/ SUSTAINABILITY

Securing the value of our digital assets

Society is generating a vast amount of digitally encoded information, upon which it depends. Some of this will be needed for a long time.

For those who have responsibility for digital preservation (DP) may well have asked themselves the following:

- How long must the resources be committed?
- How can I justify the resources needed for digital preservation?
- How can I estimate the resources needed and how can I keep these manageable?
- How can I plan to cope as the volume of digital assets increases over time?

APARSEN has collected, evaluated and developed the key answers to these questions.

HOW LONG MUST THE RESOURCES BE COMMITTED?

What are the main requirements for economically sustainable digital preservation?

Definition of economic sustainability: "*The set of business, social, technological, and policy mechanisms that encourage the gathering of important information assets into digital preservation systems, and support the indefinite persistence of digital preservation systems, enabling access to and use of the information assets into the long-term future*".

Blue Ribbon Task Force[1]

The report of the Blue Ribbon Task Force pointed out the need to *"leave open options for future stakeholders"* recognizing that if the current option is *"unsustainable over time, decision makers would need to revisit their options and make a different choice."*

"When the future value of an asset is uncertain, the likelihood of long-term preservation can well depend on current use cases. Preserving materials with clearly defined current uses implicitly creates the option of having the assets available for as-yet-unknown uses that may emerge in the future. Therefore, to the extent possible, value propositions should focus on the benefits generated for current users— about whom decision makers are under- standably most concerned—rather than focusing too much on benefits to future generations and unknown future uses. But sometimes this is just not enough."

Therefore the good news is that resources do not need to be committed forever.

However one must have done the work to leave options open for others.

For **tangible measures of success** we need:

- Recognition of the benefits of DP on the part of key decision-makers;
- Incentives for decision-makers to act in the public interest;
- A process for selecting digital materials for long-term retention;
- Mechanisms to secure an on-going, efficient allocation of resources to DP activities;
- Appropriate organization and governance of DP activities

The underlying assumption is that **DP is ultimately a managerial issue** and strategic planning and management is the key to secure the long term

competitive survival of activities and ultimately of the host organisations.

This approach allows one to address:

- The multiplicity of strategic **objectives** to be met simultaneously
- The diversity of **players**, roles and tasks involved
- The **long term** (continuous) nature of DP business processes
- The multiplicity of **scarce resources** to be managed
- The **changing settings**, parameters and paradigms

[1] *http://brtf.sdsc.edu*

HOW CAN I ESTIMATE THE RESOURCES NEEDED?

APARSEN has performed an analysis of cost parameters, tested those cost models, identified the relationship between costs and benefits and the links to other work for securing long term and easy access to preserved content.

An in-depth analysis of the mapping of cost parameters to the ISO 16363 standard on audit and certification of trustworthy digital repositories identifies and focusses on areas for further investigation and development. The results are presented following a gap analysis to provide indications of where future assessments and reviews could be undertaken. This is particularly valuable where cost models are still under development or where further projects are undertaken.[2] Three cost models were tested providing an insight into how effective cost data, from digital repositories or archives, is in providing the costs of digital preservation services or workflows when applied to different cost models.

Recommendations are given on future developments of the three cost models with advice on the creation of new models. A review of costs in relation to benefits is presented.

Organisational Infrastructure

Governance, organisational viability	Organisational structure and staffing	Procedural accountability, preservation policy	Financial sustainability	Contracts, licenses and liabilities
	CMDP, DANS, DP4LIB, KRDS, LIFE3	DP4LIB, KRDS, LIFE3, PRESTOPRIME	DP4LIB, KRDS, LIFE3, PRESTOPRIME	DANS, DP4LIB, DRDS, LIFE3

Digital Object Management

Ingest acquisition of content	Ingest creation of AIP	Preservation planning	AIP preservation	Information management	Access management
CMDP, DANS, DP4LIB, KRDS, LIFE3	CMDP, DANS, DP4LIB, LIFE3, PRESTOPRIME	CMDP, DANS, DRDS, LIFE3	CMDP, DP4LIB, LIFE3	DANS, DP4LIB, LIFE3	CMDP, DP4LIB, DRDS, LIFE3, PRESTOPRIME

Infrastructure and security risk management

Technical infrastructure risk management	Security risk management
CMDP, DP4LIB, KRDS, LIFE3, PRESTOPRIME	CMDP, DANS, LIFE3, PRESTOPRIME

Bad news

None of the models can be verified and none of them address the whole range of issues that should be addressed.

Good news

We have a good idea of all the areas that are missing and that need further investigation.

---

[2] *At the time of writing, for example, the Coordination Action, 4C (Collaboration to Clarify the Costs of Curation)*

HOW CAN I JUSTIFY THE RESOURCES NEEDED FOR DIGITAL PRESERVATION?

The *"Riding the Wave"* report addressed the question of how Europe can gain from the rising tide of scientific data. Estimates of the value of big data[3] include €250 Billion per year for Europe's public sector administration. To estimate this value McKinsey says that they considered *"only those actions that essentially depend on the use of big data – i.e. actions where the use of big data is necessary (but usually not sufficient) to execute a particular lever."* In order to extract the value a number of issues must be addressed including legal, security related, technological, and organisational. McKinsey noted that *"legacy systems and incompatible standards and formats too often prevent the integration of data and the more sophisticated analytics that create value from big data"* and that there is an increasing *"need to integrate information from multiple sources."*

The image opposite expresses the view that preservation requires continued usability for identified communities from which value can be generated. The value proposition lets one build a business case. The business case allows one to put a business model in place which gives the opportunity to provide flows of resources for preservation.

APARSENs multidimensional approach to sustainable DP practices:

- The Infrastructural dimension through the analyses of suitable *Storage solutions enabling reliable permanent access*

- The operational dimension through the identification, evaluation and proposal of *competitive DP services*

- The monitoring and performance management of DP activities through *cost models and their components*

- The underlying business dimension of DP sustainability by means of modern *management approaches and instruments such as value propositions, business cases and business modelling*.

In a competitive world a compelling Value Proposition briefly presents:

1. The way in which your preserved assets become **relevant for your targeted demand segments**, in terms of:

   a. The main and secondary **benefits** for them and for other stakeholders. Also displaying benefits for users and providers (win-win scenarios);

   b. The actual **short term and long term Impacts** derived from the **usage** of the preserved assets

2. The differential aspects of your preserved assets offering, answering the questions of why your targeted customers should come to you instead of to others with competitive offerings

3. The relevant metrics providing a measurable image of the value of your services, incorporating socioeconomic results (cost savings, innovation, wealth creation, employment…).

3 http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
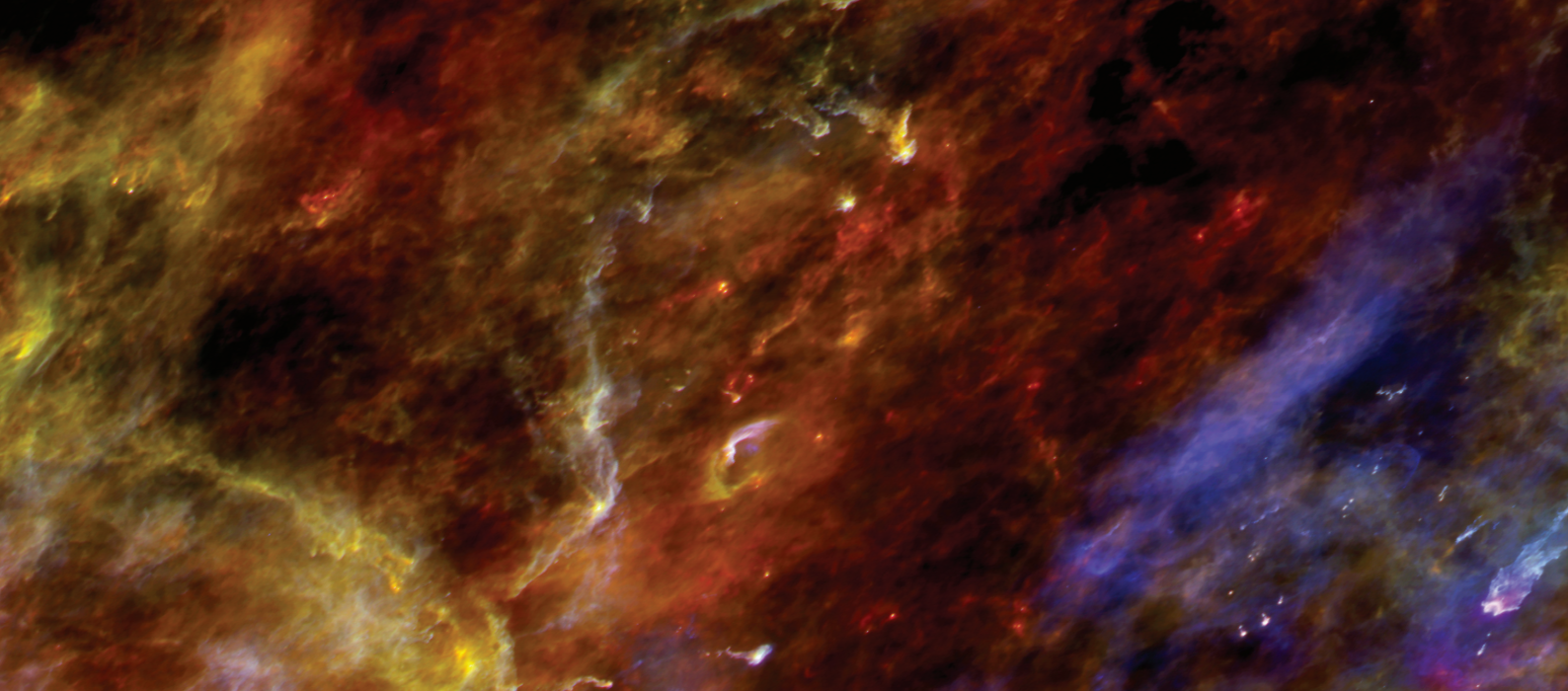
Exemplar business cases analysed by the APARSEN team, as well as the results from APARSEN survey on DP preparedness allowed us to draw up guidelines for deploying sustainable DP policies. The exemplar cases represent the DP intermediate demand side; as most of the institutions currently assume responsibility for the entire DP process, receive digital content from third parties, have partnering strategies for deploying DP initiatives and providing preserved content to stakeholders. The selected institutions (most of them participating in APARSEN) have developed expertise and services in DP, have shifted from tackling preservation at the creation end of the digital life-cycle to embedding preservation features increasingly earlier in the creation process (production context), as well as representing a variety of institutional approaches in terms of type and size.

With this in mind, the key element which all cases have in common is that those responsible for DP claim that DP is at the heart of the organisation's strategy and key investment decisions.

Implementation of DP varies significantly from case to case in terms of financing methods (organizational budget, projects, revenue sources), methodological approach (policies vs. projects), scope (mandate vs. benefits/value driven), technological solutions (in-house vs. commercial solutions, autonomous vs. shared infrastructures) and relevant expertise (use of own staff vs. collaborations/sharing resources). All these approaches are present in the cases examined, though the dominant one seems to be that of sharing resources, costs and expertise through short term collaborations and/or networks.

In terms of business models for DP, the organization's budget is the main funding source for DP in these pioneering cases. Nevertheless a mixed funding model is applied in most of the cases, but to a different degree each case, supplement the organisation's budget by external income sources such as project funding, charges from providing services, grants etc.

APARSEN field work has found that extremely basic business models are currently applied only to a few cases but even when applied it is not always done in a formal manner.

Business models are the conceptualisation of how Value propositions are to be implemented in real world conditions: *"A business model describes the rationale of how an organization creates, delivers, and captures value (economic, social, cultural, or other forms of value). The process of business model construction is part of business strategy; business models are used by managers inside companies to explore possibilities for future development.*

The essence of a business model is that it defines the manner by which the business enterprise delivers value to customers, entices customers to pay for value, and converts those payments to profit: it thus reflects management's hypothesis about what customers want, how they want it, and how an enterprise can organize to best meet those needs, get paid for doing so, and make a profit".

The landscape for DP practices is not simple one. There is a wide array of Threats and Opportunities arising from the current political, economic, social and technological trends configuring the competitive context for deployment of sustainable DP practices:

- **Consolidation of the information society:** everything and everyone will involve the digital in future. There will be new patterns of knowledge acquisition coupled with rising demand from industry and diverse business sectors, ranging from energy production equipment to pharmaceuticals and from financial services to nationwide healthcare systems

- **A data deluge:** explosion of digital content and formats such as social media content, multimedia

- **Misconceptions around** *"digital lasts forever"*: lack of awareness around ownership and responsibility for preserving digital content. There is a perception that simply converting from an analogue format will ensure long term preservation, even though digitally encoded information is itself not simple to preserve.

- **Memory institutions, such as libraries, are becoming digital organisations,** and recognise themselves as such in their own mission statements (BL, KB, etc.).

- **Lack of evidence:** There is an increasing deployment of evidence-based policy, along with a mounting demand on the public sector to provide *"more with less"*. Evidence-based policymaking practices demand measurability of results, impact assessments and return on investment. Such evidence may come too late and, unfortunately may result in a "digital disaster".

- **Compliance with institutional mandates for the leading providers of knowledge are proving not to be adequate:** the monopolies in Public Sector Information supply are falling one after the other and being replaced by other sources such as general purpose search engines or specialised providers of information.

- **Being part of public sector is not a 100% guarantee for long term support:** Increasing requirements for demonstrating return on investment (ROI), increasing political demands for proper economic/employment/wealth creation. Austerity in Public expenditure, cross sector/generalised budget cuts/ freezes/ *"Fiscal Cliffs"*, deaths/abrupt termination of entire lines of public sector driven activities and institutions.

- **Inadequacy of regulatory frameworks:** By its nature digital preservation is a long-term economic activity. Development of this kind of activity requires a favourable, robust and stable legal and institutional framework as the basic conditions to survive. In the digital realm, this requirement is far from being satisfied. Regulatory frameworks are constantly changing and will keep on changing: they form a structural component for future DP management and sustainability (e.g. what to preserve?).

HOW CAN I KEEP THE REQUIRED RESOURCES UNDER CONTROL?

Competitive digital preservation services

Preservation services are an important aspect of sustainability of digital preservation. They are services offered to an organisation (repository, archive) carrying out long-term preservation of digital material, and assisting that organisation in carrying out the tasks relating to preservation. Unless a repository is attempting to conduct all of its preservation activities alone and unaided, there is potential for some kinds of services to be of benefit. The party providing the services might be external to the preserving organisation (a commercial supplier, for example), though it need not be. The point is that services gain their benefit from being widely applicable across a user base, meaning that development effort is for the benefit of all, costs can be shared and knowledge can be pooled.

Services may be of many kinds including automated services in the information technology sense, or services provided by people such as consultancy and audit.

Some examples of preservation services:

▪ A registry repository of representation information, available for use by different repositories in the same or different domains.

▪ A fully functional preservation platform covering ingest operations, backups and integrity checking, and probably other operations as well.

▪ Carrying out audit and certification of a repository to give confidence in its long-term ability to preserve its holdings.

APARSEN has surveyed the landscape of preservation services, mapping services that exist or are under development in a number of ways. One of the mappings is to the ISO/CCSDS standard for audit and certification of trustworthy digital repositories (ISO 16363), which was used for the cost models. The rationale is that the standard elaborates on what a repository must do in many areas in order to be considered trustworthy, and assistance with these functions could in many cases be offered as services.

The conclusions have some important implications for services as a component of sustainability.

▪ There are some areas where there is less coverage of preservation services— for example, in areas of organisational infrastructure or in handling the designated community for preservation.

▪ The domain specificities of services are not well understood.

▪ There is scope for deeper description of services, as part of the infrastructure for preservation.

In some areas, though, there are services already available or under development. The APARSEN Virtual Centre of Excellence will offer services developed by its own members, to other members or externally.
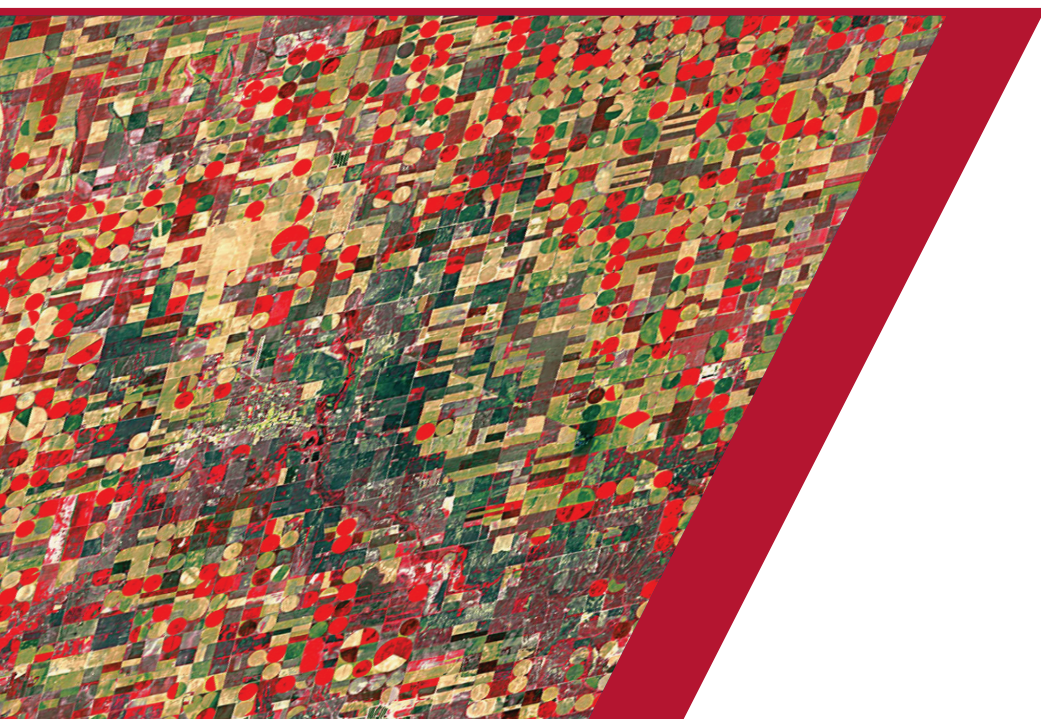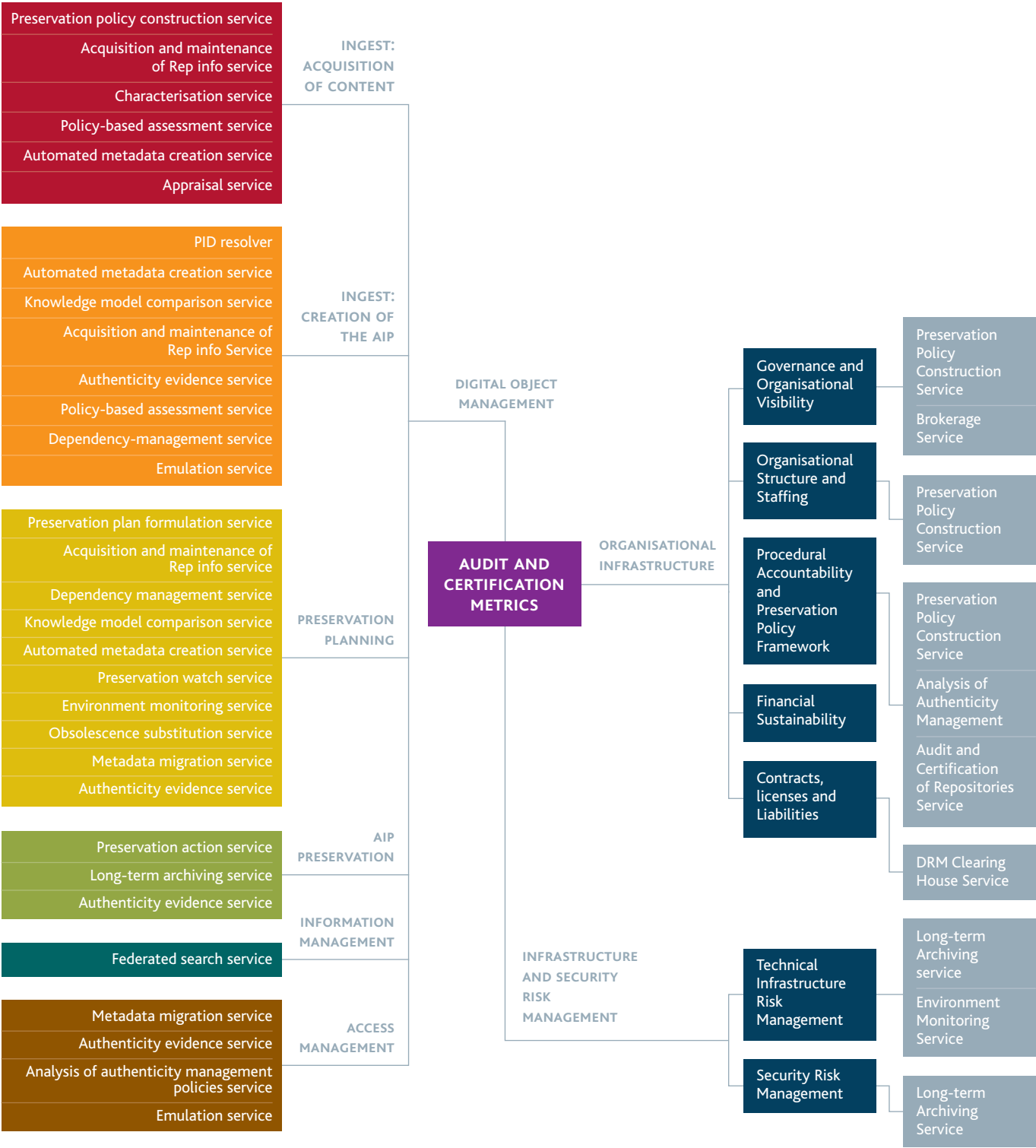
FIGURE 1

Association of preservation services with metrics from standard for audit and certification of repositories

HOW CAN I PLAN TO COPE AS THE VOLUME INCREASES OVER TIME?

Storage solutions

Storage is a central component in any preservation solution, and requires special functionalities in order to adequately address the needs of a preservation system. Storage solutions, on the basis of specific requirements, may vary substantially, e.g. in required capacity, number of objects, size of a typical object, geographical locations. Furthermore, depending on the nature of the data and its usage pattern, performance needs may vary greatly.

New technological approaches are required that meet the legal, business, cost, and scalability requirements of the 'digital age' for long-term retention of digitally encoded information. Recent advances in solutions for Storage-as-a-Service, whereby storage is supplied by a storage provider in a pay-per-use mode, raise the option of adopting new technical solutions such as Cloud Storage as the basic storage for preservation systems. A key aspect is to identify whether such technologies can not only address the cost and scalability needs, but also become a platform for trusted storage, addressing auditability, provenance and a desired level for data integrity.

The APARSEN project analysed the storage solutions currently in use. This mainly considered technical issues and the reliability of storage media. These are important topics since, according to our preliminary studies, an adequate quantification of the risks and the definition of appropriate management and monitoring policies are quite seldom undertaken. The analysis of storage solutions and of storage management policies allows one to gain a better understanding of the problems, and to select a set of relevant issues which need further investigation.

A preliminary list of best practices about ways of supporting digital preservation solutions through storage capabilities can consist in:

- store your data in geographically different places

- use RAID technology to store your data redundantly

- do not use only one storage technology and medium

- be aware of vendor lock-in

- choose storage technologies according to your requirements (regular user access vs. dark archive)

- hierarchical storage management could be a good solution

- establish a disaster recovery policy and test this scenario regularly

- the used hardware should regularly checked and monitored

- data integrity should regularly checked for example by performing checksum checks

- implement policies for data storage management

- every change of the data (for example by repair actions) has to be documented (provenance)

- use secure and encrypted connections to protect your data during the transfer

The recommendations have been collected through a survey on digital preservation across storage solutions run among several key organisations. The survey had the overarching objective to broaden the community feedback on the topic 'relevance of storage solutions with respect to digital preservation of the data', and provided a deeper insight on the topic.
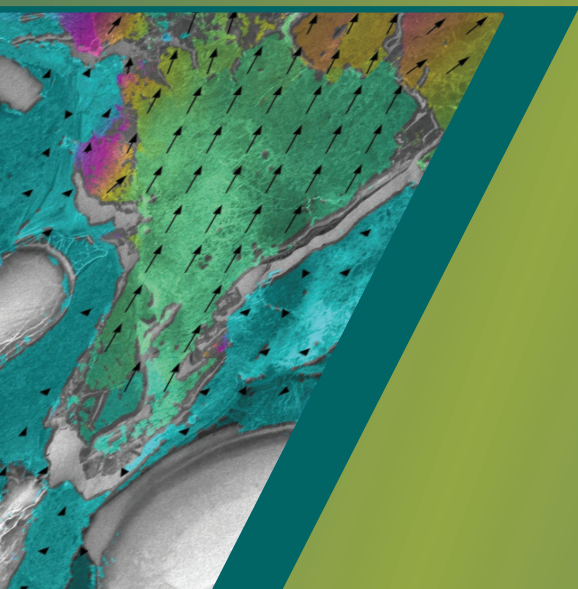
APARSEN brings together expertise from across Europe including partners from industry, cultural heritage organizations, research bodies and membership organizations and will bring coherence, cohesion and continuity to research into barriers to the long-term accessibility and usability of digital information and data. It is defragmenting ideas about and has created a Virtual Centre of Excellence for digital preservation.

Airbus Operations SAS *France*

Alliance Permanent Access (APA) *Netherlands*

Austrian National Library (ONB) *Austria*

Centre Informatique National de L'Enseignement Supérieur (CINES) *France*

Consorzio Interuniversitario Nazionale per l'Informatica (CINI) *Italy*

Data Archiving and Networked Services (KNAW-DANS) *Netherlands*

Deutsche Nationalbibliothek (DNB) *Germany*

Digital Preservation Coalition (DPC) *UK*

European Organisation for Nuclear Research (CERN) *Switzerland*

European Space Agency (ESA) *France*

Fondazione Rinascimento Digitale (FRD) *Italy*

Forschungsinstitut für Telekommunikation (FTK) *Germany*

Foundation for Research and Technology-Hellas (FORTH) *Greece*

Globale Informationstechnik GmbH (GLOBIT) *Germany*

Helmholtz Association (HA, represented by Alfred Wegener Institute) (AFPUM) *Germany*

IBM Israel, Science and Technology Ltd *Israel*

InConTec GmbH (ICT) *Germany*

INMARK *Spain*

International Association of Scientific, Technical and Medical Publishers (STM) *Netherlands*

Koninklijke Bibliotheek (KB) *Netherlands*

Luleå University of Technology (LTU) *Sweden*

Science and Technology Facilities Council (STFC) *UK*

Secure Business Austria (SBA) *Austria*

Space Research Institute of the Russian Academy of Sciences (IKI RAN) *Russia*

Tessella *UK*

The British Library (BL) *UK*

The Stichting LIBER Foundation *Netherlands*

Tieteen Tietotekniikan Keskus Oy (CSC) *Finland*

University of Essex, UK Data Archive *UK*

University of Patras (UPAT), Library & Information Center (LIC) *Greece*

University of Trento (UNITN) *Italy*

APA
RSEN

**Alliance Permanent Access to the
Records of Science in Europe Network**

www.aparsen.eu

Administrative Coordinator:
Simon Lambert
simon.lambert@stfc.ac.uk

Technical Coordinator:
David Giaretta
director@alliancepermanentaccess.org

Join us on

in **http://www.linkedin.com/
groups/APARSEN-3764755**

f **https://www.facebook.com/
APARSEN**

t **https://twitter.com/#!/
APARSENproject**

Dr David Giaretta, MBE
Director

All correspondence should be sent to UK Administration Office:

Alliance for Permanent Access	t: +44 (0)1935 872660
2 High Street	m: +44 (0) 7770 326304
Yetminster	*director@alliancepermanentaccess.org*
Dorset DT9 6LF	
UK	**www.alliancepermanentaccess.org**