Project no. 269977

APARSEN
Alliance for Permanent Access to the Records of Science Network

Instrument: Network of Excellence

Thematic Priority: ICT 6-4.1 – Digital Libraries and Digital Preservation

D11.5 Report on a Common Vision of Digital Preservation

Document identifier:	**APARSEN**-REP-D11_5-01-1_1
Due Date:	2014-12-31
Submission Date:	2015-02-13
Work package:	WP11
Partners:	ALL
WP Lead Partner:	APA
Document status	Released
URN	urn:nbn:de:101-2014121602

Abstract: This is the final report in the series which provides the summary of the Common Vision on issues related to digital preservation reached within the APARSEN project, as the basis for a coherent view for the Virtual Centre of Excellence (VCoE) which gives a common structure to training, services, tools and consultancy.

Delivery Type	Report
Author(s)	David Giaretta and all partners
Approval	
Summary	Final report in series on the APARSEN common vision of digital preservation
Keyword List	Digital preservation, Centre of Excellence
Availability	☒ PUBLIC

Document Status Sheet

Issue	Date	Comment	Author
0.1	2014-08-01	Initial update of D11.3	David Giaretta/ Simon Lambert
1.0	2014-12-12	Candidate final version	David Giaretta
1.1	2014-12-31	Internal review Completion of section 2 Final editorial checks	René van Horik (internal reviewer), David Giaretta, Simon Lambert

Project information

Project acronym:	**APARSEN**
Project full title:	**Alliance for Permanent Access to the Records of Science Network**
Proposal/Contract no.:	**269977**

Project coordinator: Simon Lambert/David Giaretta

Address:	STFC, Rutherford Appleton Laboratory Chilton, Didcot, Oxon OX11 0QX, UK
Phone:	+44 1235 446235
Fax:	+44 1235 446362
Mobile:	+44 (0) 7770326304
E-mail:	simon.lambert@stfc.ac.uk / david.giaretta@stfc.ac.uk

CONTENT

EXECUTIVE SUMMARY

The Common Vision developed in APARSEN has a very specific purpose, namely to bring coherence to the offerings of the Virtual Centre of Excellence (VCoE), helping to reduce or eliminate fragmentation in digital preservation. This coherence is a unique feature of the VCoE, and so it is important that it is well founded on the range of work done within the APARSEN project.

The Common Vision is not only of interest as a foundation of the VCoE, however. As the result of an intensive and concerted effort of integration by the more than 30 partners of APARSEN, with their diverse and complementary expertise and perspectives, it is a resource for the digital preservation community at large.

This document describes the Common Vision developed during the APARSEN project. It is closely connected with the blueprint for the VCoE (D11.6), which describes digital preservation (DP) services, products, know-how and expertise as well as needs which exist in the APARSEN consortium and the communities represented by the APAPRSEN partners as stakeholders of these communities.

The components of the Common Vision are, in increasing level of detail:

- a diagram of the lifecycle of digital preservation, which is founded on ideas of value of digital assets, and provides a visual representation of the most important connections (Figure 1)
- a summary of the overall flow of information and ideas (section 3)
- detailed descriptions of the information, services, tools and supporting evidence associated with the various parts, and which can be used by the VCoE (section 4)
- a collection of 'marketing tag lines' to help explain why this Common Vision is important (section 5)

In addition, there is a glossary which brings together, and shows the relationships between the many terms in the various collections of terminology concerned with digital preservation. This is available on the public website[1] and uses SKOS[2] to relate the terms in a consistent way.

[1] http://www.alliancepermanentaccess.org/index.php/knowledge-base/dpglossary/

[2] http://www.w3.org/2004/02/skos/

1 INTRODUCTION

The Common Vision of APARSEN is one of the main tangible results of the APARSEN network's effect in reducing fragmentation in the endeavour of digital preservation. The organisation of the work of APARSEN into topics and streams was an initial way of giving structure to the existing fragmentation of development efforts, by looking beyond the superficial differences to extract the underlying themes that recur across the field of digital preservation. Within that overall framework, work proceeded in the project towards a vision of what must be done, not only of how it can be structured.

As the Description of Work of the APARSEN project puts it:

> 'This [the Common Vision] brings together the various aspects of digital preservation investigated within the project into a coherent overall understanding of digital preservation which underpins and informs activities, advice and training in digital preservation, as well as delineating areas which require further research.'

This is especially important because providing (in particular) training or advice from a variety of source—the contributors to the VCoE—requires the consistency which the Common Vision provides.

It is important to note that we take the view that the Common Vision does not need to be *universal*, as there might be areas of digital preservation out of its scope. Nor does it imply *unanimous* acceptance of all its content, though it should be broadly agreed.

Previous versions of this report, released annually during the lifetime of the APARSEN project, summarised the progress made according to the mechanisms and processes set up in the project. Now that the project has reached its end, these details are of less interest—it is the final result that matters. Suffice it to say that all partners, and especially work package leaders, had a role in creating the vision, and that it was presented in many fora to allow for engagement with the wider community. This engagement validated the vision and promoted its broad acceptance beyond the APARSEN consortium members. Although the Common Vision has a definite purpose in giving coherence to the offerings of the VCoE, it should obviously make sense to those outside APARSEN, as some of them will be potential members, partners or customers of the VCoE.

Section 2 contains newly written material that attempts to summarise what sort of thing the common vision is, how it was constructed—not in terms of the details of working groups and time schedules but of how it built on the emerging results of APARSEN—and its purposes. In section 3 a summary of the Common Vision is given, which leads into section 4 which has the details of the assets and evidence base for the application of the Common Vision, as well as the gaps which have been identified. A collection of marketing tag lines are provided in section 5 while section 6 illustrates the role of the Common Vision in the offerings of the VCoE, supported by the scenarios, of which a few are provided in the Annex.

2 THE NATURE, CONSTRUCTION AND PURPOSE OF THE COMMON VISION

The common vision is intended to be a distinctive output of the APARSEN project, and as such it is worth enquiring what gives it its distinctiveness, and how its origins in APARSEN have formed its shape and content. Therefore, before expounding the vision in detail we will examine what sort of entity it is, how it has been constructed and the ends it serves.

The span of the common vision is the endeavour of digital preservation as a whole, conceived (as the APARSEN Description of Work puts it) as a view of the landscape, with all the multiplicity of perspectives, scales and orientations that the metaphor implies. The landscape is what we observe from a multitude of vantage points, each with its own particular perspective, be it the researcher in advanced preservation techniques, the practitioner with a digital archive in their care, or the commercial vendor concerned to maximise the attractiveness of their product. The landscape may be seen very differently from these viewpoints, yet no one doubts that there is a single reality of geology, geography and ecology that makes it what it is: the agglomeration of requirements, approaches, models, techniques, tools, experiences and evidence that has been built up through many years of research and practice in digital preservation.

Here we see the first challenge for the common vision: mapping the landscape in such a way as to reveal the underlying forms as well as the superstructures, not clinging to what may be made out from a few necessarily restricted viewpoints, but putting all together into an integrated vision that subsumes these partial views into the whole. This is precisely what is meant by the ‘defragmentation’ that has been reiterated as one of the principal goals of APARSEN.

We should not be misled, however, by the word ‘landscape’ into thinking of something static and without evolution. Here the secondary aspect of the word ‘vision’ comes into play: a vision of the future, of the ideal, of how things could or should be. Digital preservation, like other information technologies, though governed at some level by the laws of logic, mathematics and physics, is an engineering science, concerned with the principled and efficient construction of what is useful. Hence the vision must have room for recognition of where there are lacunae in the available understanding, methods and tools, of where there are imperfections and limitations in the technology; and be able to articulate goals and aspirations for their enhancement. But it is important to acknowledge that the vision is not thereby just a ‘roadmap’ of research and development, plotting out the expected milestones so as to coordinate efforts and direct funding. A roadmap will be influenced by the wider landscape, but it is not the landscape; and the distinctive common vision of APARSEN recognises this.

Having outlined the essence of the common vision, it is clear that its construction is a collective enterprise, drawing on the diverse knowledge bases of a wide range of actors within digital preservation. It is also something to be done in depth: not merely a loose and transient consensus between a few individuals at a conference workshop, nor even the sort of vision that a keynote speaker at the same conference might present, no matter how persuasive. It is rather something that has to be worked at, to be built from the collective knowledge and experience of a wide range of researchers and practitioners, brought into contact over an extended period; which is precisely why it has been developed within a Network of Excellence over four years.

Turning to how the nature of the vision is to be formulated and transmitted, it follows from the above that it will be a complex thing of distinct levels and perspectives, representing a unified whole but not reducible to simple statements or diagrams. To reconcile this complexity with the need to transmit straightforward messages, so as to disseminate the vision and allow wider audiences to grasp and critique it, an overall structure is beneficial, provided this is not confused with the vision itself. In APARSEN this is provided by the diagram of the lifecycle of digital preservation founded on the

notion of value of digital assets (Figure 1), and which provides a visual representation of the most important connections.

To emphasize again: this diagram is not the entirety of the vision, but a framework to structure the vision. However it does imply some commitment: it is not a mere coordinate system onto which features may be mapped. One commitment has already been mentioned: the centrality of value, or more precisely value propositions, addressing the question 'Why preserve a certain digital collection and who would be willing to pay for it?' A further commitment is to the OAIS Reference Model, though in a much expanded context. OAIS is recognised as a fundamental reference in digital preservation, because it represents a thoroughgoing effort put into the working out of a coherent exposition of exactly what a trustworthy digital repository must do. OAIS has nothing to say about business models or business cases *per se*, though the related standard 'Audit and certification of trustworthy digital repositories' (ISO 16363) does include a section on financial sustainability. But the OAIS way of thinking runs through the structure, as expressed in the encircling line in Figure 1 'OAIS concepts and terminology plus extension'.

We can now answer the question of what is distinctive about the APARSEN common vision. It has been noted that there are many lifecycle models that might seem similar; but the APARSEN vision is distinguished first by being comprehensive in its coverage of domains, of approaches and methods, and the knowledge bases on which it drew; and second in adopting a unifying value-based view. It is obvious that digital preservation is an activity with a cost, and one that potentially extends into the indefinite future; what then is the justification, expressed as value, for incurring that cost? This makes value, in particular the ways to *increase* value—i.e. value propositions and business cases—an extremely important consideration to be included in the overall approach to digital preservation. There are technical, management and organisational issues which must be addressed to complete the picture. By basing the common vision on the recognition of the importance of the value proposition, it was possible to bring together all aspects of work in digital preservation, as well as to embed the work in the demand for digital preservation.

The structuring of the vision is of course entangled with design of the APARSEN project itself. This is no surprise: even though the diagram had not been conceived when the original proposal was written, the components of the diagram have long been well known as aspects of digital preservation, and were reflected in the way the project was organised. In many cases individual work packages correspond precisely to parts of the diagram, or there is some overlap, as shown in the table below.

Component of common vision diagram	**APARSEN WPs related**
Preservation (activities) including ingest, store, plan, take action	WP23, WP25, WP27
Usability	WP22, WP25, WP27
Value	WP11, WP36
Business case	
Business model	
Authenticity	WP24, WP22
Rights	WP31
Governance	WP35
Costs	WP32

Selection	WP26 (in part)
Tools	WP16
Services	WP21
Training	WP42, WP43
Audit	WP33

Although the structure of the common vision matches to a large extent the structure of the APARSEN project, it is certainly not true that everything in the vision was already implicit even before a single project meeting had been held or deliverable written. The work packages, taking account of their diversity in scope—though unified by a common approach, as summarised below—fed their findings and results into the vision. How this was done was motivated by the statement in the Description of Work that the vision produces 'a coherent overall understanding of digital preservation which underpins and informs activities, advice and training in digital preservation, as well as delineating areas which require further research.'

The several aspects of this desideratum—coherence, overall understanding, suitability for basing activities, enabling identification of areas for further research—had some implications for the individual work packages: that they must adopt consistent terminology, have consistent interfaces or spheres of responsibility, be sufficiently definite in their findings, and identify gaps in the current state of things. The first two of these are conscious decisions of the project workers, while the others arise from the work done itself.

The basic APARSEN approach at work package level, set out in the Description of Work and followed in their own ways by the individual work packages, was a sequence of Scope ➔ Evaluate options ➔ Integrate/recommend ➔ Spread and sustain. This simple approach was capable of accommodating work packages that undertook research or development work themselves (where the options to be evaluated were precisely those being developed, alternatives being either entirely absent or inadequate), and those that had no such aims but surveyed what existed within their scope and drew conclusions. The fundamental questions behind every work package were: 'What have we got now (and how can it be used)?' and 'What do we still need?' Answering the second question yields the gaps, whether from the judgement of project workers or from the findings of their own developments—'This is how far we reached, but …'. It is likely that in future filling these gaps will be driven by 'market' demand (in a wide sense, i.e. including social and political needs) and sustainability requirements rather than from a 'pure research' motivation.

To illuminate these points, we can present again the table above, now indicating through what mechanisms specifically the work packages fed into components of the common vision: whether surveys of communities, overviews and analyses of existing tools, conclusions of development work undertaken, …. (Note that it is understood that all work packages drew conclusions and made recommendations of some kind or another; the 'type of input' refers to the *modus operandi* of each work package in establishing the bases for these conclusions and recommendations).

Component of common vision diagram	**APARSEN WPs related**	**Type of input from WPs**
Preservation (activities) including ingest, store, plan, take action	WP23, WP25, WP27	WP23: Case studies, survey WP25: Overview, case studies, model

		WP27: Survey
Usability	WP22, WP25, WP27	WP22: Model WP25: Overview, case studies, model WP27: Survey
Value Business case Business model	WP11, WP36	WP11: Integration across project WP36: Questionnaire, case studies
Authenticity	WP24, WP22	WP24: Survey, model
Rights	WP31	WP31: Survey, questionnaire
Governance	WP35	WP35: Questionnaire
Costs	WP32	WP32: Survey, questionnaire
Selection	WP26 (in part)	WP26: Survey, questionnaire
Tools	WP16	WP16: Repository
Services	WP21	WP21: Survey
Training	WP42, WP43	—
Audit	WP33	WP33: Case studies

Finally we turn to the fundamental question of the purpose of the common vision. The answer is twofold: for the benefit of the APARSEN project itself, and for communities beyond the project. The vision has been put to use in APARSEN to ensure coherence of the approach and outputs of work done where different strands of the project's activities were brought together. It would be possible for the individual work packages to have pursued their own agendas without reference to each other, but such individualism (indeed, fragmentation) would render the task hopeless when it came to creation of dissemination material, organisation of external events to present the APARSEN results, and crucially the definition and delivery of training courses. Furthermore, the same coherence helped to define the range or portfolio of value-added services to be offered through the Virtual Centre of Excellence, such as consultancy, advice and technical services. At one level, even the consistent adoption of a common glossary provides a unifying force, but the vision goes further than that in covering the state of the art and expected developments.

Beyond the APARSEN project, the common vision is expected to have resonances in the digital preservation community at large. Actors in the community will be stimulated to relate their own work and experiences to the most closely related areas of the vision; to consider how the concept of value intersects with their own motivating forces; to ask which of the gaps they recognise and how they might be filled; and in general, both to refine and to take novel perspectives from what APARSEN has created.

3 COMMON VISION OUTLINE

This section provides a brief overview of the proposed model structuring the Common Vision. Subsequent sections describe each part in more detail.

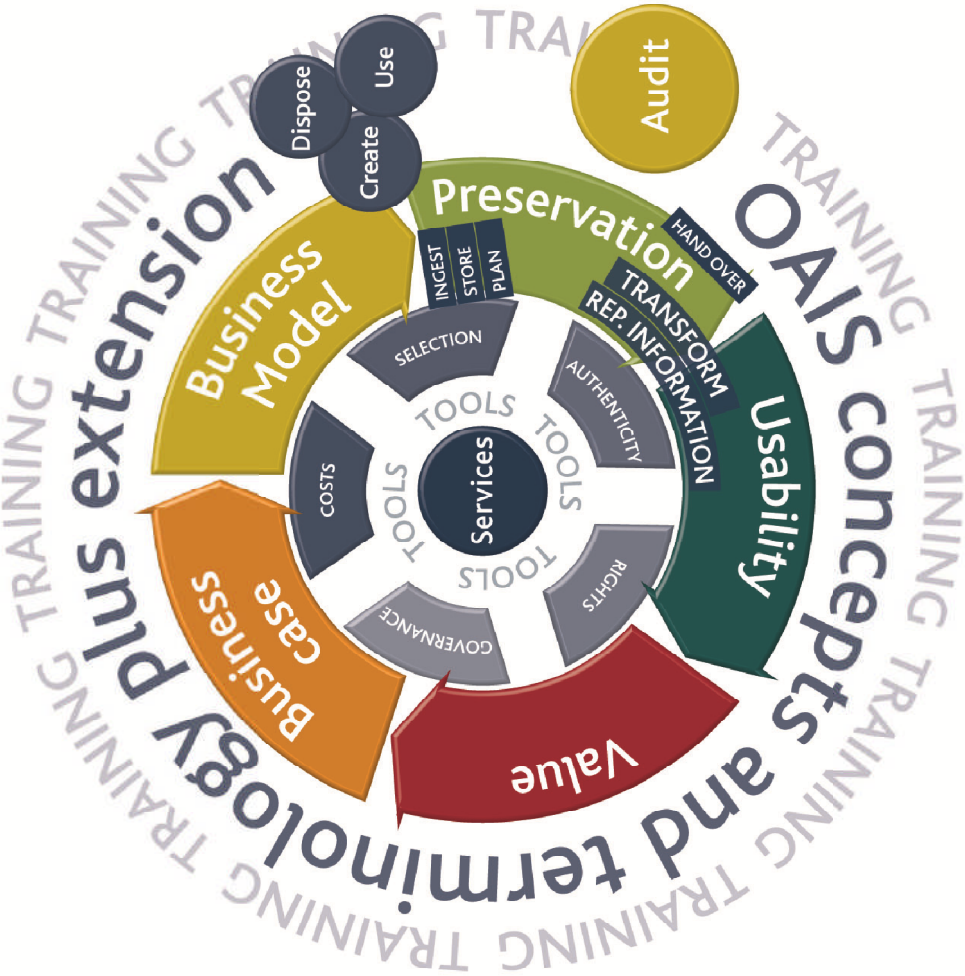

Figure 1 The Common Vision representing the digital preservation lifecycle

The diagram above illustrates the basic sequence of activities to implement a sustainable business process centred in the preservation of digital objects, to be embedded in the overall business cycle of organisations responsible for securing the future usage of such assets.

Note that the focus here is on digital preservation. There is a large number of other models ([1],[2],[3]) with which one may be tempted to compare; these tend to be focussed on the creation of digital objects and the publication of results, or the academic lifecycle, but those models tend to ignore the business model aspects, i.e. how to implement the delivery of digital preservation value proposition over time.

It should be borne in mind that in reality there may be a number of iterations. For example to create a Business case, Value may be re-visited and revised as may be Usability; these iterations are omitted in the flow shown above in Figure 1 for the sake of clarity.

The activities may be summarised as follows:

- Preserve the object by a variety of sub-processes
  - Ingest
  - Store

- Plan preservation, including identifying the designated community (ideally this should be done at the earliest opportunity – certainly before the creation of the digital objects, if we want to secure the best conditions for future usage and we must secure a proper value justification to secure financial resources flows)
- The basic preservation steps to counter changes are:
  - create adequate Representation Information for the Designated Community and/or
  - transform to another file format if necessary or
  - if preservation cannot be carried on by the current organisation then hand over to the next organisation in the chain of preservation
- Evidence about the authenticity of the digital objects must also be maintained, especially when the objects are transformed or handed over.
- Confirmation of the quality of preservation can come from an Audit of the repository (with possible certification)

- Usability
  - Digital objects and digital collections should remain usable, i.e. one (human or artificial agent) should be able to understand and use the digital material. This is closely related to task performability. Various tasks can be identified and layered, e.g. rendering (for images), compiling and running (for software), getting the provenance and context (for datasets), etc. In every case task performability has various prerequisites, (e.g. operating system, tools, software libraries, parameters, representation information etc.). These prerequisites are termed Representation Information in OAIS and the minimum amount of Representation Information needed is determined by the definition of the Designated Community.
  - Additional Representation Information may be created to enable a broader set of users to use and understand the digitally encoded information
    - Other communities may use different analysis tools and it may be convenient to transform the digital object to a more convenient format. This will itself require its own Representation Information (RepInfo); the semantic RepInfo may be unchanged but new structural RepInfo will certainly be needed.
  - The digital objects should also be discoverable in some sensible way – bearing in mind that some information will be publicly available whereas other information will be restricted.
- Value proposition – The portfolio of Value propositions will provide the core of the answers to “Why preserve a certain digital collection and who would be willing to pay for it?”
  - Value propositions must be created by the identification, classification and quantification of the expected benefits which may be obtained by the targeted communities of customers and users from the continuous usage of the preserved objects, which in turn depends on the needs of the users and the usability conditions created for such preserved objects
  - the digital objects will probably be more useful to one type of user community than to another, and this may change over time. These differences and changes must be addressed by a portfolio of Value propositions (as well as by the design and implementation of adequate business models)
  - rights may be associated with the digital objects, perhaps arising from the value or potential value of the object. These rights can generate revenue, and the revenue generation in turn depends on the business model used.

- Business case
  - There is an increasing demand from decision makers to justify: (1) the need for objects to be preserved, (2) the benefits derived of their usage, (3) the costs involved in the preservation, as well as (4) other resources required for preservation
  - Its implementation will be addressed by one or more business models
  - There will almost certainly be options for trade-offs between costs, risks and capabilities
- Business model
  - The business model lays out the business logic, i.e. how the value proposition is consistently delivered to the beneficiaries.
  - Decisions about the mix of sources providing the financial resources required for implementing and operating the preservation business process will be based on the characteristics of the users and customers base (the target groups), the competition in the provision of the preserved assets as well as in the nature and dynamics of the formulated business case.
  - The resources may be used at the very start to create new digital objects, which will presumably have been created for a specific purpose and which then may be either disposed of or be preserved.
  - A selection process will be needed to decide what is to be preserved. This will presumably be based on business case and risk considerations. It may also depend on the interest of other possible curators of the information.
  - This financial resourcing may be (perhaps should be) part of the budgets needed to create the digital objects. However some or all of the digital objects created may be disposed of rather than preserved.

Each of these steps will be assisted by the use of tools and/or services, such as the ones the VCoE should be able to offer.

The underpinning components are first the use of a consistent terminology, the OAIS terminology with extensions to cover those aspects outside the OAIS remit and second the training modules covering all aspects of the Common Vision.

4 DETAILED DESCRIPTIONS

In the following sections each block of Figure 1 is described in more detail. These descriptions each include a table showing the assets base i.e. the sources of information such as APARSEN Work Packages, other projects or software, relevant for the block, and specific items such as documents, software or evidence of effectiveness. These will support the design and implementation of the VCoE range of products and services to be offered to the targeted groups, such as training, specialised technical tools and consultancy services. Finally each block has a collection of gaps – areas where we know there is a paucity of information; these are areas where we should tread carefully and where possible seek to fill in such gaps.

Before moving to the individual blocks, there are some areas of investigation that do not relate to these individually, but cover a range of blocks. The contributions of these areas are presented first, in the same tabular format.

Preservation services

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Understanding coverage and availability of preservation services for all parts of the lifecycle	APARSEN WP21	Detailed analysis of comprehensive range of high-level services, in a standard format; but must be kept up to date to maintain its value	Knowledge of involved partners
		Mapping of services to metrics of standard for audit/certification of Trustworthy Digital Repositories (TDR)	Mapping to audit/certification standard

Gaps
Less coverage of services in areas of organisational infrastructure and infrastructure and security risk management (Relates to blocks of vision: rights, governance, business case, business model)
Domain-specific services (or understanding of domain dependencies)
Deeper description of preservation services and their interrelationships

4.1 PRESERVATION: INGEST

Moving data from the point of creation or initial use to a preserved version may be done in a number of ways. The key aim is to ensure that information about the digital objects, which almost certainly only the creators may have, is transferred. For example the Representation Information – not just the formats (structure) but also the semantics and other RepInfo such as the software needed to use the digital objects. In OAIS terminology, Archival Informational Packages (AIP) must be created.

It should be noted that one of the reasons to create AIPs is that if at some point the repository cannot continue in its preservation activities for this information, for example if the repository closes down, then it is possible to hand over the AIPs to the next in the chain of preservation. This ensures that all the information needed to allow the digitally encoded information to be preserved is handed on and nothing is forgotten.

An important decision to be taken by the repository is the definition of the Designated Community – those types of users for whom the repository guarantees usability.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Planning hand-over from creators	PAIS and PAIMAS standards	Templates for defining the hand-over process	
		APARSEN deliverable D26.1	The report explains the importance of the mutually dependent goals annotation, reputation and data quality. Annotations do not necessarily have to be available at the time of ingestion but it is probably the time when most annotations are added and thus a major issue of ingest. The report gives an overview of issues which should be considered when developing a research data repository as well as when annotating research data.
Definition of Designated Community	APARSEN WP25	CASPAR deliverables and published papers	
	SCIDIP-ES GIS service [14]	Tool to define Designated Community and associated user feedback	

Creation of AIPs	SCIDIP-ES Packaging, plus Preservation Strategy Toolkit and RepInfo Toolkit.	Tools for creating AIPs. Support from evidence collected by CASPAR. SCIDIP-ES user feedback	
	Preservica (Ingest workflow and standalone packaging application) and KoLibRI, RODA, Rosetta, etc.	XIP metadata schema based SIPs, AIPs (& DIPs).	Demonstrable ingest workflows available in preservation systems. (See D14.1 for screenshots of ingest workflows from partner test environments).
Association of Persistent Identifiers (PI) for people and digital objects on ingest	APARSEN WP22	PI Interoperability Framework	
	SCIDIP-ES HAPPI toolkit	PI creation	
Automation of the extraction and harmonization of the embedded (or implicit) metadata from the various file formats	Tool PreScan that was developed in the context of the CASPAR project	The Pre-Scan tool itself. Experience in building tools that extract the embedded metadata in from digital files, and produce harmonized warehouses of metadata.	It has been tested. The results are reported in the publication Yannis Marketakis, Makis Tzanakis, Yannis Tzitzikas: PreScan: towards automating the preservation of digital objects. MEDES 2009: 404-411
Capability to ingest a great variety of data	PANGEA Data Publisher for Earth and Environmental Science	PANGEA has long standing experience in ingesting data from a wide variety of customers.	

Gaps
No common shared technical implementation for how to package up SIPs (or AIPs) for ingest in to any repository. All ingest processes use their own bespoke implementation to create SIPs, resulting in minimal if any integration with other repository systems. Typically requires bespoke software engineering tasks to export / transformation / import AIPs from one repository system into another. No simple interchange of AIPs between systems.
Automation of the extraction and harmonization of the embedded (or implicit) metadata from the various file formats. Although this approach can produce automatically big warehouses of harmonized metadata, at no cost, it is not widely known or used.

4.2 PRESERVATION: STORE

The AIPs must then be stored. The individual components of an AIP may be stored separately, for example details of Provenance may be stored in a database while Context may be provided by published documents and the Content Data Object may be stored in a separate repository.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Selection of storage systems	APARSEN WP23: Storage solutions	Survey results from a small number of repositories	
	SCIDIP-ES Storage service	Test results and user feedback	
		AIPs that remain accessible to this day and have been held in digital repositories for 10+ years	Records held in Digital Archives like those at the TNA, although we note that these records may not be proper AIPs.
Design of system to be able to cope with the required scale	APARSEN WP27: Scalability	Survey results	
Different dimensions of scalability	APARSEN WP27	Survey results	
Storage and managed control of AIPs	Preservica	Cloud based Preservation as a Service capability including passive storage at 99.999% availability performance	
		WP14 – Common Test Environments	Demonstrated access to AIPs held in partner test environments, and/or other digital preservation systems, repositories and test environments.
Storage of context	TIMBUS	Context model	
Data storage and long-term archiving for earth observation data	APARSEN WP23	ESA's G-POD storage elements and Multi-Mission Facility Infrastructure	

Gaps
Most implementations do not share a common design e.g. implementation of the interfaces implied by OAIS. If the AIP interfaces are not provided then it may be difficult to use effectively in the transfer of information plus PDI from one repository to another.
Store not only digital objects, but also the complete workflow / process that generated the object, including its context (technical, legal etc.) and all relevant information.
Systems which are more generally available which include capabilities such as the following: G-POD: • Intelligent caching over large on-line archives • Optimization of data circulation for caching purposes: data granularity, network configuration, seeding strategy; • Caching strategies based on use-patterns, trying to "guess" which data will be requested next. For ESA MMFI: • Preservation strategies, like the periodic migration of digital products to new information technology. • Encapsulation by self-describing items in the OAIS model Modular design, open architecture and streamlined interfaces to permit an easier substitution of one or more of its elements

4.3 PRESERVATION: PLAN

If nothing changes then no preservation activities would be needed. However we know that many things change including hardware, software, environment and the knowledge base of the Designated Community.

The PARSE.Insight project collected a large amount of evidence about the threats which the community recognises as important and which must be countered.

A mantra concerning the strategy to preserve digital objects and which is often heard in the library world is “emulate or migrate”. This works well for digial objects for which the semantics can be largely ignored – usually rendered objects which are visually inspected (or read) by humans, in other words documents and images. However this does not work well in situations where semantics is important, for example scientific data and sophisticated/distributed/heterogeneous commercial/business/industrial data.

The CASPAR project collected evidence about techniques which are applicable to the preservation of a large variety – in principle all –types of digitally encoded information. In particular the semantics (meaning) as well as the structure must be captured as parts of the Represen tation Information. In addition software, including for example emulators, should also be captured.

As noted previously, another part of the preservation plan should be to identify the next in the chain of preservation to hand on to if needed.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Create plans for preservation	SCIDIP- Preservation Strategy Toolkit and RepInfo Toolkit. APA/APARSEN website	SCIDIP-ES software – open source	CASPAR evidence, SCIDIP-ES user feedback.
	Preservica preservation workflows	Cloud-based preservation planning	
	SCIDIP-ES orchestration/broker service	Software to distribute notifications about changes plus information about possible curators and experts	CASPAR evidence, SCIDIP-ES user feedback [13].
	PLATO/SCAPE	Planning tool	
Evaluate preservation capability	APARSEN deliverable D14.1 Common Test Environments	Process for evaluating the capability of disparate systems to perform preservation actions on a wide and diverse	D14.1 and evaluation spreadsheet

		set of digital object types.	
Policy-based planning	APA/APARSEN website with reference to the recommendations of policies able to support the definition of preservation plans	Collection of user scenarios and suggestions of best practice as well as evidence about preservation efficacy.	
Dependency management	CASPAR	Dependency management framework	
Preservation planning conceived in terms of performability	APARSEN WP25	Task performability service based on dependency management approach Model and framework for emulators and converters	Results from the evaluation of the prototype called Epimenides, http://139.91.183.63:8080/epimenides/, developed within WP25 The applicability of the proposed approach has been investigated over the practices and systems of some partners (e.g. DANS in the data archiving domain).
Management of persistent identifiers in preservation planning			

Gaps
There are many specialised pieces of RepInfo for which specialised tools would be useful but which do not yet exist.
The quantity of differentiated policies implied in the preservation plans must be assessed and made available.
Many dependencies, which emerge when exchanging digital objects between different systems and services, are not resolved by standardization. Standards are not widely adopted and sometime discrepant standards are in use in different organizations. More flexible methods for tac kling these dependencies and managing intelligibility gaps are needed.

4.4 PRESERVATION ACTIVITIES

The basic steps in preservation to counter changes are:

- create adequate Representation Information for the Designated Community and/or
- transform to another format if necessary or
- if preservation cannot be carried on by the current organisation then hand over to the next organisation in the chain of preservation

The mantra is therefore "collect Representation Information, transform or hand on to the next in the chain of preservation" rather than "emulate or migrate". Evidence about the authenticity of the digital objects must also be maintained, especially when the objects are transformed or handed over (see below). Confirmation of the quality of preservation can come from an Audit (with possible certification)

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Definition of Designated Community	APARSEN WP25	Deliverable D25.1. In D25.1 some possible solutions to fill these gaps (below) have been proposed. An example in the domain of preservation metadata is represented by the broad adoption of international standards and abandoning local solutions and ad-hoc metadata schema. The most promising standard we have identified is PREMIS Data Dictionary for Preservation Metadata. http://www.loc.gov/standards/premis/	Several sources of evidence of the effectiveness of the implementation of the PREMIS data dictionary have been presented at the iPRES2013 workshop titled "PREMIS Implementation Fair 2013". http://www.loc.gov/standards/premis/premis-implementation-fair-agenda-2013.html Apart from the aforementioned paper, a tool based on this approach (called RIMQA) has been implemented and experiments are reported in that paper.
	SCIDIP-ES GIS	Software to help define the Designated Community (DC) and implications of changes to the DC.	
Perform preservation actions	Preservica Preservation workflows	Cloud based preservation actions	
Evaluate Preservation	D14.1 Common Test Environments	Process for evaluating the capability of disparate systems to perform preservation	D14.1 & Evaluation spreadsheet

capability		actions on a wide and diverse set of digital object types.	
Creation of RepInfo	APARSEN WP14 SCIDIP-ES RepInfo Toolkit, Preservation Strategy Toolkit, Registry, Gap Identification service	Deliverable D14.1 CASPAR evidence SCIDIP-ES software and User feedback	
Emulation	KEEP (emulation software) ENSURE (Virtual machines)	Software (check licences)	
Transformation	OPF related SCAPE Various e-science projects	Details of software	
Handover	SCIDIP-ES Brokerage/Orchestration service	Examples of hand-over of digital objects	
Audit	APARSEN WP33	Spreadsheet to capture evidence about quality of preservation	
	SCIDIP-ES certification toolkit	Tool to perform self-evaluation	
Selection of interoperability approaches and solutions which can have impact on preservation activities	APARSEN WP25	D25.1, in particular the matrix of interoperability solutions, gap analysis and recommendations	
How to curate the specificity of the various ontology-based metadata, while the	APARSEN WP14	Experience in both theory and its applicability (including tools) A paper that describes the approach: Tzitzikas, M. Kampouraki, A. Anastasia,	

ontologies evolve (this is important for e-Science)		Curating the Specificity of Ontological Descriptions under Ontology Evolution, Journal on Data Semantics, (accepted for publication in 2013).	
Interoperability	See section about Usability (WP25)		

Gaps
Interoperability. Several interoperability gaps have been identified and classified in D25.1. In particular the following domains have been investigated: 1) Identification systems (for digital objects, authors and datasets) 2) Library classification systems 3) Library Linked Data 4) Metadata 5) Ontologies and Vocabularies 6) Data Provenance 7) Preservation tools 8) Exchange standards 9) Preservation Frameworks 10) Semantic annotation services 11) e-Science infrastructures. Some of the identified gaps can have strong impact on preservation strategies and activities. Some examples follow. 1) Lack of cross-organization coordination in the definition of metadata preservation schema to capture, maintain and share information about provenance, authenticity, preservation activity, technical environment, rights management and so on. This has led to the development of a set of metadata reflecting the particular needs and requirements of the specific community that authored them. 2) Lack of a scalable infrastructure for the efficient planning and application of preservation strategies for large and heterogeneous data collections. 3) Many different suites and preservation tools are in use in different communities (e.g. iRODS, LOCKSS). The isolation from each other represents an obstacle for inter-institutional preservation and interoperability.
Although critical for e-science, the community is not aware about the loss of specificity that happens when world models (ontologies, taxonomies, thesauri, controlled vocabularies) evolve over time.

4.5 AUTHENTICITY

Authenticity is a fundamental issue for the long-term preservation of digital objects: the relevance of authenticity as a preliminary and central requirement has been investigated by many international projects. Some focused on long-term preservation of authentic digital records in the e-government environment, and in scientific and cultural domains.

In the OAIS reference model authenticity evidence is part of the PDI (Preservation Description Information), it mostly deals with fixity and provenance, but other aspects, such as Persistent identifiers, interoperability, data quality and non-technical provenance, e.g. reputation, are involved as well.

Managing authenticity implies the early collection of evidence along the whole lifecycle of the digital resource, and possibly their standardized organization over time according to well defined policies, in order to lay proper bases for interoperability.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
State of the Art	APARSEN WP24 Extensive state of the art that addresses the main international projects in the field, as well as the standards, recommendations and guidelines for keeping and preserving digital objects. Includes an extensive reference list and an appendix where all the major projects in the area are reviewed.	Deliverable D24.1 APARSEN Internal Deliverable: ID2401 Resorting to inference allows one to derive provenance from elementary evidence, thus simplifying management and diminishing storage requirements.	The work described in C. Strubulis, Y. Tzitzikas, M. Doerr and G. Flouris, Evolution of Workflow Provenance Information in the Presence of Custom Inference Rules , 3rd International Workshop on the role of Semantic Web in Provenance Management (SWPM'12), co-located with ESWC'12, Heraklion, Crete, June 2012
Authenticity management	APARSEN WP24 Systematic methodology for the management of authenticity of digital objects along their lifecycle. The methodology is complemented by a set of operational guidelines, i.e. a sequence of steps that should be followed, to get to the definition of an adequate *authenticity management policy* that is to formalize the rules according to which authenticity evidence should be collected, managed and preserved along the digital resource lifecycle.	- Deliverable D24.1 - M. Guercio, S. Salza. Managing Authenticity through the Digital Resource Lifecycle. In: M. Agosti et al. eds. Digital Libraries and Archives, IRCDL 2012 Revised Selected Papers. Comm. in Comp. and Inf. Science, vol. 354, p. 249-260, Berlin, Heidelberg: Springer-Verlag, doi: 10.1007/978-3-642-35834-0	

Case studies	APARSEN WP24 APARSEN authenticity methodology and guidelines have been tested in several real life environments. The case studies have proved the robustness, flexibility and effectiveness of the methodology. This documentation may be a useful starting point for those aiming at improving authenticity management in their repository.	Software - Deliverable D24.2 - S. Salza, M. Guercio. Authenticity Management in Long Term Digital Preservation of Medical Records. In: R. Moore et al., Proc. of the 9th Int. Conf. on Preservation of Digital Objects, Toronto, Oct. 2012. ISBN: 9780991799701.	The case studies have proved the effectiveness of the methodology and guidelines in a large variety of contexts.
Policies for authenticity evidence management	APARSEN WP35: Survey and recommendations	- Deliverable D35 (under development).1	
	APARSEN WP35 InterPARES Framework definition and assessment of authenticity services	- i-TRUST: recommendations, agreed European and transnational teams, for the assessment of authenticity management in LTDP repositories	
Authenticity services	SCIDIP-ES authenticity toolkit. This is an implementation of the APARSEN methodology within the SCIDIP-ES project.	Software to capture evidence to support claims of authenticity	The implementation of the interoperable framework proves the applicability of the methodology at implementation level.
	APARSEN WP21 Two different services have been proposed: a consultancy service to assist in analysis and design of authenticity management policies, and an authenticity evidence service based on an interoperable framework that can be customized to implement the policies (including consultancy support)	- Deliverable D21.1	
Implementing authenticity management	APARSEN - SCIDIP-ES cooperation Cooperation between these two projects has led to the implementation of and Interoperable framework to	- SCIDIP-ES Authenticity Toolkit: a software framework, based on REST principles and Java technology.	

policies	support the management of authenticity evidence in a LTDP repository.	- L. Briguglio, S. Salza, M. Guercio. Preserving Authenticity Evidence to Assess Provenance and Integrity of Digital Resources. In: P. Nesi, R. Santucci. Information Technologies for Performing Arts, Media Access, and Entertainment, Proc. ECLAP 2013. LNCS, vol. 7990, Springer-Verlag, ISBN: doi: 10.1007/978-3-642-40050-6_7	
Audit	APARSEN WP33 Case studies about audit and certification of authenticity management policies in digital repositories that have been audited as part of WP33 activities.	D33b	
Persistent identifiers	APARSEN WP22 DIGOIDUNA project (https://digoiduna.wordpress.com) ODIN project (http://odin-project.eu) Building an interoperability infrastructure for persistent identifiers systems for digital objects, authors, contributors and other related resources (e.g. organizations) as a crucial step to develop trust-enabling services such as authenticity, citability and provenance certification services.	WP22 survey on PIs systems D22.1 scenarios on authenticity Interoperability Framework (IF) for PIs systems developed within WP22 High Level Expert Group (HLEG) feedback and evaluation of the IF for PIs Results from the two APARSEN Workshops on interoperability between PIs systems	One of the main results of WP22 has been the definition of an interoperability framework for PIs. The trust criteria of eligibility to the framework have been recognized and agreed by experts as the crucial aspect to adopt the framework as the core interoperability layer for authenticity and provenance services. The main evidence comes from the results of the evaluation of IF by the High Level Expert Group in WP22. D22.2 and D22.3 include a discussion of these results and a revision of framework based on this expert feedback. Authenticity and provenance services are also discussed in

			details.
Modelling and exchanging provenance information	APARSEN WP24 Core ontology for provenance, mappings with other schemas that can capture provenance. Inference rules for propagating provenance.	- ID2401 - C. Strubulis, Y. Tzitzikas, M. Doerr and G. Flouris, Evolution of Workflow Provenance Information in the Presence of Custom Inference Rules , 3rd Intl. Workshop on the role of Semantic Web in Provenance Management (SWPM'12), co-located with ESWC'12, Heraklion, Crete, June 2012.	

Gaps
Insufficient gathering of authenticity and provenance evidence during the digital resource lifecycle. Case studies have shown that the management of many repositories is not sufficiently aware of the relevance of the issue.
Lack of an acknowledged standard for the interchange of authenticity evidence. The issue is crucial, since in many cases preserved digital objects undergo changes of custody during their lifetime.
Lack of adequate definitions for transformational properties, to be used to check authenticity and integrity when the representation of digital resources undergoes transformations.
Application of provenance mapping and rules to tracing of authenticity evidence with large numbers of generations of large number of objects
Secure logging, which is important in terms of confidence in the evidence which is presented needs to be converted into a practical method
Data quality, which is very far from being generally solved
An automated way of evaluating evidence about authenticity is not available. Additional kinds of annotation are available, ranging from publications in refereed journals to grey literature.
Development of automated processes for making interoperable and sustainable large and differentiated groups of authenticity evidence information
The lack of an integrating infrastructure for discovering and locating digital resources (identified by different PIs) which ensures some basic fundamental criteria of trust as part of the process of assessing the authenticity of the retrieved resources.
The community is not aware that for tackling the problem of incomplete provenance one has to resort to inference (otherwise the storage requirements will be prohibitive, and the correction of errors is laborious).

4.6 USABILITY

Rendered objects tend to be looked at by humans, often one person looking at (or hearing) one rendering of a digital object at a time. Generating results and combining information from multiple objects is done within a human mind. These rendered objects of course have value and that value can be increased by various means including pointing to related information. Note that to do this often means treating the information as data (as for example Google does).

When one deals with information which is normally not rendered a number of different issues must be considered. Such information is instead processed, often combining information from multiple sources. This requires various types of Representation Information, particularly types which can be treated computationally. For example a machine readable description of a digital object, such as DRB [12], can be used to extract selected pieces of information (numbers, text etc.) from that digital object, to be combined with information extracted from other digital objects. Virtualisation techniques are often used to increase automated use – including data, hardware and software.

One way to increase the usability of a digital object is therefore to add RepInfo which makes the digitally encoded information understandable to a broader set of users, beyond the Designated Community. Note that the Designated Community is special in that the repository has guaranteed to ensure that the digital object can be understood and uses by that community – this defines the minimum amount of RepInfo the repository must have – matching the knowledge base of the Designated Community.

The repository can choose to add as much additional RepInfo so match the knowledge base of any other community – but does not guarantee to continue to do this into the future. This (wider) community will be able to understand and use the information and this value may justify the maintenance of the RepInfo into the future.

The digital objects should also be discoverable in some sensible way – bearing in mind that some information will be publicly available whereas other information will be restricted.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Creation of RepInfo to match knowledge base of a community	SCIDIP-ES tools and services	**Tools and Systems** From SCIDIP-ES: 1/ RepInfo Toolkit, 2/ Preservation Strategy Toolkit, 3/ Registry, 4/ Gap Identification service	CASPAR evidence base
Interoperability Objectives and Approaches	APARSEN WP25	Deliverable D25.1: Survey of interoperability objectives and approaches. It contains an extensive collection of initiatives, projects, solutions, scenarios, recommendations.	

Automated Reasoning for Interoperability	APARSEN D25.2 and SCIDIP-ES software	**APARSEN D25.2**: A modelling approach that enables automatic interoperability reasoning. It is converter and emulator aware. It can show how to achieve interoperability by combining existing software. In brief D25.2 proposes a modelling approach that enables task performability checking, which in turn can reduce the human effort required for periodically checking or monitoring whether a task on an archived digital object or collection is performable, and consequently whether an interoperability objective is achievable. Such services can also assist preservation planning, especially because now converters and emulators can be modelled and exploited by the dependency services. Overall, the methodology for capturing, modelling, managing and exploiting the various interoperability dependencies can be considered as a significant contribution to the VCoE: expertise in designing and realizing novel inference services for task-performability, risk-detection and for computing intelligibility gaps. Furthermore, the implemented system (which is already web accessible) can be used for disseminating the results of this work, as well as for investigating and planning future operational applications of this approach, either in the context of single organizations (e.g. the DANS case), or in the context of the VCoE (e.g. as an advanced semantic registry).**Tools and Systems**	

		1/ APARSEN WP25: **System Epimenides** that proves the feasibility of the approach. The user evaluation was also positive. 2/ **CASPAR Gap Manager**: Positive experience from CASPAR 3/ **SCIDIP-ES Gap Identification Service**	
	SCIDIP-ES software and User feedback	SCIDIP-ES RepInfo Toolkit, Preservation Strategy Toolkit, Data Virtualisation Toolkit, Registry, Gap Identification service	CASPAR evidence
Virtualisation techniques	Models of Virtualizations - Hierarchy of tasks as described in APARSEN D25.2	Tools and Systems - SCIDIP-ES Virtualisation toolkits - SHAMAN results	Experience of application of virtualisation techniques
	SHAMAN	Multivalent software	
Discovery techniques	APARSEN WP22	Ontology and tools for PI	
	SCIDIP-ES Finding Aid Preservica EUDAT	Search and browse capability	
Searching and accessing information across PIs domains and connecting this information across multiple services and infrastructures for e-science communication	APARSEN WP22 DIGOIDUNA	PI Interoperability Framework PI resolvers	Two services for searching and accessing distributed information have been developed as part of the WP22 research activities. Using a trusted PI as input the services allow to navigate the network of information connected to the identified entity. For example, providing a PI for an author, it is possible to find his publications and accessing to those which are available on trusted repositories.

			ORCID and ISNI joint statement on interoperation VIAF and ISNI interoperability initiative
	ODIN	ODIN Proof of concepts in two domains: High-Energy Physics and Humanities and Social Science	Preliminary models on data exchange and workflows
How to achieve interoperability without solely relying on standards.	APARSEN WP25	The methodology, the methods and the tools that are described in D25.2. In brief D25.2 proposes a modelling approach that enables task performability checking, which in turn can reduce the human effort required for periodically checking or monitoring whether a task on an archived digital object or collection is performable, and consequently whether an interoperability objective is achievable. Such services can also assist preservation planning, especially because now converters and emulators can be modelled and exploited by the dependency services.	
Generic comparison functions for detecting what has been changed	APARSEN WP25	Methods and tools for comparing RDF Graphs (which is currently the lingua franca for metadata management).	
How Information Objects are defined	APARSEN WP25	The theory presented in the CoRR paper: Martin Doerr, Yannis Tzitzikas: Information Carriers and Identification of Information Objects: An Ontological Approach. CoRR abs/1201.0385 (2012)	
Interoperability Objectives and Approaches	APARSEN WP25	Collection of initiatives, projects, solutions, scenarios, recommendations, etc.	
Data are made available in	PANGEA		

a structured way enabling complex searches	(http://www.pangaea.de)		

Gaps
Interoperability. Lack of an interoperability infrastructure, which can ensure a unique point of search and access to information, which is identified by different PIs and distributed across systems and disciplines.
Lack of interoperability and coordination between PIs and Linked Open Data initiatives.
Cooperation among the major PI systems on governance, policies and technology is still in its infancy stages and this is especially true in the case of PI systems for different type of entities (e.g. contributors and datasets).
Lack of common access methods e.g. interfaces to images and tables, in end user software (data virtualisation)

4.7 DIGITAL RIGHTS MANAGEMENT (DRM)

Intellectual property rights (IPR) and access rights of various kinds may be associated with the digital objects. According to OAIS, the digital archive has to "obtain sufficient control for preservation". The problems of assuming sufficient control of the Content Information and Pres ervation Description Information, when they are largely digital, are addressed in three related categories, as follows:

1. copyright implications, intellectual property and other legal restrictions on use;
2. authority to modify Representation Information;
3. agreements with external organizations.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
DRM is used in a broad and a narrow sense: can be implemented inside or outside the file that is being protected	APARSEN WP31	an extensive overview of DRM techniques and tools and their risks for the long term preservation	
DRM mechanisms that are built into file formats like the protection for viewing, copying, printing and altering can prevent necessary digital preservation actions	APARSEN WP31	a guideline for dealing with DRM protected material and its preservation	
There is no standard	APARSEN WP31, tools for generating technical metadata	overview of appropriate best practices from the workflow perspective, i.e. the	

workflow in the Ingest workflow to check if the digital objects are protected	and file analysing like FITS, JHOVE, ExifTool, WP21 (Characterization service)	use of automated tools for the detection of DRM protected material and its preservation	
Preservation of digital rights that are associated with digital objects	APARSEN WP31, PREMIS, METS rights, Europeana Data Model	Overview of metadata standards that allow the including of digital rights information Mets Rights: Rights Declaration Schema http://www.loc.gov/standards/rights/ PREMIS Rights entity http://www.loc.gov/standards/premis/	OAIS: http://public.ccsds.org/publications/archive/650x0m2.pdf, page 3-1, 3-2.
Use of persistent identifiers in DRM	LCC project		

Gaps

4.8 VALUE

Value is usually created by the usefulness of the digital object, which in turn depends on its usability. However there are many other reasons for assigning value for example having local control over critical pieces of information even if no use is expected in the foreseeable future.

The value may be "potential" value – in that there is no certainty in that value and perhaps some evaluation period would be needed or one might need to create an "option of having the assets available for as-yet-unknown uses that may emerge in the future" [1].

An important aspect of the value is an estimate of resources which might be attracted, for example commercial payments for use or advertising revenue or academic value; alternatively the value may be in terms of penalties which might be avoided, for example legal penalties or the costs of replacing the objects if lost.

The digital object may be of value because it cannot be re-created (for example evidence of Climate change) or because of the cost of (re-)creation of the object (for example the data gathered by the LHC).

The digital object will probably be more useful to one type of user community than to another, and this may change over time.

Rights may be associated with the digital objects, perhaps arising from the value or potential value of the digital object.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
How to assign value to digital objects	BRTF report [6]	D36.2	
	LIFE project	LIFE tools	
		DP Impact [10]	
Calculating the Value of Digital Objects at Risk	Tessella DVAR	DVAR calculator http://preservica.com/resource/using-the-dvar-calculator/	
Value vs. legal mandate as main driver for implementing DP practices	APARSEN WP36 DP Impact	D36.1 DP preparedness in Scientific libraries. D36.2 Exemplary business cases DPImpact chapters 4 & 5	Description of currently in place policies and practices in pioneering DP practices
In terms of cost models, may be covered by impact or benefits assessments which link to value. See KRDS model.	APARSEN WP32 cost modelling	Further assessment of D32.1 and D32.2 could be undertaken to provide training, consultancy – not strictly although could be considered a service perhaps, but we could provide guidance and advice on best tools available	

		given specific situations to assess value of DP. This would, however, need to be investigated further. In terms of tools, we could refer to specific models which are currently available.	
Persistent identifiers as enablers of value in scientific data e-infrastructures (SDIs) in particular for developing value-added services on top of scientific data and contents. These services deal with many aspects of the e-science landscape including data and information access, knowledge discovery, Citability, quality assessment and provenance.	DIGOIDUNA ODIN	DIGOIDUNA final report, in particular SWOT analysis and recommendations http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/digoiduna.pdf ODIN deliverables available at http://odin-project.eu/project-outputs/deliverables	Preliminary models and results on data exchange improvements and workflows Results will be published soon at http://odin-project.eu/project-outputs/

Gaps
The major gaps are related to the identification, classification and quantification of the benefits and impacts the usage of preserved digital objects may produce. In turn this also requires the identification of the targeted users of the preserved objects (e.g. researchers or students) and the customers/purchasers of such collections (e.g. companies or universities) Very limited work done on value in terms of cost models.
Persistent identifiers. 1) A common agenda among key stakeholders towards the design and implementation of a governance model and an integrating infrastructure for managing PIs in SDIs in which technological, economic, social and political factors are taken into account; 2) Common policies on the governance of PIs and integrating technical solutions; 3) Mobilization of technical, human and financial resources to trigger a wider demand of usage and exploitation of e-Science results based on PIs. 4) Suitable business models and organizational mechanisms to ensure long-term sustainability of the implemented solutions. Lack of interoperability between identification systems for data on one hand and for contributors on the other hand. This can cause 1) the inability to follow interconnections between datasets and contributors as a method for data discovery; 2) the inability to share and connect identifiers of contributors and authors between different user communities; 3) inability to uniquely identify datasets attributed to a particular contributor and contributors to a particular dataset.

4.9 GOVERNANCE

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Collect good governance structures and data policies	APARSEN WP35	D35.1: Exemplar good governance structures and data policies	
Policies for preservation	WP35: recommendation	D35.1	
	InterPARES Trust and conflicting rights (2013-2017)	Framework for policy evaluation (under definition)	
DP practice implementation, a managerial issue	APARSEN WP36	D36.2	Strong presence of DP implementation based on projects, not as a general policy

Gaps
Fragmentation of policies involved for digital preservation and related taxonomies
Lack of orientation in the conflicting environments relevant for ensuring data access and data management
Competitive Integration of DP as a cut across activity within overall organisational structures and workflows

4.10 BUSINESS CASE

The Business case is needed to justify the resources required for preservation. It will be built on the value of the objects and the costs involved in their preservation.

It will probably be embedded within a particular governance system and business models, defining risks about assumptions of value. There will almost certainly be options for trade-offs between costs, risks and capabilities

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Creation of business case	APARSEN WP36	Deliverables D36.1 and D36.2	
	BRTF [6]		
	SHAMAN		
	SPRUCE	Business case toolkit	
	DP Impact [10]		
	Use a risk based assessment of digital assets to determine archive strategy	http://preservica.com/resource/using-the-dvar-calculator/	
Business case as the backbone of sustainability conditions	WP36	D36.2, especially chapters 3 & 4	Lack of business cases based on quantitative evidence of usage of preserved assets, actual quantifiable benefits and thus value measurements
Ingest fees	PANGEA	PANGEA has long standing experience in calculating ingest costs which are perceived as the cost driver of the project.	

Gaps
Organisational culture modernisation

4.11 COSTS

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
How much will preservation cost	APARSEN WP32	Deliverables D32.1, D32.2 These show that the existing cost models are so far untestable however they provide a view based on ISO 16363 of cost areas that should be considered	
	Various projects investigated in WP32		
	DP Impact [10]		
Cost models are rarely used to assess the costs of DP activities within organisations. Organisations use internal resources or knowledge.	WP32	Work carried out and provided in D32.1 and D32.2 could be (and has been used) for training, consultancy – not strictly although could be considered a service perhaps, but we could provide guidance and advice on best tools available given specific situations. This would, however, need to be investigated further. In terms of tools, we could refer to specific cots models which are currently available.	
Many different cost models being used and being newly invented The real issue is turning cost models into cost recovery models	Work in the 4C project (Neil Grindley, JISC) And in RDA (Data Publishing - cost models) (4C project: http://4cproject.eu)	4C wishes to build a registry where you can look up cost models as used by others RDA is working on cost recovery models, also taking Public-Private-Partnerships into consideration.	
Ingest costs	PANGEA	PANGEA has long standing experience in calculating ingest costs which are perceived as the cost driver of the project.	
Costs in scalable preservation systems, including the use of cloud storage			

Gaps
Gaps identified as very few cost models available and those which are available tend to specific to the creators needs and more significantly to the organisation creating the model.
Some gaps may be filled by the 4C project, a coordination action on a 'Collaboration to clarify the costs of curation'. The Curation Costs Exchange tool may be of relevance (see: http://www.curationexchange.org) Reliable cost models for support of persistent identifiers.

4.12 BUSINESS MODEL

Decisions about the mix of sources providing the financial resources required for implementing and operating the preservation business process will be based on the characteristics of the users and customers base (the target groups), the competition in the provision of the preserved assets as well as in the nature and dynamics of the formulated business case.

The resources may be used at the very start to create new digital objects, which will presumably have been created for a specific purpose and which then may be either disposed of or be preserved.

A selection process will be needed to decide what is to be preserved. This will presumably be based on business case and risk considerations. It may also depend on the interest of other possible curators of the information.

This financial resourcing may be (perhaps should be) part of the budgets needed to create the digital objects. However some or all of the objects created may be disposed of rather than preserved.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Business models for preservation in general	APARSEN WP36	Deliverable D36.2	
Going beyond financing DP from budgets and projects	APARSEN WP36	D36.1 financial sources for DP D36.2 especially section 2.6	DP being primarily funded by short term budgets
Sustainability and Revenue Models for Online Academic Resources (10) DP Impact [10]			

Gaps
Business modelling for DP in scientific libraries and scientific data curators organisations

4.13 SELECTION

A selection process will be needed to decide what is to be preserved. This will presumably be based on value, business cases, costs and risks. It may also depend on the interest of other possible curators of the information.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
	Records Management processes	Best practice guidance for selection and disposal	
Policies for selection	APARSEN WP35: recommendations	Deliverable D35.1	

Gaps
Lack of awareness of stakeholders on the complexity and centrality of data appraisal

4.14 AUDIT

There are a number of options for gaining some kind of recognition of quality of preservation. This may be of importance in terms of funding, for example providing a competitive advantage for provision of preservation capabilities.

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
How to obtain some level of audit and certification for trustworthiness in terms of preservation.	APARSEN WP33 test audits	D33.1B	
		ISO16363 and ISO16919	International reviews of the standards
Accumulating evidence about quality of preservation activities	SCIDIP-ES certification toolkit	Software	
Creating a standardized approach for supporting quality and feasibility of documentation	APARSEN WP24	D24.1 and D24.2	

Gaps

4.15 GLOSSARY

A common vocabulary as a component of a Common Vision of Digital Preservation is very important to avoid confusion, especially when communicating with and between a number of different disciplines/cultures. There are a (growing) number of candidate vocabularies.

The glossary is being developed following the normal pattern within APARSEN by collecting information, in this case of the terminology in use based on a number of available preservation glossaries, and then bringing in some coherence and consistency.

This collection started simply as an alphabetical list, collecting terms from many existing glossaries/vocabularies, but as a following step an organisation/grouping based on SKOS (Simple Knowledge Organisation System) has been put in place. This is available on the public website at http://www.alliancepermanentaccess.org/index.php/knowledge-base/dpglossary/. In addition we will ensure that we work with new developments such as the UNESCO related multi-lingual database at http://www.web-denizen.com/.

The origins of the various terms are tracked e.g.:

- OAIS = OAIS glossary (also in http://www.alliancepermanentaccess.org/index.php/knowledge-base/member-resources/digital-preservation-glossary/)
- DPC = http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts
- APARSEN = a term that has been introduced by one of the APARSEN project team members
- ANZ = http://archives.govt.nz/advice/continuum-resource-kit/glossary/definitions-full-list
- SNIA = http://www.snia.org/education/dictionary
- INTERPARES = http://www.interpares.org/ip2/ip2_terminology_db.cfm

This glossary shows the wording convention within the APARSEN project and network to avoid misinterpretation when it comes to discussing technical terms and concepts within our collaborative work efforts.

We are using OAIS as the fundamental vocabulary, with additional terms from other vocabularies for new concepts as needed. Where there some overlap in meaning, instead of simply building a collection of terms and their definitions from various sources, a SKOS approach has been used to indicate relationships (broader, narrower, related etc.) between the terms [4].

For example:

Access_Rights_Information	
In scheme:	OAIS
Definition:	The information that identifies the access restrictions pertaining to the Content Information, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the Submission Agreement, related to both preservation (by the OAIS) and final usage (by the Consumer). It also includes the specifications for the application of rights enforcement measures.
Pref Label:	Access Rights Information
Narrower term:	*Permissions*
Narrower term:	*Restriction*
Narrower term:	*Access_privileges*

Narrower term:	*Access_INTERPARES*
Narrower term:	*Access_privileges_code*
Narrower term:	*Access_restrictions*
Narrower term:	*Access_restrictions_code*
Narrower term:	*Access_Rights*
Narrower term:	*Access_rights_INTERPARES*
Broader term:	*Preservation_Description_Information*
Related term:	*Access*
Related term:	*Digital_Rights_Management*

Where glossaries have used the same name for a concept, we attach the name of the originating glossary to make each name unqiue.

Very importantly each term can be referred to by a URI, for example the term "Access_Rights_Information" may be referenced as

http://www.alliancepermanentaccess.org/index.php/knowledge-base/dpglossary/#Access_Rights_Information

This will allow its use in external documents or websites or ontologies or linked data. We anticipate that other lists of terminologies related to digital preservation will be produced and these we will attempt to integrate into this SKOS glossary.

The HTML for the web page and the RDF file which is also available on the page are created using an Excel file with macros.

The relationships between terms express the views of the APARSEN members but others can submit their ideas for additional relationships, corrections to the existing relationships and additional terms via a form attached to the glossary page.

4.16 FORMAL QUALIFICATIONS AND TRAINING IN DIGITAL PRESERVATION

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Accredited Continuing Professional Education (CPE) in DP	WP42/WP43, APARSEN Online Training Portal (OTP) including repository of training materials, registry for formal qualification opportunities, online course learning management system (LMS); DP training and qualification services	GLOBIT Educational Portal, GLOBIT training and qualification services; APARSEN OTP, APARSEN D42.1 survey of existing initiatives and curricula, training materials collated/ produced by APARSEN, curricula for selected disciplines	
Academia/ Higher Education (HE) curricula and courses	See above	See above	

Gaps
Structured descriptions of opportunities; There should be a shared, standardized vocabulary for the description of curricula and courses; Digital Preservation as a discipline is lacking generally accepted competence/job profiles for formal qualifications; no up-to-date register of for formal qualification opportunities of European scope

4.17 EXTERNAL COMMUNICATION AND AWARENESS RAISING

Asset base			
Issue	**WP/Project/Tools/Services**	**Asset**	**Evidence**
Community of stakeholders in DP	WP44 Interactive Map (IM) of Stakeholders in Digital Preservation including underlying stakeholder register; APARSEN Online Training Portal (OTP) with community features; APARSEN Facebook account; APARSEN newsletter	GLOBIT Association Management systems and services, GLOBIT Customer Relationship Management (CRM) systems and services; APARSEN WP44 stakeholder register; DP initiatives and organisations cooperating with APARSEN; users of APARSEN services; APARSEN newsletter subscribers; organisations gathered in APARSEN surveys	
APA/APARSEN contacts for dissemination and exploitation	See above	See above	

Gaps

5 MARKETING TAG LINES

A number of brief marketing tag lines are presented below.

- Digital preservation forms the bedrock on which our increasingly digital society is being built
- Digital Preservation minimises Information Society amnesia
- Digital Preservation multiplies the opportunities for extracting value of digital content over time
- Digital Preservation minimise the losses of value of digital content over time
- Digital Preservation will ultimately be an standard component of modern management processes
- Digital preservation requires resources and these must be repeatedly justified; we must work together to hone our arguments
- Digital preservation is business-driven and should according to strategy fit and objectives be adequately funded/resourced
- We can work together to find ways to add value to digitally encoded information
- Evidence should form the basis of decisions we take on preservation
- We have many tools and techniques for preservation and we must be able to choose the right ones for each challenge
- Authenticity is often elusive but we have the basis on which the required evidence can be collected
- Repositories should prepare for certification even if they do not undergo formal audits
- The total cost of preservation is uncertain but we can provide the options on which decisions can be based digital curators are the unsung heroes of the information society
- Digital preservation is an important and interesting adventure which requires lifelong learning

6 INTEGRATION OF THE COMMON VISION IN THE VCOE ACTIVITIES

In the context of APARSEN, the Common Vision will provide a unifying base shaping the activities of the Virtual Centre of Excellence, allowing its members to share a view of the challenges of digital preservation, as well as underpinning the offerings of the VCoE. However it is hoped that the Common Vision will enjoy wider influence and acceptance in the digital preservation and data holding communities.

The offerings of the VCoE have been grouped into four headings: Consultancy, Tools, Services and Training. The Common Vision (and the associated the assets – the Knowledge base) underpins all of these, as described next.

6.1 CONSULTANCY

The Common Vision provides the basis of consistent and coherent advice and information about digital preservation. It is underpinned by the asset base which has been collected. Even the information about gaps provides us with a view of the "known unknowns" in digital preservation.

6.2 TOOLS

Tools provide fundamental capabilities such as creation of Representation Information, evidence about Authenticity, building business cases, estimating costs and planning preservation strategies.

The web site [5] contains information about tools and this information is picked up by the SCIDIP-ES [13] "Representation Information Toolkit" [14].

An important question is what tool is likely to be useful for a particular preservation-related task. The information on the website [5] attempts to bring together evidence about the usefulness of particular tools for various types of data. APARSEN D14.1 summarises information about this approach.

6.3 SERVICES

Services are widely applicable ways of sharing information and capabilities, and are used by many tools. They provide a way of sharing information and capabilities. APARSEN D21.1 reports on information collected on a variety of services, mapped to a landscape based on ISO 16363.

Some services can be developed as part of an international effort with other projects and stakeholders (i.e. InterPARES Trust and conflicting rights, International Council on Archives)

6.4 TRAINING

APARSEN D43.1 contains a list of training topics. These are essentially consistent with the Common Vision presented in this document. However here we can present the training material within a coherent, comprehensive and consistent framework.

The training modules will be planned in detail within this framework, illustrated in part in the following table.

Course/Sub-Modu	Course title	Vision for course	Sources - see D11.3	Length (mins)	Stated Goal	Participant	Benefits for participant	Key features for participant
1	Introduction	This set of modules will give an overview of the whole cyclical model from the APARSEN view. It consists of 5 main modules which will themselves provide an introduction to the more detailed course of the same name.		Sum of modules and sub-modules	The participant will wish to take the more detailed courses.	Division directors and middle management responsible for any type of repository	The participant should be made aware of issues that may affect their repositories or may enable their repository to justify their current preservation activities, or enable them to move into new areas in competition with other repositories.	Ideally the issues should include ones they are not familiar with.
1.1	Preservation	The preservation module gives a rapid overview of the issues that need to be addressed for a variety of different types of digital object, together with a brief introduction to how these link to OAIS and basic preservation techniques		sum of sub-modules			The participant will be in a position to see how his/her repository can improve and take advantage of new opportunities	Training should provide a number of examples which provide a context into which he/she can fit current and likely future issues. There should be a number of specific digital objects used in the course, and a display of bit/hex sequences.
1.1.1	Overview	Present a number of digital preservation scenarios showing the problems that have been faced.	User scenarios	15	This will give the participant the ability to compare his/her experiences with those of other organisations. In a face to face course there would be opportunities for interaction		If this is the only module taken it should give the participant an overall impression of the issues that may be of concern.	This may be the only module a higher level manager may view.
1.1.2	Problems in preservation	Presents some specific problems in preservation including issues with data aswell as documents. Most of the threats which are shown later should be covered.	User scenarios	15	The aim is to ensure that the participant can put his/her immediate problems into a broader context, by showing a variety of preservation issues		The participant will see what problems have not yet been addressed in his/her own repository	Module should show a variety of preservation problems which may not have been familiar to participant
1.1.3	Types of digital objects	Shows a number of different types of digital objects, covering many of the types from D14.1. Should also show some of hexadecimal dumps and compare some of the document formats including docx and ods etc.	WP14	20	Opens the eyes of the participant to the different types of digital objects.		The participant should be in a position to understand the problems that need to be addressed if he/she is to preserve new types of digital objects.	participant should show at least some digital objects which participant are not familiar with
1.1.4	Outline of OAIS concepts	A brief overview of the Functional and Information models and their relationship. Also the responsibilities of a repositories will be discussed.		20	participant should understand the reason OAIS has these models.		participant gains much better understanding of OAIS, a very important standard	Module should show the reasons for the various models in OAIS
1.1.5	Threats	This module provides a brief overview of the threats which PARSE.Insight has identified. The participant will be able to compare the issues he/she is concerned about.	PARSE.Insight	15	The participant will be able to identify threats which he/she may not yet have thought about.		participant will be better prepared in terms of the threats to his/her digital holdings	Module should show the various threats and the depth of agreement across disciplines.
1.1.6	Basic techniques	Introduce participants to a variety of preservation techniques including adding RepInfo (including emulators), transformation, hand-over		20	Ensure that participants are familiar with a a number of preservation techniques - at least at a high level.		participant will gain a better understanding of a wide range of preservation techniques.	
1.2	Using	This module will give the participant an overview of the varieties of ways in which digital objects are/can be used.		20	participants should appreciate that there are many different types of digital objects, used in many different ways, many of which they have not yet had dealings with		The participant has a better understanding that there is a range of techniques	Must include an example for ldocuemnt world as well as example of science data. The examples should not need detailed discipline knowledge
1.3	Value	This module provides participants with examples of ways in which value can be attributed to various types of data		15	participants should be able to think in general terms about the value of their own digital holdings		participant will see how others value their holdings and have a better appreciation of their own repository's value.	Module should show some specific and understandable ways that value has been assigned i.e. don't just say "it is valuable" but explain why and to whom and how this might be quantified (if possible)
1.4	Business cases	Provide participants with examples of business case		15	participants should be able to understand the need for business cases		participant will gain a better appreciation of the types of business cases that are used.	Module should contain a few different vbusiness cases.
1.5	Business Models	General discussion of types of business cases		15	participant should be able to think in general terms of the business case which supports their repository		participant should gain an idea of the possible business models	Must make a clear distinction between business cases and business models

Figure 2 Extract from plans for training "pills"

7 REFERENCES

[1] Life Cycle Models for Digital Stewardship, by Bill LeFurgy, 2012, see http://blogs.loc.gov/digitalpreservation/2012/02/life-cycle-models-for-digital-stewardship/

[2] Review of Data Management Lifecycle Model by Alex Ball Univ Bath, 2012, see http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf

[3] Data Lifecycle Models and Concepts by CEOS, 2012, see http://www.ceos.org/images/DSIG/Data%20Lifecycle%20Models%20and%20Concepts%20v13.docx

[4] APA/APARSEN digital preservation glossary - see http://www.alliancepermanentaccess.org/index.php/knowledge-base/dpglossary/

[5] Evidence-based digital preservation tools repository, see http://www.alliancepermanentaccess.org/index.php/knowledge-base/tools/tools-for-preservation/

[6] BRTF "Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information". Blue Ribbon Task Force on Sustainable Digital Preservation and Access. Final Report, February 2010. See http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

[7] Beagrie, N., Eakin-Richards, L. and Vision, T. "Business Models and Cost Estimation: Dryad Repository Case Study". Austrian Computer Society (OCG), 2010. See http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/beagrie-37.pdf

[8] Guthrie, K., Griffiths, R. and Maron, N. "Sustainability and Revenue Models for Online Academic Resources" An Ithaka Report, May 2008. See http://sca.jiscinvolve.org/wp/files/2008/06/sca_ithaka_sustainability_report-final.pdf

[9] European Commission, Information Society and Media Directorate-General. "The Future of the Past – Shaping new visions for EU-research in digital preservation". Luxembourg, May 2011. Cultural Heritage and Technology Enhanced Learning. See http://cordis.europa.eu/fp7/ict/telearn-digicult/future-of-the-past_en.pdf

[10] Inmark Estudios y Estrategias, S.A. "DP Impact. Socio-Economic Drivers and Impact of Longer Term Digital Preservation". Final Report, June 2009. European Commission. See http://cordis.europa.eu/fp7/ict/telearn-digicult/dpimpact-final-report.pdf

[11] Knowledge Exchange, c/o Danish National Library Authority. "Report on Knowledge Exchange workshop. Price of keeping knowledge: financial streams for digital preservation" Amsterdam, 17 January 2013. Large member base and diverse sources of income, see http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fPrimary+Research+Data%2fWorkshop+Price+of+Keeping+Knowledge%2fReport+KE+workshop+Price+of+Keeping+Knowledge+17+January+2013.pdf

[12] Data Resource Broker www.gael.fr/drb

[13] SCIDIP-ES project website http://www.scidip-es.eu

[14] SCIDIP-ES interactive platform for information about and access to the tools and services available from http://int-platform.digitalpreserve.info/

8 ANNEX 1: SCENARIOS

A number of scenarios are being collected in this section. The aim is to have a variety of scenarios, some generic and some very organisation/sector specific, for example scientific repositories, national archives, commercial organisations with archives etc.

A common format will be developed to bring out similarities and differences between organisations. These will be used to check the Common Vision and as material for the training and consultancy offered by the VCoE.

8.1 SCENARIO 1

A repository manager must create a sustainable plan for the preservation of a dataset. Steps include:

Activity	WP
Arrange training on preservation of datasets covering all the elements in APARSEN - using consistent terminology Check on consistent vocabulary	WP42/WP42
agree on the designated community for the dataset	WP25
create preservation strategy	also SCIDIP-ES
-- identify preservation tools which are likely to be appropriate for this type of data	WP14
-- identify services which can help in preservation -- Identify standards which can help in preservation	WP21 and SCIDIP-ES WP13
-- identify optimal storage system to use	WP23
-- identify persistent identifiers to use	WP22
collect evidence of authenticity	WP24
-- integrate provenance which came with the dataset into the repository's provenance tracking system. Essentially preservation (interoperability and use) of the provenance information	WP24
-- check fixity information	WP24
collect evidence about the quality of the data	WP26
estimate cost of preservation and negotiate with funders about quality of service and acceptable risks	WP32
create business case	WP36
-- identify potential wider set of users and potential benefits	WP25/WP36
attempt to obtain trustworthy status in order to help attract further deposits of data	WP33

8.2 SCENARIO 2

The User accessing the SCIDIP-ES data is able to retrieve the data itself and all the additional information needed to understand, analyse and process it. Steps include:

Activity	WP
User, belonging to a Designated Community, searches for data required.	
User selects a specific dataset	
Representation Information which enables the dataset to be understood by the Designated Community is made available (if possible)	WP25 plus SCIDIP-ES
-- if there are gaps in the Representation Information then new RepInfo may be collected by the Repository	WP14 plus SCIDIP-ES
-- the data may be transformed to make it easier for the user	WP21 plus SCIDIP-ES
-- the data plus RepInfo may be packaged together of reuqired	SCIDIP-ES
-- check fixity information	WP24
-- identify persistent identifiers to use	WP22
Provide evidence of authenticity	WP24
Provide any available evidence about the quality of the data	WP26

8.3 SCENARIO 3

Changes take place in the preservation chain. The changes could affect the user community, the data itself (new processors, new formats), or the associated RepInfo (new auxiliary files available, new calibration campaigns). The Preservation Network Model (PNM) and/or the RepInfo Network associated to the data must be updated correspondingly. Steps include:

Activity	WP
The Data Producer/manager is informed (e.g. by data users via the Orchestration Service) that some changes in Designated Community (DC) knowledge could limit or prevent the ability to retrieve or to use a dataset properly in the future. He uses services and toolkits to update the archive and to ensure the continuity of data usability. Changes could happen in the following contexts (the following examples are only an indication): SubCase 1 - Change in the DC (e.g. new DC) SubCase 2 - Change/Error in the RepInfo Network SubCase 3- Change in technology (e.g. sw obsolescence, IPR, policies) SubCase 4 - Evolution of the PNM	SCIDIP-ES
Data Manager accesses the PNM and performs an analysis of the new situation. Dependencies to be investigated by means of the Gap Manager and the RepInfo Registry can include:	SCIDIP-ES
-- Changes in the Designated Community, which means new DC to be added; changes in the DC knowledge. By means of the Gap Identification Service and the RepInfo Registry Service, he updates modules and dependencies.	WP25 plus SCIDIP-ES
-- Preservation Network Model needs to be updated, using the Preservation Strategy Toolkit and the RepInfo Registry Service.	SCIDIP-ES
-- Changes in the RepInfo network, using the RepInfo Toolkit and the RepInfo Registry Service.	SCIDIP-ES
-- Change in technology: some elements inside the PNM become out of date, are no longer maintained and some sort of migration/update and/or replacement is needed. The data manager performs the changes needed and uploads the updated PNM and a required RepInfo. These steps use the Gap Identification Service, the RepInfo Registry Service, the Preservation Strategy Toolkit, the RepInfo Toolkit, the Storage Service and the Packaging toolkit. The new AIPs are stored by the Storage Service.	SCIDIP-ES
-- May need to Transform the data	SCIDIP-ES
Collect evidence of authenticity	WP24
Identify additional business case opportunities	WP36

Further scenarios are available on the public website[3] and a selection will be appended here.

[3] http://www.alliancepermanentaccess.org/index.php/knowledge-base/tools/tools-for-preservation/list-all-user-scenarios/

8.4 SCENARIO 4

1. Introduction: different sectors and long-term preservation

A number of rationales to put effort in long-term preservation of digital objects can be distinguished. These rationales are related to the mission and strategy of organisations. A number of sectors, in which organisations are active, can be distinguished.

- Research data centres (data archives, university repository)
- Cultural heritage institutes (libraries, museums)
- Corporate organisations (companies)
- Government archives (national archives, city archives)

(Research data centres, cultural heritage institutes and corporate organisations are represented in the APARSEN project).

The main rationales for research data centres, such as data archives, to put effort in long-term preservation of digital objects are to create durable links between research outcomes and related data, to enable the replication of research and to enable secondary analysis of existing data sets.

For cultural heritage institutes long-term preservation is relevant to secure future access to digitized and born-digital cultural heritage objects. Examples of future access are the creation of (digital) exhibitions, the education of students and to inform the general public.

Long-term preservation for corporate, commercial organisations implies the management of digital assets that are essential for its core business. In order to protect its market share this sector by default is keen on keeping the access to digital objects closed and controlled. This in contrary to the other sectors, such as the research data centres, for which open access and transparency is obvious.

The archives or government bodies, such as the national archives, are the memory of the government and the activities concerning the long-term preservation of digital objects are related to provide durable access to this memory.

Sector	Rationales for long-term preservation of digital objects
Research data	1. Enable linking between research results and related data sets 2. Enable replication of research 3. Enable secondary analysis
Cultural heritage	1. Enable future access to digital cultural heritage 2. Enable the creation of new cultural heritage objects
Corporate world	1. Enable the management and protection of digital objects related to the creation of products and services of the corporate body.
Government	1. Enable the reconstruction of the proceedings of the public administration

2. Scenarios relevant for the long-term preservation of digital objects for research data centres

(A scenario is “a story about what happened in the future”)

In this section a number of actions are described that support the rationales or scenarios for putting effort into long-term preservation of digital objects as given in the table above. In case not stated explicitly the activities are relevant for all three scenarios.

Activity

01. Persistent identification

Persistent identification involves the creation, management and connection of persistent identifiers to data objects (or fragments of data objects).

02. Trusted archiving / Managing authenticity

Research data objects are stored in a Trusted Digital Repository (TDR). A TDR is a repository whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future.

03. Linking / referencing

In order to enable the (re) use of digital objects stored in a TDR it must be possible to create links / references to data objects or parts of them. The PI is an important part of this link/reference. Other components of the link/reference are bibliographic references to the digital objects.

04. Licensing

Usage licenses contain the conditions for (re) usage of data sets. User license management is an important action related to the long-term preservation of digital objects as it determines who, and in what conditions is gained access to the data objects. The licenses should obey a number of regulations such as privacy regulations and codes of conducts. The Open Access licence paradigm is the de-facto default license norm. But other forms of licensing are also used if applicable.

05. Searching / browsing

In order to gain access to data objects for (re) use and/or replication, it is obvious that the data objects must be found. For this search and browse functionality is required. TDR's consist of these functions.

06. Migration of data objects

Migration concerns the conversion of data objects into newer formats in order to avoid format obsolescence. Data objects formatted in obsolete formats are not accessible anymore. Migration activities typically are part of the proceedings of a data archive. Migration might be necessary to make research replication and (re) use of data sets possible, as it is possible that the original file format is not accessible anymore.

07. Version management

Version management is an activity within the competencies of a research data centre. Research data sets can be updated while an earlier version is (re) used in a given research project. Both the original and the updated version must be made available in the TDR.

08. Quality assessment

Quality assessment concern activities related to the correct data applicability and interoperability. Examples are harmonization of metadata elements in order to facilitate harvest systems and a facility to review the quality of data objects so users can assess its quality based on a judgement given by peers.

09. Storage management

Activities related to the storage of data objects concern the development and maintenance of data storage facilities. Determinants in this action are issues such as scalability, speed, cost, etc.

10. Business case definition

The cost/benefit issues of long-term preservation of digital objects concern a wide range of activities. The goal is to define a business case that fits the purpose and available resources of the TDR and its users.