with 'microservices' is that they allow flexible combinations of specialised solutions for preservation depending on the requirements of a DCH institution. As an example, 'Microservices' for digital preservation are currently used in the open archival information system Archivematica.[24]

[24] http://archivematica.org/wiki/index.php?title=Development_roadmap

III: OBJECTS TO ADDRESS

As discussed earlier in this handbook, preservation is a complex activity. This is not only because of the increasing complexity of digital objects and their growing number; it is also because the contexts of active user need to be re-created, which means sustaining not only the data, but also any specific software which was used to work with it, and the technological infrastructure.

The gradual expansion of preservation towards various types of objects is presented in the following figure:

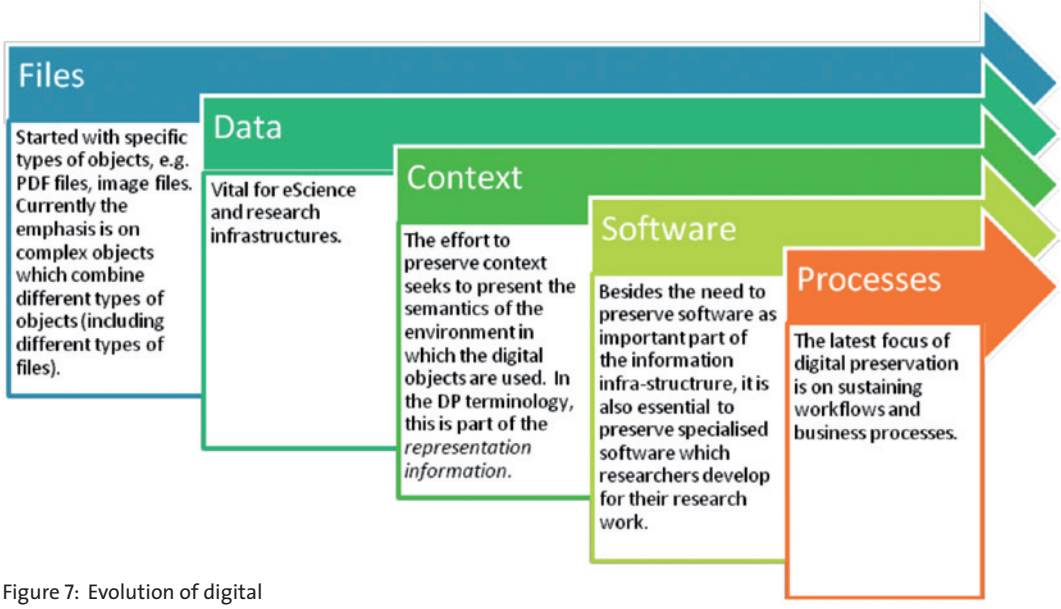


Figure 7: Evolution of digital objects addressed by digital preservation

Source: Digital Preservation Services: State of the Art Analysis by Raivo Ruusalepp and Milena Dobreva (report for the DC-NET project, available at http://www.dc-net.eu

All these different types of digital objects are relevant for digital preservation within cultural heritage institutions as well as in Humanities and Arts research. Although in many cases the emphasis is on the preservation of computer files, it is important to analyse the need to preserve software, the context of digital objects necessary for their future use, and any processes which also need to be preserved.

IV. TYPE OF SERVICE ARCHITECTURE

As mentioned above, the OAIS reference model provides the basic archiving workflow, but it does not articulate clearly how distributed archiving architectures can or should be arranged. E-Infrastructure service architectures vary significantly and do not allow for a uniform mapping of preservation tools and services to a single architectural model. Conceptualising and modelling of joint service architecture have been undertaken by only a few recent initiatives, and remain in a developmental phase.

The EUDAT project presented the architecture of a conceptual model that integrates various infrastructures with vast amounts of research data, and adds services for curation and trust in addition to the interface to users. This architecture illustrates a process that will have to be accommodated in the future by most preservation work, where solutions for preservation and curation can be used to support multiple different infrastructures.

As it stands, this model represents basic stakeholder needs in the research area: ensure the trustworthiness of data, provide for its curation, and permit an easy interchange among the generators and users of data. These could also be said to be basic needs in the cultural heritage community, and the EUDAT project's conceptual model can, therefore, serve as a basis for further development in the cultural heritage sector.

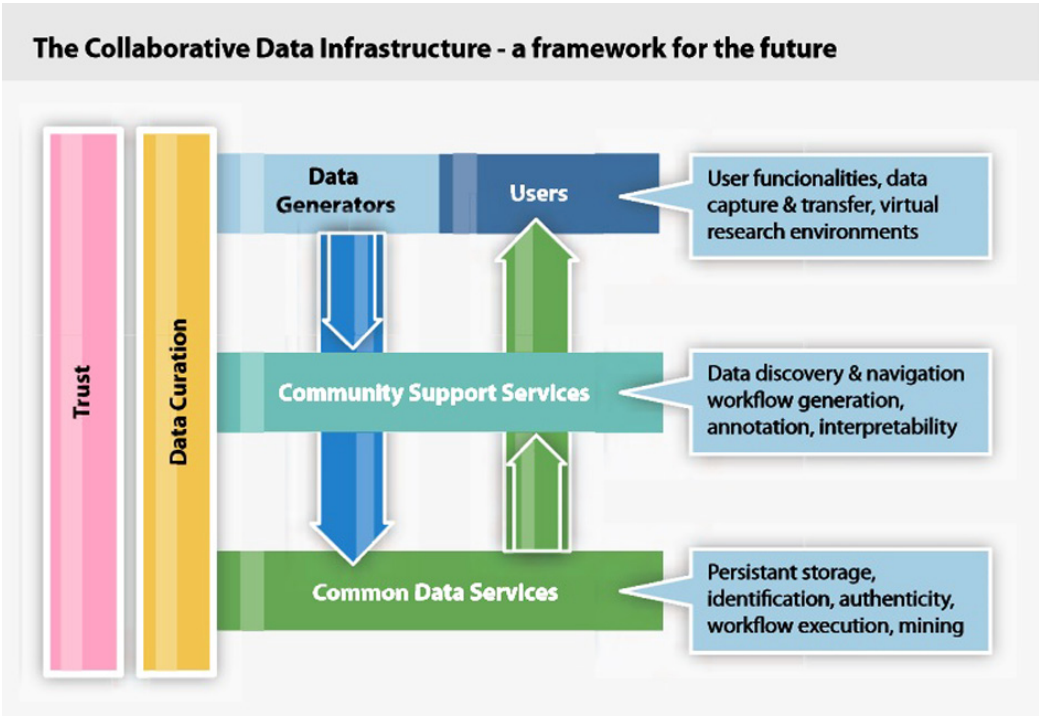


Figure 8: The collaborative data infrastructure - a framework for the future; from Riding the Wave, p. 31

Improvements and adjustments of the model have already been made in, for example, the area of research data. The Data Archiving and Networking Services (DANS) in the Netherlands has developed from the EUDAT conceptual model a federated data infrastructure with three layers of roles and responsibilities for the various stakeholders (The Front office – Back office model). [25]

[25] See www.dans.knaw.nl

Since preservation is part of the lifecycle of digital objects, it has implications for the processes and professionals within the institutions. The organisational structure of cultural heritage institutions varies and understanding their specific requirements from the distributed preservation infrastructure is a challenge that is not so easy to handle. It is sometimes argued by DCH institutions that the uniqueness of their digital holdings requires tailor-made approaches. A comparison of digital preservation provision across major European national libraries and the German Computer Game museum, made some years ago, showed significant differences in the type of holdings which need to be preserved, collection policies, preservation systems and standards used.[26]

It is undoubtedly true that continuing investment in in-house preservation systems will contribute to the lack of interoperability and fragmentation of resources into "digital silos". Stand-alone solutions that are not transferrable and interchangeable lead to fragmentation and do not offer economies of scale. Instead, shared solutions for the creation, storage and use of digital resources, including the e-Infrastructures, will become the major component of the future knowledge economy.

In order to move ahead from the current state toward shared, decentralised solutions, it is important to define key institutional requirements in a standardised way. The use of enterprise architecture models is one possible approach because enterprise architectures seek to address system complexity while aligning technological developments with the institutional needs. There are a number of approaches for defining enterprise architectures; one of the popular ones is the Open Group Architectural Framework (TOGAF)[27] and its eight-stage Architecture Development Method that helps to manage requirements within complex systems.

[26] The National Library of France is developing its in-house preservation system SPAR, which is OAIS-compliant and based on the use of METS and PREMIS-compliant metadata; The Royal Library of the Netherlands uses the e-Depot system which is based on the IBM DIAS and uses extended Dublin Core bibliographic metadata; The German National Library deployed a combination of tools including kopal-DIAS, koLibRI and has developed its own preservation metadata format, LMER (KEEP, 2009, 54 59;. *Preliminary document analysing and summarizing metadata standards and issues across Europe* (KEEP project deliverable D3.1). Available: http://www.keep-project.eu/ezpub2/index.php?/eng/Products-Results/Public-deliverables

[27] http://www.opengroup.org/togaf/

An earlier framework that looks at the various roles within an organisation and helps to summarise perspectives of various stakeholders on basic modalities of the organisation is the Zachman framework.[28] An adaptation of the Zachman Framework for the digital preservation domain has been carried out by Raivo Ruusalepp and Milena Dobreva in a report conducted for the DC-NET project.[29]

Service architecture as a technical area is very close to service-oriented architecture (SOA), which is a software design and software architecture design pattern based on pieces of software that provide functionality as a service that is easy to combine into different kinds of applications. Services mean in this case not services for the users but services in terms of written functions, ready to be used by programmers and by other applications. SOA can be seen in a continuum: from older concepts of distributed computing and modular programming, on to current practices of mashups, SaaS, and cloud computing, which some see as the offspring of SOA. In the DCH-RP roadmap, aiming at the use of e-Infrastructure, SOA can clearly be regarded as a concept from which to derive inspiration.

V. LEVEL OF MATURITY

Tessella[30] has described in a so-called Maturity Model how durable storage, information management and preservation provide increased levels of sophistication aiming at a complete digital preservation strategy.

The term Maturity Model is used to imply layers of sophistication in processes. The first layer must be complete before graduating to the next. For example, in digital preservation, there is no point having a good information management system if you do not have secure storage.

[28] Zachman, J. Concise Definition of The Zachman Framework. http://zachman.com/about-the-zachman-framework

[29] See Digital Preservation Services: State of the Art Analysis by Raivo Ruusalepp and Milena Dobreva (report for the DC-NET project, avilable at http://www.dc-net.eu)

[30] Preservica – white paper (July 2013) http://preservica.com/resource/praesent-ante-stiam-white-paper/

The Digital Archiving Maturity Model has three main parts:

Durable Storage (layers 1-3 in the Model) provides increasing levels of safety and security in the storage of the raw bits used to hold information. A level 3 compliant system implies you can be confident that your information will not be lost and that it has not been manipulated.

Information Management (layers 4-5) ensures that the preserved raw bits are organised. These layers have a hierarchy, descriptive metadata, and security, and they have a set of powerful tools to allow upload, management, search, browse and download.

Information Preservation (layer 6) is critical for information that must be retained for more than the lifetime of the application that created it. It ensures the file formats in which the information is held remain relevant to the applications available at the time the information is required, thus enabling it to be used immediately.

A simple storage archive would fulfil durable storage (layers 1-3) but no more, and a content management archive the information management parts (levels 4 – 5). A specialist digital preservation platform would fulfil all 6 layers.[31]

[31] See also Safety Deposit Box (http://www.digital-preservation.com/sdb) and Preservica Preservation as a Service (http://www.preservica.com)

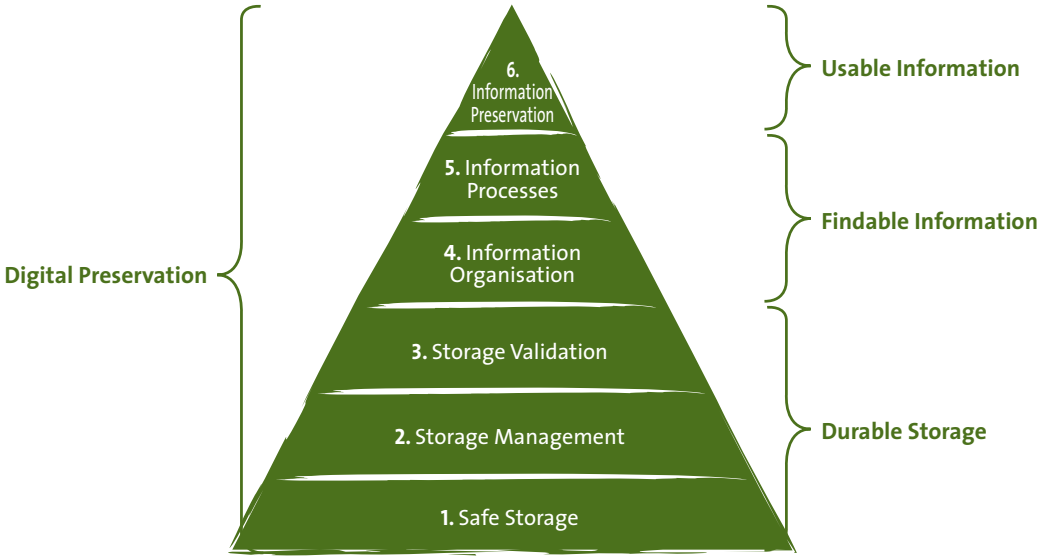


Figure 9: The Digital Archiving Maturity Model

Standards and licenses

The extensive use of relevant and open standards is a vital pre-requisite for the cultural heritage community when promoting interoperability, encouraging widespread access and controlling costs in its digital preservation programmes, regardless of whether they are built on in-house or distributed solutions. Extensive reviews under the auspices of earlier EU financed projects like Minerva (2008), Athena (2009) and Linked Heritage (2011) have already categorized and described many of the standards that are most applicable or recommended in the area of the DCH-RP project. However, moving into the field of distributed digital preservation services makes it increasingly important to understand and communicate the license

agreements and terms of usage associated with digital resources, whether these are 'born digital' or are digitised representations of other cultural heritage artefacts[32]. The Linked Heritage project reported seven overall license types relevant here and broke them out further, for example describing at least four variations of the Creative Commons (CC) licenses in routine use. Worth mentioning is also a highly structured method for license expression, namely ONIX-PL. This is not a license in itself but rather a machine-readable framework for conveying licensing and usage terms, conditions and prohibitions.

[32] The Linked Heritage project investigated this topic www.linked-heritage.org. The results are summarised in DCH-RP deliverable D3.2 Standards and Interoperability *Best Practices Report* which presents (and refers to) standards, best practices, and identifiers of interest for the Digital Cultural Heritage sector; the deliverable is available in the web space for the roadmap (see chapter 7 below).

5.3 The main components of the roadmap

5.3.1 A vision

Distributed preservation solutions are becoming more and more common, but there is an apparent lack of basic concepts that the DCH community has agreed on for implementing distributed preservation solutions, such as architectural design or best practice. There is obviously no commonly agreed vision of distributed digital preservation architecture relying on e-Infrastructures. Such a vision is an important piece in the puzzle and, therefore, urgently needed.

The overall vision for the DCH-RP roadmap is to implement a federated infrastructure, dedicated to supporting the application of open science in the arts and humanities, which will make digital cultural heritage accessible and usable long-term. This will be done by exploiting and integrating what already exists and by creating only those parts that are not yet available. The key to success is to use existing e-Infrastructures for research and academia (including NRENs, NGIs and the newer data infrastructures) as an efficient channel for the delivery of advanced services also to the digital cultural heritage. Connecting these facilities to the DCH sector will also contribute to

developing the research capacities of this sector. This is simplified by the fact that DCH data and scientific data have overlapping layers of information and therefore can be expected to have rather low barriers for sharing common services.

5.3.2 Timeframe

The DCH-RP roadmap should make it possible for each cultural heritage institution to define its own practical action plan with a realistic timeframe for the implementation of its stages.

- Short-term (2014-2015)
  The short-term action plan is proposed in order to initiate the development of a preservation services infrastructure on a level that will be self-sustainable and continue to progress on its own. This further progress is defined in terms of two further proposed time spans:
- Medium-term (2016-2017), i.e. two years after the end of DCH-RP, and
- Long-term (2018 and beyond) for the logical continuation of the DCH-RP work.

5.3.3 Appraisal and selection

In a digital environment, decisions taken when digital objects are created have significant implications for the preservation of those objects. The link between access and preservation is far more explicit than it is for paper and other carriers of information, as access to a digital object can be lost within a short period of time if actions are not taken right from the beginning to guarantee that it is preserved. A useful approach is, therefore, to incorporate appraisal and retention functionalities into the design of information systems applications.

While many of the principles from the traditional preservation environment can be applied in a useful way, policies and procedures will need to be adapted to the digital environment. In a paper-based

environment, the decisions to select for preservation, and to actually preserve, can be taken separately and within a timeframe which may span decades. The relatively brief lifetime of digital objects before they become inaccessible makes it necessary for decisions about selection and preservation to be taken simultaneously.

For networked digital resources, where access does not necessarily require bringing these resources physically into a collection or holding, appraisal and selection is quite different from that observed for traditional collections and holdings. Whereas in a non-networked environment acquiring a resource normally means keeping it, in the networked digital environment it is possible to provide access to a resource without undertaking any preservation commitment either short or long-term. Access is instead provided by making copies/mirrors for access in combination with hyperlinks, online catalogues, and other kinds of discovery aids. Appraisal and selection criteria are in this case often based on the number of requests and/or on cost-benefit considerations.

Employing evaluation criteria and selection procedures for all potential digital acquisitions ensures that collection development is carefully prioritised and sustainable. Such a review will normally be required for digital objects acquired before institutional policies and procedures were in place. One of the first steps that an institution undertakes in implementing a digital preservation policy may therefore be to quantify its current digital holdings and assessing preservation risks.

Over time the need may also arise to review collections and collections policy to reflect changing needs. The necessity of making early decisions on selection for preservation in a digital environment may result in a need for future reviews in the preservation lifecycle. However, for digital objects selected for long-term preservation such reviews should be conducted under strictly controlled circumstances.

5.3.4 Major areas on which to concentrate

The roadmap exercise as such aims to produce an instrument that will facilitate the work of policy makers as well as management activities within cultural heritage institutions. To achieve this, the roadmap concentrates on four areas which identify the policy domains that require intervention:

Harmonisation of data storage and preservation: would allow integrating in common environments the curation of research data with other digital objects – two domains which are currently addressed separately.

Improved interoperability: includes better integration of preservation within the overall workflows for digitisation and online access. In a way this is a set of measures to avoid building 'digital silos' within the organisation, for example when digitisation is carried out without taking into account needs for preservation, and/or when accessibility online is considered separately from preservation; an area of importance that needs to be integrated is the selection of what to preserve (see section 4.3.2 above)

Establishment of conditions for cross-sector integration: a key condition for maximising the efficiency of successful solutions, transferring knowledge and know-how.

Governance models for infrastructure integration: a necessary condition for successful institutional participation in larger e-Infrastructure initiatives, and aggregation and re-use of digital resources.

These four areas were selected in order to help consolidating experience gained in individual institutions and to merge it into useful knowledge for the cultural heritage sector as a whole. For each area a set of prioritised actions are suggested (see chapter 6.2).

5.3.5 A sustainability plan for the roadmap

The basic objective of a sustainability plan for the roadmap is to ensure that a clear direction is set. This means first of all providing the roadmap with an end point. But the roadmap also needs a structure that makes it possible to maintain it in some form until this end point has been reached. In section 5.3.4 above a timeframe has been set for the DCH-RP roadmap.

A second objective is the sustainability of the distributed services for digital preservation to which the roadmap leads. Each cultural heritage institution has to summarise steps taken in the short-term stage of the roadmap and transform them into a plan for how to implement the roadmap. We call such a plan "Preservation as a service" (PraaS). Depending on the conditions and considerations outlined in the roadmap, this plan (PraaS) can differ quite a lot in order to suit the needs of each individual institution.

The PraaS also has to take into consideration the sustainability (political, financial, organisational, and technological) of the preservation infrastructure services that are chosen. Identification of target groups is one of the key factors to ensure that these chosen services will become permanent and not only exist as 'mayflies'. The identified target groups could belong to different areas. They could be policymakers (EC, Member States, and Regions), owners of institutions, or e-Infrastructure providers; the important point is that on some level they interact or at least share the same view on digital preservation. Important questions to consider are, therefore:

- Which scenes of action (existing, or to be created) are there on which the identified target groups can interact?
- How can these target groups be motivated to participate in the implementation and continued support of the roadmap? What are their key interests?

- How are issues on digital preservation and access considered in each of the identified target groups? How are these two concepts interlinked in each group?

The answers will indicate:

- On what level access is needed (e.g. what is it politically and economically possible to request) by the cultural heritage sector and the e-Infrastructures serving this need;
- The amount of services in long term preservation that need to be offered by the e-Infrastructures, and what level of sustainability is required (e.g. what is it politically and economically possible to request) by the cultural heritage sector.

Another key factor is cost aspects that are themselves crucial for the sustainability of the roadmap. 'Who will pay?' is a question that will always be raised – sooner or later. In chapter 6.2.2 below the issue of business models is highlighted.

Social and cultural factors are expected to play a driving role in the implementation of the roadmap but also in sustaining it. The implementation as well as the sustainability of the roadmap is a complex process that will involve many different actors that may change over time, each with different vocations, stakeholders, knowledge and technical skills, and economical capabilities.

A sustainability plan for the roadmap has considered practical actions that should be taken to avoid those social and cultural factors becoming obstacles for the implementation and sustainability of both the roadmap itself but also for the outcome of the roadmap exercise.

5.4 A condensed version of the roadmap

5.4.1 Short-term (2014 – 2015)

This section presents a condensed version of the roadmap in a short-term perspective.

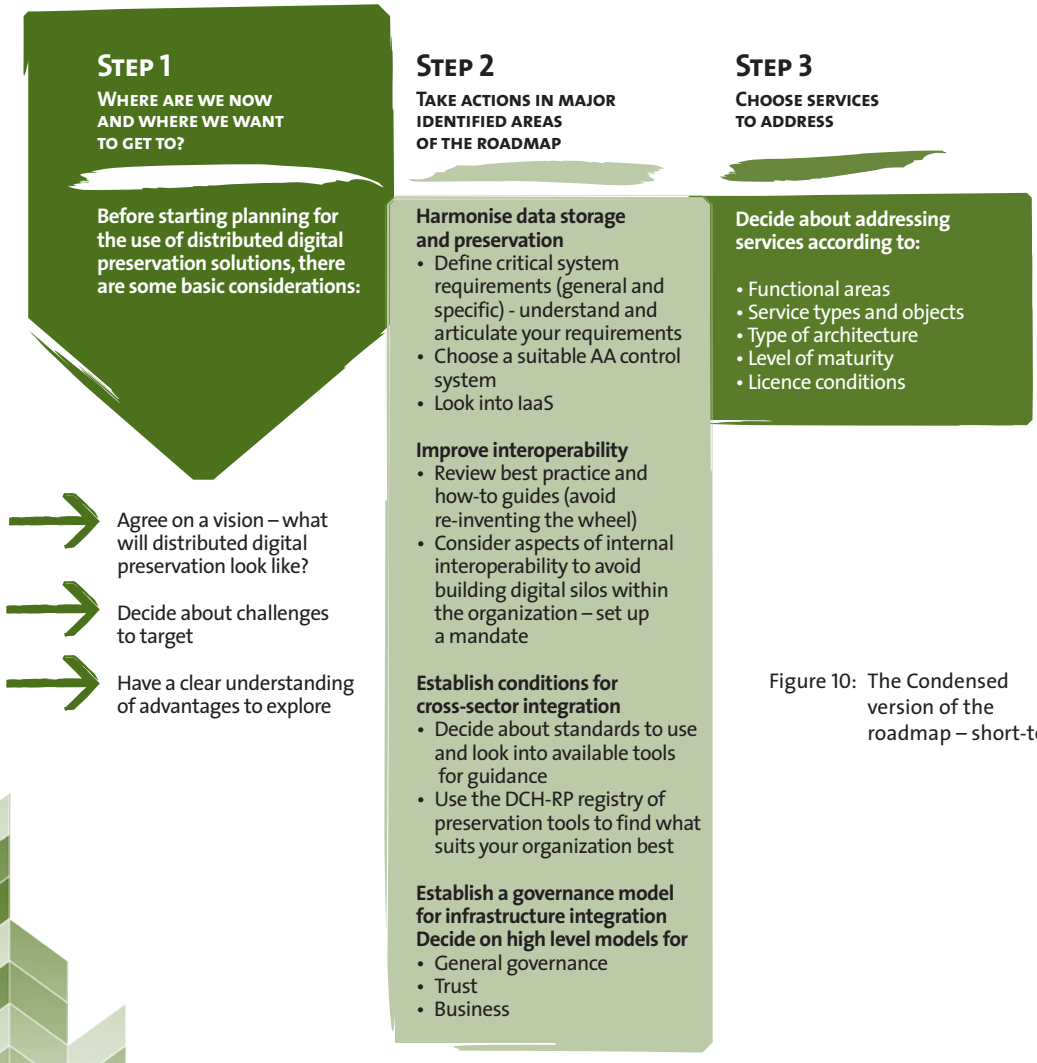


Figure 10: The Condensed version of the roadmap – short-term

When the short-term steps are completed, the following will be achieved:

- A clear view of both the actual situation and the main goals for using distributed digital preservation services;
- An initial set of requirements, functional as well as technicaand administrative;
- An idea about services to address (types, objects, architecture etc).

5.4.2 Medium-term (2016 – 2017)

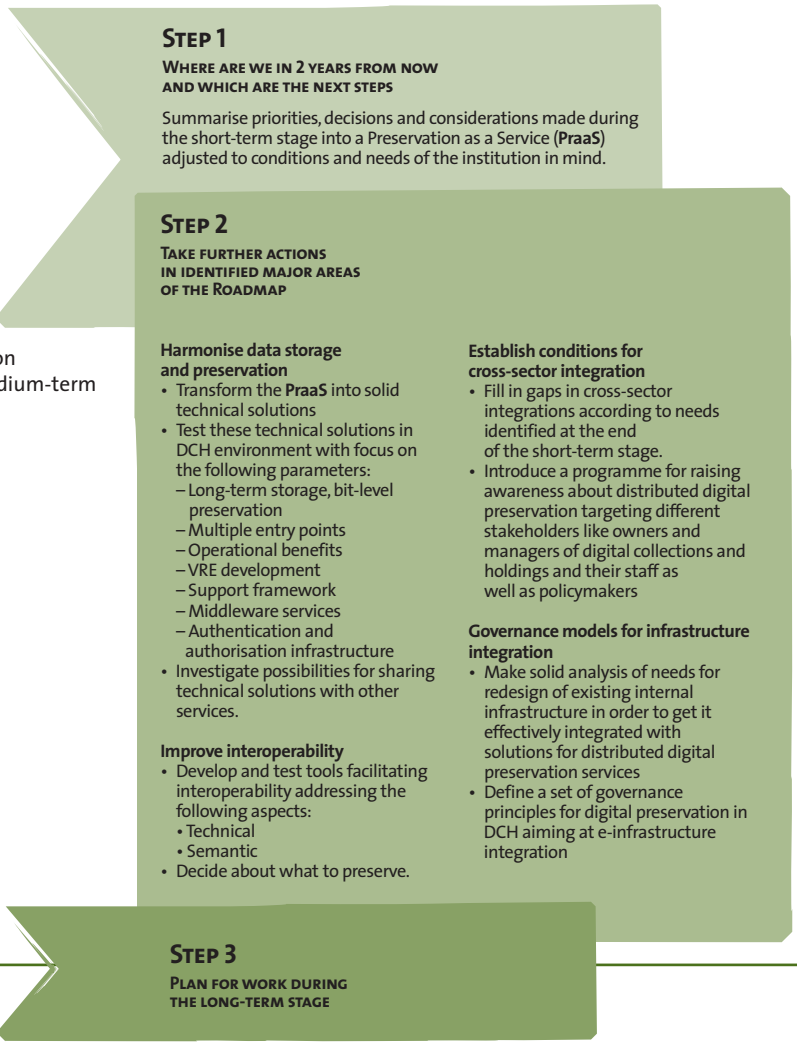


Figure 11: The Condensed version of the roadmap – medium-term

When the medium-term steps are completed, the following will be achieved:

- A sustainability plan – a PraaS – and tests of technical solutions; some e-Infrastructures should also have been identified as designated for distributed digital preservation;
- Decisions about digital resources for which distributed digital preservation services can be applied;
- Preparation of the internal organisation through awareness programmes, adaptation of internal infrastructure, and decisions on governance principles.

5.4.3 Long-term (2018 and beyond)

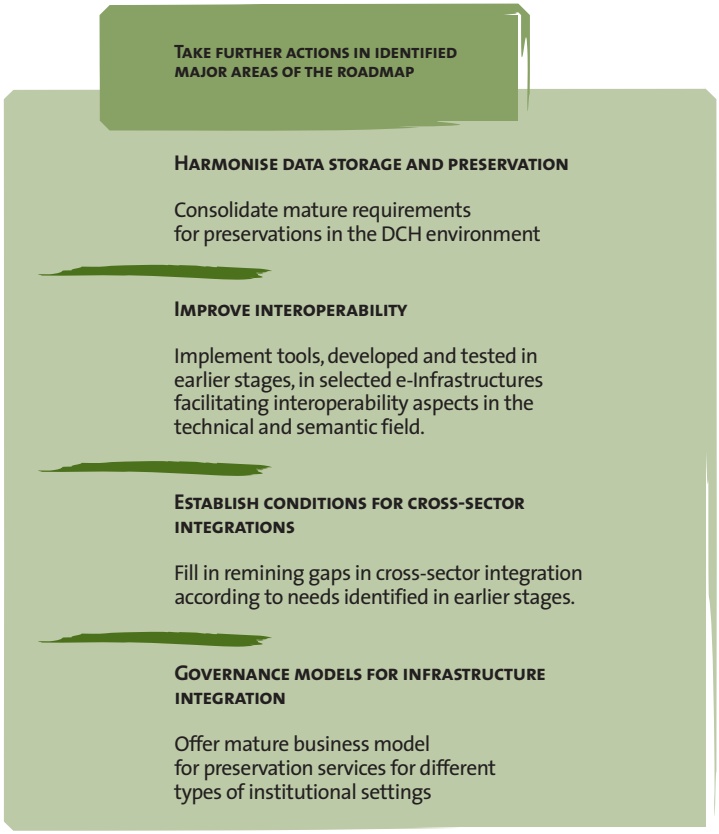
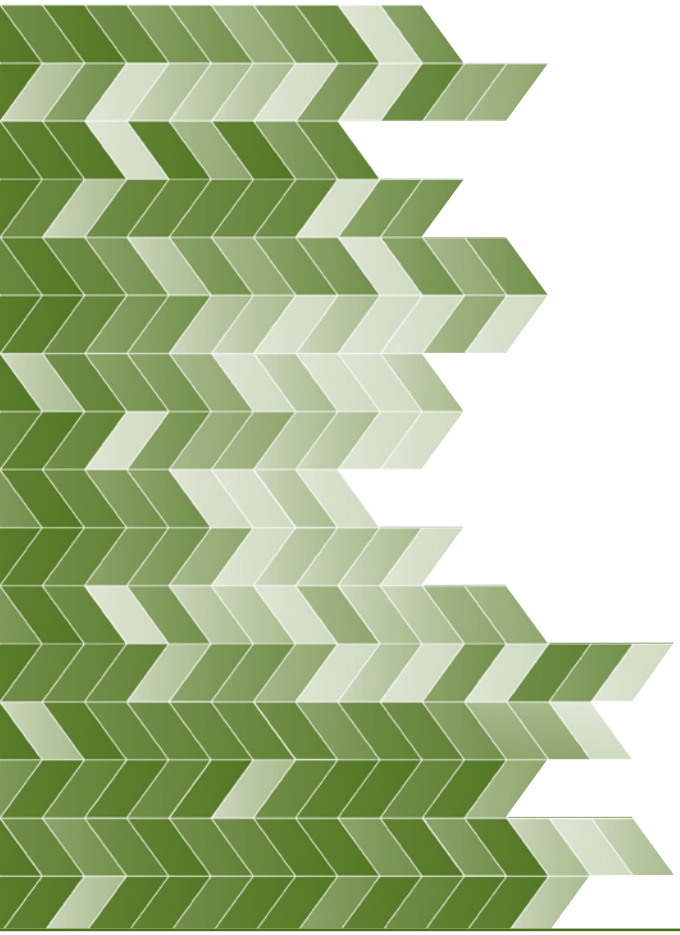


Figure 12: The Condensed version of the roadmap – long-term

By the end of this stage, selected e-Infrastructures tools and services that have been developed and tested in earlier stages will be implemented.



6 > AN ACTION PLAN

6.1 Establish a value chain

The DCH-RP project has looked into other domains, to see if there are experiences concerning value in distributed digital preservation that are transferrable to the DCH domain. Apparently, very little has been done so far, but in the e-journal preservation community much has been achieved in terms of evolving mechanisms and organisations to look after digital preservation. The technical, organisational and financial challenges have been proved to be solvable, given strong commitment from the communities involved. The key issue appears to be the ways in which these communities have organised themselves to bring about long-term agreements and infrastructures to make preservation happen.

Cost will clearly be a key variable when deciding whether or not to contract out digital preservation to an external service provider. But there are also other factors to consider, and the advantages and disadvantages of each of them need to be balanced against the overall mission of the institution. For example, legal provisions regarding privacy or confidentiality may influence whether outsourcing is appropriate or not. The extent to which the potential advantages of using distributed preservation services can be maximised and the potential disadvantages minimised is also dependent on the possibilities for dedicating staff resources to preservation activities. The costs for these staff resources need to be added to the overall contract costs when calculating the cost benefit of using distributed services for digital preservation. However, one should be aware of that most of these costs will be or should be incurred even when digital preservation is not outsourced.

The Digital Preservation Coalition has listed a number of issues and potential advantages and disadvantages of using distributed services in digital preservation activities.[33]

[33] See Preservation Management of Digital Materials: The Handbook http://www.dpconline.org/advice/preservationhandbook

Issue	Potential advantage of using 3rd party services	Potential disadvantage of using 3rd party services
Limited practical experience in preserving complex digital objects over time	Avoids the need to develop costly infrastructure (particularly important for small institutions). Allows the institution to focus on other aspects of service provision. Provides specialist skills and experience which may not be available within the institution. If there are economies of scale, outsourcing may well be cost effective. Allows action to be taken in the short to medium term, pending development of infrastructure.	Without some practical experience and expertise, it will be difficult to develop and monitor effective contracts. Without practical experience it will also be difficult effectively to communicate the requirements of the organisation (or to assess whether they are technically feasible or not). Danger of either not developing or losing skills base. There is no established bench marking. It is still too new an area. Risk of business failure. Until the market increases there may be an overdependence on one contractor. Unless there are adequate exit strategies, may be locked into an outsourcing contract longer than intended.
Access considerations	Monitoring usage may be more efficient (assuming the contractor has a demonstrated ability to deliver meaningful usage statistics). There may be synergies and cost savings in outsourcing access and preservation together.	Difficult to control response times which may be unacceptably low and/or more costly, especially for high-use items.
Rights Management	Avoids what is often a resource intensive activity for the institution.	May significantly increase the cost of the contract and/or complicate negotiations with rights holders.
Security	Contract can guarantee security arrangements required by the institution.	Lack of control, especially for sensitive material.
Quality control	A watertight contract will build in stringent quality control requirements.	Risk of loss or distortion may still be unacceptably high for highly significant and/or sensitive material.

Major advantages, specific to the DCH sector when using distributed services offered by e-Infrastructures, could for example include the following:

- Long-term preservation (i.e., bit-level preservation) and access to digital objects of different kinds, also so called "live" content (e.g., streaming audio and video collections);
- Multiple entry-points that suit a variety of user interfaces (e.g. APIs, protocols). New cloud based search engines are under development, based on multilevel nodes that can combine different data sources (documents, images, books etc) from multiple content providers;
- The DCH-community can focus on its own areas of specialisation by deploying new services for monitoring and management tools that ensure smooth and secure running of distributed operations;
- Forming a community of best practice or a Virtual Research Community that transcends discipline and national boundaries while achieving economies of scale by bringing together international communities;
- Benefitting from integration within the research and educational e-Infrastructures support framework;
- Central hosting and monitoring of middleware services;
- Simple authentication and authorisation infrastructures for large (and potentially unbounded) user groups;
- Connections to shared services in other countries and sectors (e.g. research data centres, commercial businesses, etc.).

To summarise: it is important for cultural heritage institutions to have a clear understanding of what to exploit, before taking a decision about the use of distributed digital preservation services.

Research and development on the use of digital preservation services built on distributed facilities instead of ones performed in-house has just started.[34] Some identified drivers that probably will underpin an enhanced use of distributed digital preservation services are:

- Increased flexibility in digital preservation architectures based on granular or layered structures (e.g. SaaS, PaaS, IaaS) that are easy to adapt to a variety of preservation scenarios;
- Clearly defined sets of metrics or benchmarks for comparing preservation tools and services and their performance;
- Terminology and standards that no longer reside within professional community borderlines but instead are agreed across sectors.

[34] An example is the InterPARES Trust (ITrust 2013-2018), a multi-national, interdisciplinary research project exploring issues concerning digital records and data entrusted to the Internet (http://interparestrust.org)

6.2 Actions to take

6.2.1 Harmonise data storage and preservation

SHORT TERM PRIORITIES

Today, an ever-broadening range of preservation software tools is available, and institutions can combine and tailor digital preservation components according to their specific needs and context. The typical digital preservation workflow incorporates generic tools, e.g. virus checking, metadata generators or format identifiers, specific preservation services, as well as services that relate to storage management in distributed preservation environments. The aim here is to establish the necessary conditions for various services to coexist and to be orchestrated into a suitable digital preservation "eco-system", regardless of whether the services are targeted on research data or other digital objects.

I. DEFINE AN INITIAL SET OF CRITICAL SYSTEM REQUIREMENTS

General needs and requirements in a digital preservation context
Examples (listed regardless of priority):

Miscellaneous issues

- Reliability and robustness
- Assurance of valid licensing procedures, commercial conditions, and transactions
- Open, scalable, and flexible solutions (built on open industry standards like J2EE and XML)
- Ease of use (for example, user-friendly interfaces)
- OAIS compliance
- Multilingualism

Content/information issues and metadata issues

- Mechanisms for integration and automation of appraisal and ingestion of digital material
- Automatic metadata capture and extraction
- Separation of content (information) and metadata
- Various content formats (from print-based documents to digitized images)
- Ontologies for both visual and textual concepts
- Annotation services

Performance issues

- Scalability (up to hundred terabytes or more)
- Performance for hundreds of thousands of electronic documents

Trust issues and security issues

- Authenticity and integrity of data
- Continuity (which means the handling of information, both data and metadata, for at least the next 100 years)
- Identification of digital objects which are in danger of becoming inaccessible due to changes in technology

- Security during transmissions of files between countries
- Validation (certification) of software and hardware environments required to render the digital objects

Infrastructure-related issues
- Distributed systems
- Virtualisation

Hardware-related issues
- Support of many storage media and devices
- Backup and restore

SPECIFIC REQUIREMENTS

Need for simplicity

Integrating preservation workflows with e-Infrastructures normally requires significant levels of computing and IT expertise, not always available in cultural heritage institutions. The solutions developed need, therefore, to be tested for their simplicity of installation, management and use.

Metadata

The metadata connected to a digital object is crucial for the possibilities of preserving it for future use. It has to include basic descriptive information about the file as well as information about the file format of the object. The metadata collected about a digital object helps to place it in context, as well as give specific information, which is essential for making sure the object in mind is authentic (hasn't been added to or modified in any way). This is especially important for digital files, which in contrast to print media can be easily changed in ways that may not be readily apparent. Metadata can be linked to the digital object or encapsulated with the digital object itself. Encapsulating the metadata with the object ensures that the information stays with the file, no matter where it goes. Linking the metadata but storing it separately ensures that the information about the file can be

recovered even if the object itself is lost. Depending on the actual situation, a decision about metadata has to taken before a cultural heritage institution enters into distributed digital preservation.

Storage in different locations
Archival data (master files) can often be stored offline, since they are infrequently accessed. It is best practice in many cultural heritage institutions to write digital archival data to more than one type of media and then store these in different locations.

Digital resources in continual use (surrogate delivery files) will typically be stored online. Online storage is often mirrored across multiple disks using redundant disk arrays (RAID).
Today clustered (data centre) and distributed storage systems are normally used for distributed storage. A storage cluster consists of at least two independent storage nodes, running under the control of relevant software. When one of the nodes fails, the other immediately takes over all of its duties.

A data centre is a facility housing computer systems and associated components like telecommunications and storage systems.
It generally includes services such as redundant or backup power supplies, redundant data communications connections, environmental controls (e.g., air conditioning, fire suppression) and security devices. The concept Dynamic Infrastructure is a design approach for data centres making it possible for the underlying hardware and software to respond dynamically to changing levels of demand in more fundamental and efficient ways. This concept is also known as *Infrastructure 2.0* and *Next Generation Data Centre*.

Cloud storage is often implemented with complex, multi-layered distributed systems built on top of clusters of servers and disk drives. Sophisticated management, load balancing and recovery techniques are needed to achieve high performance and availability. While there is a relative wealth of failure studies of

individual components of storage systems, such as disk drives, relatively little can be found reported, so far, on the overall availability behaviour of large cloud-based storage services. Special care has therefore to be devoted to this issue before entering into a solution based on distributed preservation.

Migration of data and metadata
A routine error-checking schedule should be implemented and a strategy drawn up for migrating data and metadata to suitable formats as necessary. If a file format is becoming obsolete and a migration is planned, archival master files should be migrated to new formats that are non-proprietary. Quality control checks should follow any migration or refreshment so that any loss of data integrity can be identified and quickly addressed.

II. NEEDS AND CONDITIONS FOR INFRASTRUCTURE FEDERATION

The need to access networked applications and remote/distributed data is evolving dramatically. Authentication and authorisation are often separated from the application and the data themselves: authentication of the users is done by the user's Identity Providers while the authorisation is done by the services based on the information received by the Identity Providers.

Access that follows this model is known as federated access and has advantages for both users and application developers. However, the usage of federated access requires that some technical and trust issues have to be solved.

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

that offers the service. In line with the objectives of the DCH-RP project, the ambition is not to establish a separate authentication and authorisation (AA) infrastructure for the DCH service and user community, but to use the most suitable AA services available in the research and education community.

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

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

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

The eCulture Science Gateway of INFN (Instituto Nazionale di Fisica Nucleare), is based on federation identities. eCulture Science Gateway was developed within the framework of the earlier INDICATE project. It was upgraded with new functions by the Italian DCH-RP partner INFN and used for the DCH-RP projects Proofs of Concept.

When joining a federation, the following entities will be encountered:

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

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

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

More information about federated access is available in the documentation available online in the web space for the roadmap (see chapter 7 below).

III. ONGOING EXPERIENCE WITH GRIDS AND CLOUD SOLUTIONS APPLIED IN CULTURAL HERITAGE INSTITUTIONS

One of the basic assumptions for the DCH-RP project is that grid and cloud approaches can offer a stable and reliable storage and computing platform to the digital cultural heritage community.

In general it seems that this community's first priority, when it comes to digital preservation activities, is storage. Other identified priorities are computer capacity for integrity checks and access to advanced virtualisation services. One conclusion is, therefore, that at least two main approaches to preservation services must be in place for distributed solutions. In chapter 5.2.2 they are referred to as the "kiosk" model and the "turn-key" model respectively. The approach that in the same section is called "micro services" could also be a fruitful approach to investigate. However, if various micro services are to be used, they must be orchestrated in a way that assures that requirements for the authenticity and integrity of preserved digital objects are not compromised.

When reviewing the limited experiences of distributed preservation of digital cultural heritage to date, the most striking observations are a feeling on the part of the e-Infrastructure side of frequent dissatisfaction on the users' behalf, and of users regularly reporting difficulties in utilising the facilities and tools offered. Therefore, a roadmap establishing future approaches and methods of preservation definitely has to put special emphasis on how to bring the e-Infrastructure closer to the users, how to make the e-Infrastructure providers more sensitive to user demands and, on the other hand, how users can better exploit the opportunities offered by the e-Infrastructure.

IV. EXAMPLES OF THE USE OF PLATFORM AS A SERVICE (PAAS) AND OF BENEFITS OFFERED BY VIRTUALISATION

Although a number of preservation tools are available, their uptake and use in practice is very hard to measure, and so is the whole market for digital preservation services. The models for evaluating market maturity are too general to fit easily a niche area like digital preservation. The Planets project conducted interviews with leading IT companies to explore the emerging market-place for digital preservation tools and services. Results of this study confirm that engagement is being led by memory

institutions and driven primarily by legislation. There is perceived high demand for technology to support automation of digital preservation processes and for consultancy, training, awareness-raising and exchange of best practice, but the overall description of the services market was as a "market in its infancy".[35]

In recent years some new distributed services in digital preservation have been introduced. One example is the Data Archiving and Networked Services (DANS). In the Netherlands a federated data infrastructure is developing with DANS as a trusted digital repository, in the first place for research data, performing back-office functions like expertise in data governance and long-term storage and accessibility.[36] Another example is Preservica, a cloud-based service to safeguard digital information. Preservica conforms to the OAIS model (ISO 14721:2003) and markets itself as providing all the tools required for building a long term digital preservation solution.[37]

Tests of existing technical solutions in a DCH environment have been carried out by the DCH-RP project. The results achieved are reported in deliverables D5.3 and D5.4, which are available in the web space for the roadmap (see chapter 7 below).

MEDIUM TERM PRIORITIES

Recommendations, best practice and lessons learned from tests of existing technical solutions – executed during the phase of short-term priorities – have to be transformed into solid technical solutions aimed at the DCH environment. These solutions will then need to be tested more specifically, addressing aspects like these:

- Long-term storage (bit-level preservation)
- Multiple entry points
- Operational benefits
- VRE development
- Support framework
- Middleware services
- Authentication and authorisation infrastructure

[35] *An Emerging Market: Establishing Demand for Digital Preservation Tools and Services*. Available: http://www.planets-project.eu/docs/reports/Planets-VENDOR-White-Paperv4.pdf (PLANETS 2010)

[36] See www.dans.knaw.nl

[37] Preservica Preservation as a Service (http://www.preservica.com)

Possibilities for sharing technical solutions with other services will also need to be investigated.

LONG TERM PRIORITIES

The main priority in this stage is to consolidate mature requirements for distributed digital preservation in the DCH environment.

6.2.2 Improve interoperability

SHORT-TERM PRIORITIES

I. IDENTIFY AND PROMOTE BEST PRACTICES

A large part of the roadmap is dedicated to best practices, presenting overviews of the most important practical guidelines and lessons learned connected with the integration between the cultural heritage community and the e-Infrastructure providers. The second round of Proofs of Concept conducted in DCH-RP has been an important instrument for capturing best practices by interacting with the DCH community. The results are presented in DCH-RP deliverable D5.4, which is available in the web space for the roadmap (see chapter 7 below).

II. ANALYSE INTEROPERABILITY ISSUES

To avoid building 'digital silos' within the organisation, the following aspects need to be considered:

1. *Technical aspects*: a storage solution should be decided upon before producing any digital output, as it is of prime importance for the following steps in an organisation's digital preservation programme; strategies for both online and offline storage should be considered for the digital resources to be stored, otherwise storage of digitised resources runs the risk of competing with limited resources for maintaining the administration platform; due to the large size of master files, an entire digital collection can be very substantial in

size, possibly requiring a mixed architecture for data storage; the size of both master files and any surrogate files have implications for the amount of storage space required and should be calculated or estimated at the outset of the project.

2. *Semantic aspects*: there are many vocabulary sources already available and it makes sense to check these out before inventing a new one. Depending on its needs an organisation might:
   - Use an existing controlled vocabulary;
   - Adapt or customise a vocabulary in use;
   - Develop its own vocabulary (not recommended though sometimes unavoidable);
   - Use an "uncontrolled" vocabulary - i.e. keywords entered by the organisation's cataloguers or its users – should not be done under any circumstances as it makes interoperability impossible or very hard to achieve.

   Of course, it can be quite reasonable to use a combination of these approaches, for example a formal controlled vocabulary plus additional keywords to assist in retrieval.

   In choosing a vocabulary, it is important to keep in mind:
   - The end users - are the terms used going to be meaningful to them?
   - The community - it makes good sense to use vocabularies that similar collections are using.
   - The nature and extent of the collection - if the collection is small, it will probably not need a detailed vocabulary.
   - Copyright issues - it will sometimes be necessary to check whether permission or a license is required to use the vocabulary in the way the organisation wishes.

3. *Organisational and inter-community issues*: while it is clear that a technical strategy is necessary to ensure digital preservation, it is also important that digital preservation is underpinned by organisational commitment.

4. *Legal issues*: the transfer of personal data has to be in line with European directives on data protection and their implementation in national legislation; harmonisation of legal frameworks in general have also to be addressed, for example concerning the issue of cross border storage and differences in legal positions regarding preservation of master files.

5. *Political/human aspects*: digital preservation is an active task, and it is imperative that the responsibility for all digital resources is firmly assigned and known to all stakeholders - digitisation projects should have, as part of their project specifications, a policy which covers:
   - Who the digital resource or collection belongs to in the organisation and who is responsible for its upkeep;
   - What the process is for deciding when and how refreshment/ migration takes place and who makes the decision;
   - Where the budget is coming from for this ongoing digital preservation investment.

MEDIUM TERM PRIORITIES

Improved interoperability is an area of action that focuses mainly on the DCH institutions' own internal conditions (see above under Short-term priorities). It is important during this stage to develop and test tools that facilitate this internal interoperability, addressing both technical and semantic aspects.

During this stage the question of what to preserve has to be raised. An important issue, called for by several stakeholders, is that analogue data carriers converted into digital ones by digitisation and "born-digital" objects have to be treated differently; mostly because digitised objects may be needed during a shorter period and sometimes are considered just as digital copies of physical objects.

One way of starting the process of appraisal is to try to get an overview of the basic conditions for preserving different media types in digital form. A first attempt is made in figure 13 below.

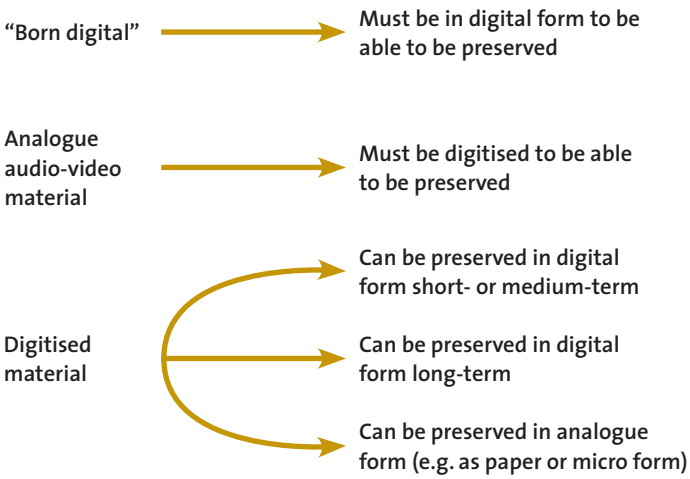


Figure 13: Basic conditions for preservation of different media types in digital form

When applied to the actual content in the collections and holdings kept by a cultural heritage institution, the results can be used in a matrix where the other axis shows for example the key factors noted in the sustainability plan (interests and intentions of different target groups, cost aspects and social and cultural factors, etc., see chapter 5.3.4 above).

LONG TERM PRIORITIES

By now, some e-Infrastructures should have been identified as designated for distributed digital preservation The main priority in this stage is, therefore, to implement in selected e-Infrastructures tools and services that have been developed and tested to facilitate interoperability aspects, in the technical as well as in the semantic field.

6.2.3 Establish conditions for cross-sector integration

SHORT TERM PRIORITIES

I. ANALYSE WHAT IMPACT EMERGING AND ESTABLISHED STANDARDS HAVE ON GRID AND CLOUD PRESERVATION ARCHITECTURES

The DCH-RP project has reviewed existing projects and initiatives as well as standards, guides, and tools, which are potentially of use for the DCH and e-Infrastructures communities when approaching digital preservation issues. One of the challenges for the DCH community is to choose among the vast number of standards that are already available. This may be problematic, especially for small DCH institutions with limited knowledge or resources in this field. There are also non-technical issues that have to be resolved. One is differences in the legal systems between countries, especially when data is covered by copyright or is classified.

The conclusion is that much work has already been done, but that more effort is still needed before these standards, guides and tools, etc can be of real help to the DCH community. For example, many of them need to be more user-friendly in order to be understandable for non-technical personnel. Furthermore, practical tests made within the DCH-RP project have shown that already developed e-Infrastructures must be modified and/or improved in order to provide a "pan-European" solution for the DCH community.

II. REGISTRY OF PRESERVATION TOOLS AND SERVICES

The development of the DCH-RP preservation services registry is a key step in the construction of the roadmap. In this regard, it should be noted that the collection and summarisation of information on services is quite an onerous task because over the last decade the number of tools and services produced within the community has been quite impressive. However, more work needs to be done on the characterisation of services in order to make

them usable in a distributed e-Infrastructure and currently there are no testing tools which would help to run systematic evaluation on the behaviour of tools – either singly or in combination.

There are a few hundred software tools on offer to support automation of preservation tasks, yet their support status, interoperability status, level of documentation, quality, and reliability are poorly documented. There continues to be inadequate support for decision-making, selecting, testing and benchmarking tools for preservation. While a number of digital preservation tools registries/collections are already in place, there is no such collection addressing grid and cloud services. The DCH-RP project's Registry of services fills this gap by presenting a registry of the services available to support preservation activities, with particular regard to the services that can better fit the requirements of the DCH sector. This deliverable is public and available on the project's homepage www.dch-rp.eu. Through integration of the work of the EU project APARSEN into the DCH-RP Registry of Services and discussions with other initiatives (e.g. COPTR recommended by the Digital Preservation Coalition[38]) the sustainability of the registry will be accomplished (for more information about the Registry, see also chapter 2.3 above).

[38] For more information on the recommendations of COPTR, see http://www.dpconline.org/advice/tools-coptr

III. ANALYSE WHICH MIX OF SERVICES BEST MATCHES DIGITAL PRESERVATION REQUIREMENTS

To find a suitable mixture of distributed services to match individual institutions' requirements for distributed digital preservation can be a challenge. The DCH-RP project has conducted a number of practical experiments in the framework of its Proofs of Concept that are meant to be used as best practice. The results from these experiments, together with some

general recommendations, are available in the web space of the roadmap (see chapter 7 below).

IV. IDENTIFY GAPS IN PROVISION AND ESTABLISH A PLAN FOR MEDIUM- AND LONG-TERM DEVELOPMENTS TO ADDRESS THE GAPS

A plan for medium- and long-term work to address identified gaps needs to be assembled at the end of the short-term stage.

MEDIUM TERM PRIORITIES

The main challenge during this stage will be to fill in gaps in cross-sector integration according to a plan made at the end of the short-term stage.

As a prelude to the actual implementation of the roadmap by memory institutions, it will be necessary to plan for an initial training/awareness phase, where the cultural heritage managers, often coming from a professional background developed in a more static environment, need to familiarise themselves with the concept of a roadmap. Management staff should improve their awareness about the need to conceive the cultural institution and its data as a living body, which should move along a 'road', to arrive at new destinations. The terminology used in the roadmap presents some elements of discontinuity with the traditional terminology that cultural heritage managers may be familiar with, which need to be explained in order to maximise the benefits from the adoption of the roadmap.

LONG TERM PRIORITIES

The main challenge during this stage will be to fill in gaps in cross-sector integration according to a plan already outlined by the end of the short-term stage.

6.2.4 Establish a governance model for infrastructure integration

SHORT TERM PRIORITIES

I. ANALYSE MAJOR INFORMATION GOVERNANCE PATTERNS AND WINDOWS OF OPPORTUNITY

The governance model decided upon must be tailored to the concept of distributed digital preservation. The following framework is suggested as an outline for how to achieve good governance:

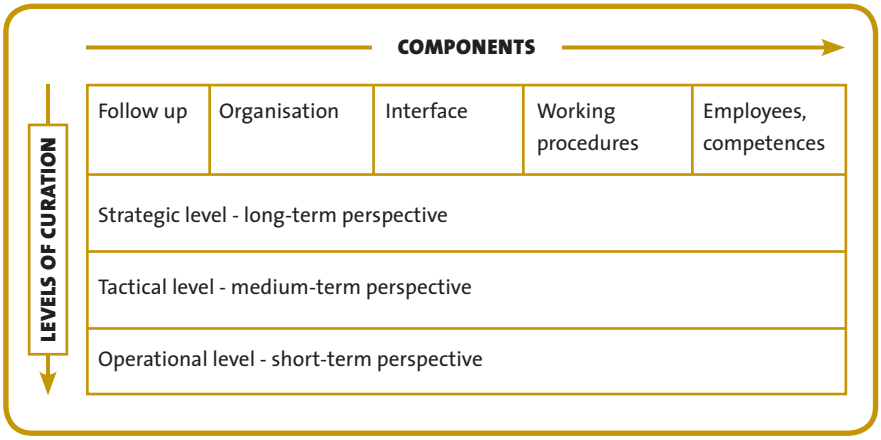


Figure 14: A framework for the governance of distributed digital preservation services

This framework consists of five components that highlight different dimensions of governance focusing on three different levels (strategic, tactical and operational). The components are:

- Follow up (including how to manage distributed digital preservation services);
- Organisation (including definitions of roles and responsibilities);
- Interface (including a forum for clients and service providers to meet);
- Working procedures;
- Employees and competences.

The levels of governance each have different focus and perspectives:

- Strategic level: aiming at securing the long-term perspective; this is done from both an internal and an external perspective through, firstly, follow up and managing a consolidated service provider portfolio, and, secondly, establishing a forward-looking relationship between the client and the service-provider;
- Tactical level: has a middle-term perspective with focus on securing services and agreements in hand and ensuring that they are up to date;
- Operational level: focus is here on securing the follow up of the daily work and that problem and incidents that arise are handled in a proper way.

Depending on which type of service is involved, the service providers can be classified as being strategic/non-strategic and providing services that are easily accessible/not easily accessible. For the cultural heritage institutions the results of such a classification will inform their approach to managing the situation.

II. EXPLORE THE ISSUE OF TRUST-BUILDING

There is no trust model of a distributed repository system in place today in the DCH domain. The only similar example in existence is what can be decribed as a "circular chain trust model" of the LOCKSS system where all partners using the software also share a trust network. The CESSDA is working on a formalised trust model for their services based on this example.[39]

[39] See DCH-RP deliverable D3.1

Single repository level trust issues are being explored by several research projects (e.g. APARSEN) for example in the context of auditing of digital repositories. The underlying concept there is that the trustworthiness of a repository can be established through an audit. This notion is derived from the 2002 RLG/OCLC report "Trusted digital repositories: attributes and responsibilities". The European trusted digital repository

audit framework sets out a three-tier model for establishing trustworthy repositories through audits:

1. Self-assessment, using the Data Seal of Approval (a toolkit developed by DANS for research data archives) or DRAMBORA;
2. Self-audit using ISO 16363 or DIN 31466 (both are originally based on the TRAC checklist that was developed by RLG and NARA);
3. Formal audit using ISO 16363 or DIN 31466 using external auditors and leading to certification.

In parallel with this initiative there is the Center for Research Libraries (CRL) in the US that still conducts TRAC audits and issues certificates to repositories and their cooperatives.[40]

[40] See http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories

Neither of these approaches is directly relevant to DCH-RP purposes, because NRENs will probably not be interested in undergoing a full digital repository audit. NRENs are for understandable reasons not very keen to become full-scale digital preservation repositories for DCH alone, because this is not really their core business. What is needed is a more flexible method of auditing of a distributed digital preservation service where a repository is outsourcing some of its services to an NREN. Such audit tools do not yet exist. In section 5.3.2 this approach to distributed preservation services is called the "kiosk" model.

There is one recent development that is highly relevant for the DCH-RP project. This is called the Distributed Digital Preservation reference model (DDP) that is trying to enhance the original OAIS model that best suits a single repository.[41] As part of the DDP model there are plans to develop a distributed trust model, but this work has not proceeded very far yet.

[41] See a guide: http://www.metaarchive.org/GDDP

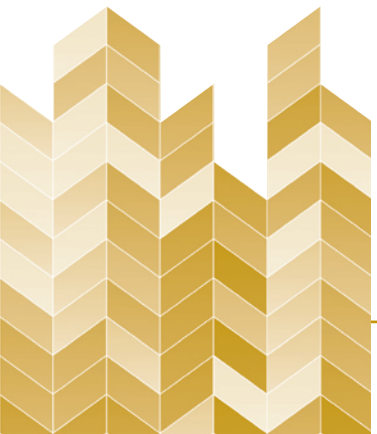
It is important to strengthen the capability of cultural heritage institutions to articulate their trust requirements. In deliverable D4.1 *Trust building report* the DCH-RP project

has outlined the design of a new trust model suitable for the use of e-Infrastructures, including recommendations for user authentication and access control system(s). The deliverable is available in the web space of the roadmap (see chapter 7 below).

III. ESTABLISH A POSSIBLE BUSINESS MODEL

A business model describes the rationale of how an organisation creates, delivers, and captures economic, social, cultural, or other forms of value. In both theory and practice, the term business model is used for a broad range of informal and formal descriptions to represent core aspects of a business, including purpose, target customers, offerings, strategies, infrastructure, organisational structures, trading practices, and operational processes and policies. There is also a clear connection between the business model used and trust-building. It is obvious that a business model based on passive preservation is not an option. While there is understandable concern that the costs of preserving digital materials will be high, it is equally important to consider the costs and implications of not preserving them. The costs of recreating a digital resource may be much higher than those for preserving it; further, the opportunity to do so may no longer exist when the digital resource concerned is needed. An increasing dependence on both digitally produced and accessed information means that there is a rapidly growing body of digital material for which there are legal, ethical, economic and/or cultural imperatives to retain the material, at least for a defined period of time and, in some cases, forever. If active steps are not taken to protect these digital materials, they will inevitably become inaccessible and unusable within a relatively brief timeframe.

Digital preservation built on a distributed model also needs a business model for the integration between the cultural heritage community and the e-Infrastructures. ICT managements have today started to implement new concepts for outsourcing, whether cloud-based or not. One of them is Vested Outsourcing.



This is a hybrid business model, based on research conducted by the University of Tennessee Center for Executive Education and funded by the U.S. Air Force, In this model both clients and service providers in an outsourcing or business relationship focus on shared values and goals to create an arrangement that is mutually beneficial to each, in contrast to traditional outsourcing and businesses relationships that, according to Vested Outsourcing, focus on win-lose arrangements.[42]

[42] http://en.wikipedia.org/wiki/Vested_outsourcing

The basic philosophy in the Vested model is "What´s in it for We", and it consists of five rules that have to be implemented in a relation-based contract, in this case for distributed digital preservation:

Focus on results and not on transactions: conform to a business model that will give both parties unanimous interest with focus both on valuable results and on a joint vision for the partnership.

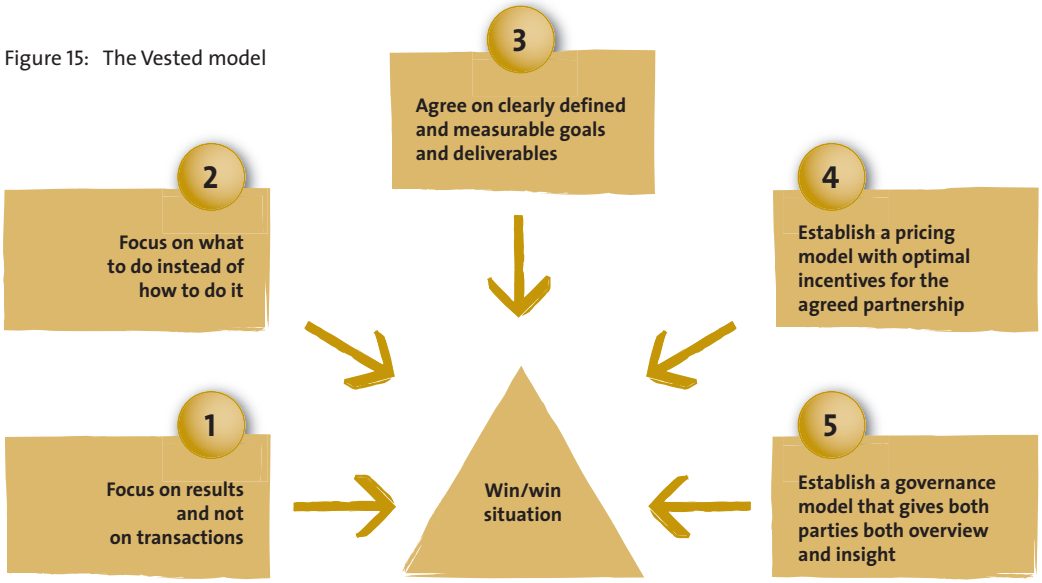
Focus on what to do instead of how to do it: this approach means to concentrate on what to achieve instead of how it shall be done. Traditional outsourcing contracts often have detailed texts on how a service provider shall provide a service. This, sometimes called the "outsourcing-paradox", can end up in a situation where the client outsources a service to an expert organisation, but at the same time describes in detail how this expert organisation shall provide its expertise. The Vested model instead points out the need for both a definition of functions and a roadmap with strategic goals for how the service provider shall support the client in achieving his or her objectives.

Agree on clearly defined and measurable goals and deliverables: traditional contracts on outsourcing often contain agreements about measuring different levels of services and how to compensate the client if the agreed levels are not reached. However, this is not the same as the client being satisfied with the results. In a result based business model, focusing on what to do, the goals and achievements must be clearly defined from the beginning.

Establish a pricing model with optimal incentives for the agreed partnership: the traditional price list is not used in the Vested model. Instead, the service provider shall be economically compensated depending on how the strategic goals are achieved. But the conditions for every pricing model are constantly changing, and both partners must, therefore, have a high degree of transparency regarding their actual costs and economic situations. Otherwise fruitful negotiations about changes of prices will not be possible.

Establish a governance model that gives both parties both overview and insight: the important part in good governance is - according to the Vested model - to focus on the partnership as such and not on the partners. The partners work with a stratified structure, usually found in governance models (see above), but instead of just one interface for communication, with one responsible person per partner, several interfaces are used, one for each specific field in the contract.

Figure 15: The Vested model



MEDIUM TERM PRIORITIES

In this stage the highest priority actions to take are:

- To carry out solid analyses of needs for the redesign of the cultural heritage institution´s existing internal infrastructure, in order to get it effectively integrated with distributed digital preservation services;
- To define a set of governance principles for digital preservation in DCH aiming at infrastructure integration.

LONG TERM PRIORITIES

Most important in this stage is the possibility to offer mature business models for distributed digital preservation services for different types of institutional settings (context and environment).

7 > A WEB-SPACE FOR THE ROADMAP

The roadmap for the implementation of a distributed preservation e-Infrastructure for digital cultural content represents the main outcome of the DCH-RP project.

By definition, a roadmap is not useful if it is not widely disseminated, validated and endorsed by the user groups that it aims to target. DCH-RP contributed substantially to the creation of a wide community of people coming from different sectors (policy makers, cultural institutions, e-Infrastructure providers, etc.) who demonstrated interest in the work done for the development of the roadmap. Now it is important to keep alive and continue to nurture this community, creating awareness about the final version of the roadmap and fostering its diffusion and implementation in Europe and worldwide.

Furthermore, a roadmap cannot be considered as a final step. It has on the contrary to be considered as a living document that needs to be continuously maintained, updated and improved as time passes, technology changes, new requirements have to be taken into account, and so on.

It is for these reasons that the DCH-RP project decided to create a dedicated web-space where it is possible to download the latest version of the roadmap, but also where it is possible for everyone to provide feedback and comments – in other words, a kind of forum dedicated to the use of e-Infrastructure services and facilities for the long-term preservation of digital cultural content.

Figure 16: Overview of sections in the web-space for the roadmap

Apart from presenting and discussing the roadmap, this web-space links also to other relevant material, information and services that are linked to the roadmap itself and that contribute to or supplement it.

In particular, a section is dedicated to the Registry of Services and Tools that was developed in DCH-RP as a practical instrument to help decision makers, DCH communities, institutions and projects to plan the implementation of their digital preservation processes. Another section is dedicated to presenting the results of a range of Proofs of Concept conducted during the project, where cultural institutions and e-Infrastructure providers worked together on concrete experiments targeted at demonstrating how e-Infrastructures can be of benefit for the DCH community, in particular for the preservation of digital cultural content.
From the end of the project, the web-space has been hosted as a section of the DCH-RP showcase in Digital meets Culture (http://www.digitalmeetsculture.net/heritage-showcases/dch-rp/dch-rp-roadmap-for-preservation).

Figure 17:
Entrance page to the roadmap in Digital meets Culture

8 > CONCLUSIONS

The work of DCH-RP has been based on a real community of researchers, curators and other professionals who share a common interest in the study of the preservation of digital cultural heritage.

This Handbook is just one product of DCH-RP. It is a synthesis of a wide set of material that has been developed and gathered throughout the project and that is available for consultation and re-use in the web space for the roadmap. This web space is not only a repository of useful information; it also provides access to supporting services and in particular to the Registry of Services for digital preservation. Both the roadmap and the Registry of Services are conceived as dynamic instruments that need to take into account changes and evolutions in the sector and for this reason contributions and comments from the community will represent a very important added value.

Naturally, the roadmap is just the first step in a long journey, during which we hope to encounter again all the collaborators who worked with us in the development of DCH-RP.

The most important lesson that the roadmap wants to disseminate is that digital preservation is a task to be planned at the beginning of any new initiative dealing with digital cultural heritage, during the design and installation of systems intended to host both born-digital material as well as content derived from a digitisation process.

Finally, implementing the roadmap will be complex and expensive; the preservation of digital cultural heritage cannot be an action to be pursued alone. It will be important to establish cooperation to integrate resources and efforts both with the other sciences and with the various e-government initiatives which share with the cultural sector the need to preserve digital content for multiple uses in today's society and for future generations.

Glossary

Specific terms and the definitions used in this publication:

Born Digital – Digital materials which are created in digital form and do not have or are not intended to have an analogue equivalent.

Cloud computing – a phrase used to describe a variety of computing concepts involving a large number of computers connected through a real-time communication network such as the Internet.

Digital archaeology – the process of retrieving a digital resource which has become inaccessible and unusable due to technological obsolescence and/or poor preservation of metadata about its format, structure and content (for digital records also its appearance).

Digital asset – the material produced as a result of digitisation; the term includes also more complex accumulations such as online learning resources, web pages, virtual reality tours and digital/visual files.

Digital curation – has wider coverage than digital preservation and involves maintaining, preserving and adding value to digital data throughout its life-cycle.

Digital preservation – a set of activities required to make sure digital objects can be located, rendered, used and understood in the future.

Digital record – any information that is recorded in a form that only a computer can process and that satisfies the definition of a record as stated in the formal regulation and/or the policy of the cultural institution concerned.

Digital resources – encompasses both digital records and digital assets.

Digitisation – the process of converting analogue data carriers (parchment and paper records, microforms, photos, film and audio and video tapes) into digital form using scanning, digital photography, or other conversion methods.

E-Infrastructure – the term used for the technology and organisations that support research undertaken through distributed regional, national and global collaborations enabled by the Internet. It embraces networks, grids, data centres, and collaborative environments; it can also include supporting operations centres, service registries, single sign-on, certificate authorities, training, and help-desk services.

Grid computing – the collection of computer resources from multiple locations to reach a common goal.

Hub – a common connection point for devices in a network (could be of different kind).

Memory institutions – a metaphor used about a repository of public knowledge; a generic term used about institutions such as libraries, archives, museums, clearinghouses, electronic databases, and data archives, which serve as memories for given societies or mankind as a whole.

Metadata – information about data required to manage, search, understand, use, and preserve it.

Mashup – in web development, a web page, or web application, that uses content from more than one source to create a single new service displayed in a single graphical interface.

NUMERIC Study – a study on statistics on digitisation of cultural material in Europe; built on the results of this study an EC- funded project, ENUMERATE led by

Collections Trust in the UK, has the task of creating a reliable baseline of statistical data about digitisation, digital preservation and online access to cultural heritage in Europe.

Ontology – a structural framework for organising information; used in artificial intelligence, the Semantic Web, systems engineering, library science, information architecture etc as a form of knowledge representation about the world or some part of it.

Persistent identifier – a long-lasting unique reference to a digital object, which could be a single file or set of files.

Virtualisation – refers in computing to the act of creating a virtual (rather than actual) version of something, including a virtual computer hardware platform, operating system (OS), storage device, or computer network resources.

Visualisation – any technique for creating images, diagrams, or animations to communicate a message. Visualisation today has ever-expanding applications in science, education, engineering (e.g., product visualisation), interactive multimedia, medicine, etc.

Abbreviations

AAI	Authentication and Authorization Infrastructure
AIP	Archival Information Package
API	Application Programming Interface
AQuA	Automated Quality Assurance Project
CHI	Cultural Heritage Institution
COPTR	Community Owned Preservation Tool Registry
CLARIN	Common Language Resources and Technology Infrastructure
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DIP	Dissemination Information Package
DCH	Digital Cultural Heritage
DC-NET	Digital Cultural Heritage NETwork
DCH-RP	DCH-RP Digital Cultural Heritage - Roadmap for Preservation
DP	Digital preservation
EC	European Commission
e-IRG	e-Infrastructure Reflection Group
EU	European Union
EUDAT	European Data Infrastructure
GRID	See Grid computing
ICT	Information and Communication Technologies
HPC	High Performance Computing
HW	Hardware
IaaS	Infrastructure as a Service
INDICATE	International Network for a Digital Cultural Heritage e-Infrastructure
MW	Middleware
NGI	National Grid Initiative
NARA	National Archives and Records Administration (US)

NREN	National Research and Education Network
OAIS	Open Archival Information System
PaaS	Platform as a service
PB	PetaBytes
PEST	Political, Economic, Scientific, Technological
PoC	Proof of Concept
PraaS	Preservation as a Service
PSNC	Pozna Supercomputing and Networking Center
RAID	Redundant array of independent disks (former Redundant array of inexpensive disks)
SaaS	Software as a Service
SCAPE	SCAlable Preservation Environments
SIP	Submission Information Package
SOA	Service Oriented Architecture
SW	Software
TDR	Trusted Digital Repository
TB	TeraBytes
VPN	Virtual Private Network
VRC	Virtual Research Community
VRE	Virtual Research Environment
VRO	Virtual Research Organization

List of figures