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

D11.3 Report on a Common Vision of Digital Preservation: Progress to Year 3

INTERMEDIATE VERSION, SEPTEMBER 2013

Document identifier:	**APARSEN**-REP-D11_3-01-0_3
Due Date:	2013-12-31
Submission Date:	2013-09-30 for this intermediate version
Work package:	WP11
Partners:	ALL
WP Lead Partner:	APA
Document status	Draft

Abstract: This intermediate version of deliverable D11.3 presents a snapshot of the common vision of digital preservation of the APARSEN project, based on input from most of the work packages within the project. The vision is presented with reference to a model of value of digital preservation, and identifies the assets, gaps and evidence accumulated by APARSEN for each of the components of the model.

Delivery Type	Report
Author(s)	David Giaretta and all partners
Approval	
Summary	
Keyword List	
Availability	☒ PUBLIC

Document Status Sheet

Issue	Date	Comment	Author
0.1	2013-08-22	Initial draft	David Giaretta/ Matthias Hemmje
0.2	2013-09-09	Clarifications and details added	David Giaretta/ Ruben Riestra
0.3	2013-09-30	First round of contributions from partners added	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

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 in Europe. The organisation of the work of APARSEN into topics and streams is one 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 proceeds in the project towards a vision of what must be done, not only of how it can be structured. As the Description of Work 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.'

The present document reports on the state of development of the common vision up to Month 36 of the project. It will be followed by a final annual report.

The other key component in the strategy to reduce fragmentation is the creation of the Virtual Centre of Excellence (VCOE). This is the subject of a separate deliverable (D11.4). The VCoE will be the legacy of APARSEN that will live on beyond the end of the funded project and will be guided by the common vision.

In the FP 7 Work programme of the European Commission, some project consortia have been urged to commit to a Centre of Competence (CoC) or a Virtual Centre of Excellence (VCoE) as one of the outcomes of their projects. The difference is defined by a CoC having expertise centrally which is made available to members and others whereas the expertise of a VCoE resides mostly in its members. Experience shows that the difference is not clear cut.

It is important to note 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.

This section sets out the methodology used to reach the common vision, which will be exposed externally to allow for engagement with the wider community. This engagement will help to validate the vision and ensure its broad acceptance beyond the APARSEN consortium members.

The Description of Work sets out some principles for the development of the vision—for example, that it should be 'clarified and embedded through open discussion, extended and vigorous argument and thorough testing'—but this itself does not define the steps that need to be taken.

The methodology covers the remaining time of the APARSEN project, and this document includes a first draft of the common vision which will be made available in September 2013, as requested by the review panel at the review meeting in March 2013.

This initial draft will form the basis for the remaining stages of integration, described in section 1.5, which will bring in the results generated in the remainder of the project, testing and expanding the details of the common vision as we proceed.

1.1 REPRESENTING THE COMMON VISION

In order to allow engagement by stakeholders the common vision must be committed to writing in an understandable and usable form at various defined stages, subject to the following constraints. Such documents, representing snapshots of the vision, are referred to below as *vision statements*.

- The vision statements must be sufficiently concise and punchy that they encourage engagement by parties outside the APARSEN partnership.
- The vision will evolve over the lifetime of APARSEN (and beyond), as outlined in this methodology.
- The vision should have some elements of aspiration—a view of the future and what needs to be done to reach it—and could therefore be labelled a 'manifesto' for digital preservation, not only a presentation of the facts as APARSEN has found them.

- The vision statements should be as precise as is reasonable, to allow debate and reaction, but should not over-specify.
- The vision statements should be backed up by evidence, to the extent that is needed to give confidence in their soundness. We have many of the leading organisations within our consortium, thus which can give their testimonials, good pieces of advice on lessons learnt, 'real world' vs. theoretical discussions, etc.
- The vision statements will likely reflect the perspectives of different types of organisation/stakeholder, thus enriching the overall content included in the common vision.
- The vision statements should support and reinforce strategic alliances among stakeholders involved in the digital preservation environment by providing scenarios and initiatives for cooperation

In addition to the vision statements, supplementary material will be prepared to help with the consultation and feedback process. These support elements include:

- questionnaires for probing reactions, say on a scale from 'Total disagreement' to 'Total agreement'
- scenarios for illustrating the issues in the vision (see below)
- presentation material for a consistent presentation of the vision at conferences and other external events, coordinated with the Spreading Excellence stream
- more extended summaries of evidence backing up the vision statements

1.2 USES OF THE COMMON VISION

The common vision will be put to use in APARSEN to ensure coherence of the approach and outputs of work done where different strands of the project's activities are brought together. These include:

- creation of dissemination material
- proposal and organisation of external events
- creation of training courses
- definition of curricula for formal qualifications
- definition and evaluation of suggested services and policies

Furthermore, and more importantly, the same coherence will be expected when it comes to defining the range or portfolio of value added services to be offered through the VCoE such as consultancy, advice and technical services. At one level, even the consistent adoption of a common glossary will provide a unifying force, but the vision goes further than that in covering the state of the art, expected developments, etc.

1.3 MECHANISMS AND INPUTS

A number of procedures to give input to the common vision are envisaged, and some have already been initiated. They are:

- The results of the work done within APARSEN itself
- High-level scenarios that illustrate the issues in an easily understandable way. Scenarios which actually relate to user application domains and if possible, to user and customers' requirements for the VCoE definition, as development of these scenarios and the related policies or services will be instrumental to prepare the ground for the VCoE. Scenarios will be developed both top-down and bottom-up: the former will ensure coverage of important areas, while the latter will be based on real needs and test the applicability of the APARSEN outputs and the conditions for securing its sustainability. Some examples of such scenarios are already included in an annex to this document and further examples are available on the public website.
- The common glossary of digital preservation terms
- Focussed workshops for APARSEN participants like the one held in Brussels in June 2012
- External consultation and feedback on the evolving vision, including with other projects in the area and existing networks/centres of competence; this should specify the stakeholders that

should be engaged (which are the relevant stakeholder communities with which we would like to agree on a common vision?)

1.4 WORKING STRUCTURES

Within APARSEN, the following structures are used to manage the development of the vision

- A Working Group within WP11 with general oversight of the vision; this group has already been created but its membership will evolve to cover all kinds of stakeholders represented within the project (universities/research institutes, vendors, national libraries, big science, research data archives, membership organisations, industrial design/engineering)
- The Working Group can launch sub-groups or task forces on specific topics and then integrate results from there or from WPs in a clear picture
- Integration of results of Work Packages in topic areas (trust, sustainability, usability and access), through involved partners' work in WP11, overseen by the Working Group
- Possibly focussed subgroups on special themes, to be created as necessary

1.5 PROCESS AND TIMELINE

The basic process is that the Working Group produces a series of vision statements following each stage of integration. These documents will cover:

- The context of the topic being integrated (i.e. what the landscape looks like now)
- The findings or developments made by APARSEN in this area
- Evidence to support the vision statements
- Aspirations for the future, based on gaps identified

The integration of each topic takes place after the completion of all deliverables from the WPs in the topic. It is within the scope of WP11 and will involve the following tasks:

- Review of major findings of each WP within the topic
- Review of scenarios: what do the common vision scenarios indicate about the needs in the topic area and the conclusions of the WPs?
- Gap analysis: what areas are inadequately covered? Where is further research or development needed?
- Cross-WP linking: how do the WPs connect together to cover the topic area?

This refers to the integration of individual topics within themselves; the overall integration across topics is part of the common vision development itself, building on the topic integrations in succession.

These draft documents will first be reviewed by the whole consortium, on a defined timetable, for critique and additions. This participative procedure will be key to ensure sufficient commonality. The vision statements will then be released outside APARSEN consortium for wider consultation. This may be done by presentation at selected conferences, by soliciting input from particular stakeholders, and by open public consultations. An annual workshop could be a way to disseminate the common vision even after the end of the APARSEN project. The vision statements will be revised and will also be summarised in the formal deliverables due at the end of each year. As the vision statements form a series they are identified as **VS1**, **VS2**, ….

The overall timeline of activities and the main players in each one is depicted in the following table

Month	**Activity**	**Responsible**
By end of June 2013	Development of scenarios to underpin vision Integration of topics 'trust' (WP14, WP24, WP26, WP33) and 'sustainability' (WP21, WP23, WP32, WP36)	All partners, coordinated by Stream Leaders Coordinator and Stream Leaders with involved WPs
1 June–31 July 2013	Working Group produces draft **VS1** and supplementary material	Coordinator as WG chair
1 July–30 Sept 2013	Internal feedback on **VS1**	All partners, coordinated by Stream Leaders
30 Sept 2013	**VS1** available (in the draft of this document) for external consultation, including by review panel	
1 Aug–30 Sep 2013	Integration of topic 'usability' (WP16, WP25, WP27)	Coordinator and Stream Leaders with involved WPs
1 Oct–31 Dec 2013	Presentation and feedback on **VS1** externally Working group produces draft **VS2** and supplementary material	Stream 4 leader? Coordinator as WG chair
1 Dec 2013–28 Feb 2014	Internal feedback on **VS2**	All partners, coordinated by Stream Leaders
By end of Dec 2013	Production of deliverable D11.3 'Report on a common vision of digital preservation: progress to Year 3'	Coordinator
1 Dec 2013– 28 Feb 2014	Integration of topic 'access' (WP22, WP31, WP35)	Coordinator and Stream Leaders with involved WPs
1 March–31 May 2014	Presentation and feedback on **VS2** externally	Stream 4 leader?
1 Feb–30 Apr 2014	Working Group produces draft **VS3**	Coordinator as WG chair
1 Apr–30 Jun 2014	Revisit and refine scenarios Internal feedback on **VS3** and supplementary material Internal workshop on common vision as a whole	All partners All partners, coordinated by Stream Leaders Coordinator
1 July–30 Sept 2014	Presentation and feedback on **VS3** externally	Stream 4 leader?
By end of Dec 2014	Production of deliverable D11.5 'Report on a Common vision of digital preservation'	Coordinator

2 APPROACH

When implemented in productive operations conditions, i.e., incardinated in the business processes of those organisations willing to secure future access to digital assets, digital preservation becomes a truly economic activity. As recognised by the BRTF. As such, preservation requires continuous flows of financial resources. Securing that such flows are allocated to this activity, either by pure funding or by a more complex business model, require a cost-benefit type of justification, given the fact that preservation must compete to other uses of these usually scarce resources.

This raises some fundamental questions: "Why and Who pays for preservation?" The relevance of these questions is not trivial at all. The lack of proper answers to these questions may have likely contributed to the decision by the EC not to keep funding research into digital preservation in H2020.

In parallel, there is also a much wider (probably global) demand that the research that has been carried out, including that done within APARSEN, is put to practical use, creating benefits and thus impacts of diverse nature – economic, societal, scientific etc – from what is preserved . This makes value, in particular ways to increase value -i.e. Value Propositions, and business cases- extremely important considerations to be included in the overall approach to digital preservation. There are technical and organisational issues which must be addressed to complete the picture.

The aim of this part of the document is to describe how the various aspects of APARSEN research come together in a Common Vision in terms of a lifecycle model which is driven by value propositions in order to provide the answer to the fundamental questions stated before. This model brings together the research activities of APARSEN, together with a number of other activities, some aspects of which have been the subject of other research projects and others of which (the gaps) have not yet been investigated. It is likely that in future filling these gaps will be driven by "market" demand and sustainability requirements rather than from pure research motivation. This is particularly true when designing the structure and operations of the VCOE. As a matter of fact, the only strong enough justification for the existence of the VCOE, following the main conclusion of the BRTF studies, is the existence of a well identified and relevant (in terms of number and economic capacity) group ("target market") of organisations who explicitly express their needs for third party assistance to secure the access and usage of digital assets in the long run.

Within this context, the role of the Common Vision, graphically depicted below, is to provide a unifying base supporting the offerings of the VCoE, in reaction to the needs and expectations of this demand.

The proposed vision provides unifying concepts and terminology for the value adding products and services to be offered through the VCoE which have been grouped under the terms Consultancy, Tools, Services and Training, expertise for which come in large part from the members of the VCoE.

Integrated into the Common Vision are the topics into which the APARSEN research has been grouped named Trust, Sustainability, Access and Usability. As will be seen below these are not given equal prominence in the common vision because they were designed as convenient groupings of the APARSEN work packages rather than being regarded as fundamental, unique, divisions of digital preservation. The first three will be seen within the preservation block of the common vision while usability is, for reasons which should become apparent, given its own block.

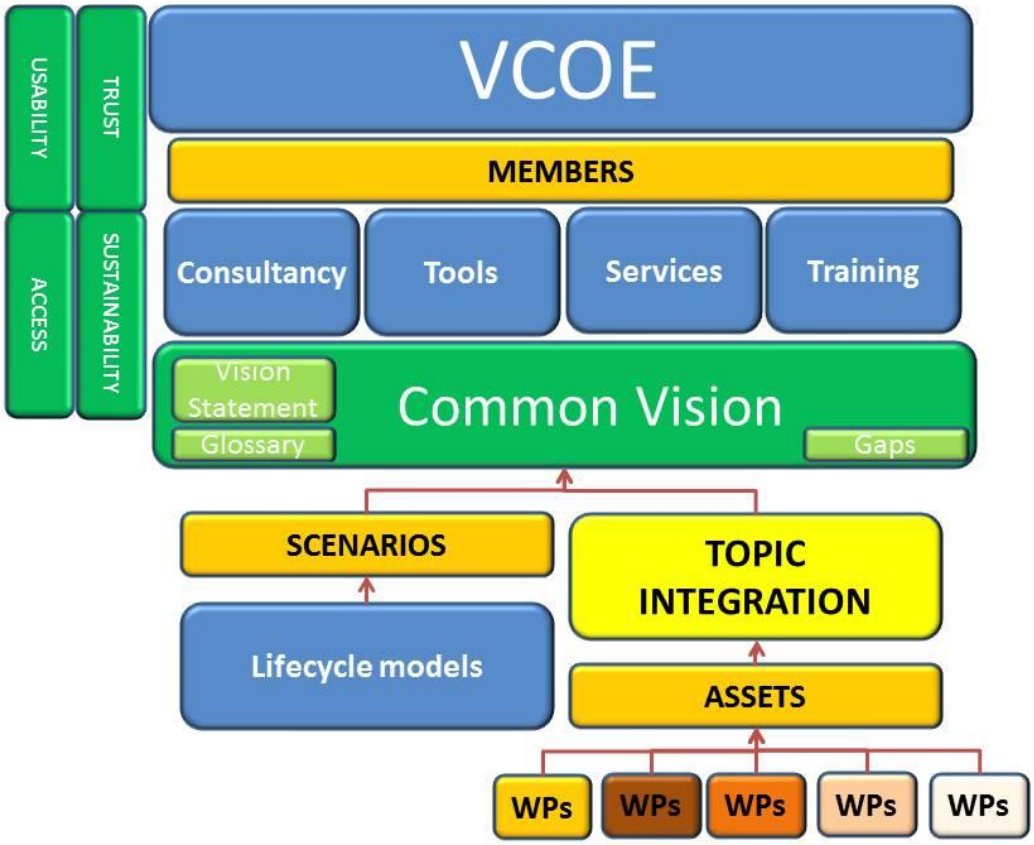


Figure 1 VCoE overview

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.

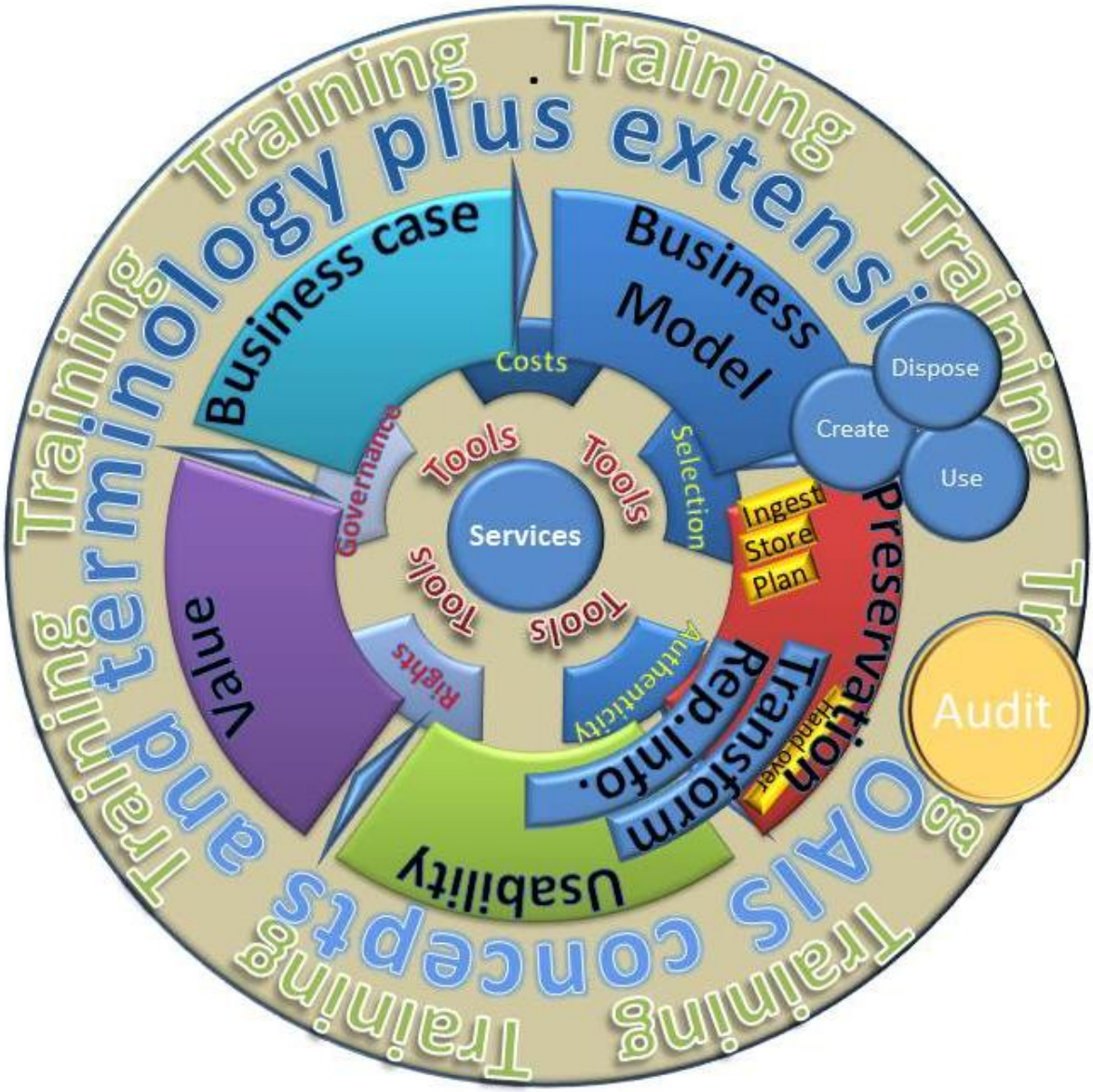


Figure 2 Common vision lifecycle

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

Note that the focus here is on 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.

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 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 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
    - 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 (with possible certification)
- Usability
    - 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; 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 – The portfolio of Value proposition/s will provide the core of the answers to "why and who pay?"
    - 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 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 objects, perhaps arising from the value or potential value of the object.
- Business case
    - This is needed to justify : the need for objects to be preserved, the benefits derived of their usage, the costs involved in the preservation, as well as other resources required for preservation
    - It will be embedded within a particular business model

    - There will almost certainly be options for trade-offs between costs, risks and capabilities
- 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.

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 sub-sections each block of the Figure 2 is described in detail, including 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, which 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**
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
		Mapping of services to metrics of standard for audit/certification of trustworthy digital repositories

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

Evidence	
Source of evidence	**Summary of evidence**
Mapping to audit/certification standard	Identification of gaps as above
Knowledge of involved partners	

4.1 PRESERVATION: INGEST

Moving data from the point of creation or initial use to preservation may be done in a number of ways. The key aim is to ensure that information about the digital objects, which 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 on 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**
Planning hand-over from creators	PAIS and PAIMAS standards	Templates for defining the hand-over process
Definition of Designated Community	APARSEN WP25	CASPAR deliverables and published papers
	SCIDIP-ES GIS service	User feedback
Creation of AIPs	SCIDIP-ES Packaging, plus Preservation Strategy Toolkit and RepInfo Toolkit.	CASPAR evidence, SCIDIP-ES user feedback
	Preservica (Ingest workflow and standalone packaging application)	XIP metadata schema based SIPs, AIPs (& DIPs).
Association of Persistent Identifiers (PI) for people and digital objects on ingest	APARSEN WP22	PI Interoperability Framework
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	Experience in building tools that extract the embedded metadata in from digital files, and produce harmonized warehouses of metadata
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.

Evidence	
Source of evidence	**Summary of evidence**
Ingest workflows in Systems like Preservica, KoLibRI, RODA, Rosetta, etc	Demonstrable ingest workflows available in preservation systems. (See D14.1 for screenshots of ingest workflows from partner test environments).
Demonstrable AIP packages	AIPs formulated to OAIS specification (high level) and held in repositories and test environments
Tool PreScan (www.ics.forth.gr/isl/PreScan)	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
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.

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.

Asset base		
Issue	**WP/Project/Tools/Services**	**Asset**
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
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
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
Common AIP structure to allow true interchange of managed digital repository content, so that the event history / audit trail for the digital object is maintained across all systems, maintaining provenance.
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.
For 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

Evidence

Source of evidence	Summary of evidence
WP14 – Common Test Environments	Demonstrated access to AIPs held in partner test environments, and/or other digital preservation systems, repositories and test environments.
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.

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 which is often heard in the library world is "emulate or migrate". This works well for 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 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 Representation Information. In addition software, including for example emulators, should also be captured.

As noted previously, another part of the 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**
Create plans for preservation	SCIDIP- Preservation Strategy Toolkit and RepInfo Toolkit. APA/APARSEN website Preservica preservation workflows	CASPAR evidence, SCIDIP-ES user feedback. SCIDIP-ES software – open source Cloud-based preservation planning
	SCIDIP-ES orchestration/broker service	Software plus information about possible curators and experts
	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 set of digital object types.
Polic-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
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 tackling these dependencies and managing intelligibility gaps are needed.

Evidence	
Source of evidence	**Summary of evidence**
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).	The prototype will be evaluated by APARSEN partners (in particular those involved in WP25) and external experts.
APARSEN WP14 / D14.1 Process for evaluating capability	D14.1 & Evaluation spreadsheet

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**
Definition of Designated Community	APARSEN WP25	Deliverable D25.1
	SCIDIP-ES GIS	
Perform preservation actions	Preservica Preservation workflows	Cloud based preservation actions
Evaluate Preservation capability	D14.1 Common Test Environments	Process for evaluating the capability of disparate systems to perform preservation 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 licence)
Transformation	OPF related SCAPE Various e-science projects	Details of software
Handover	SCIDIP-ES Brokerage/Orchestration service	Examples of hand-over
Audit	APARSEN WP33	
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 ontologies evolve (this is important for e-Science)	APARSEN WP14	Experience in both theory and its applicability (including tools) A paper that describes the approach: Tzitzikas, M. Kampouraki, A.

		Anastasia, 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.

Evidence	
Source of evidence	**Summary of evidence**
APARSEN WP14 / D14.1 Process for evaluating capability	D14.1 & Evaluation spreadsheet
In D25.1 some possible solutions to fill these gaps 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	Experiments reported in the aforementioned paper.

4.5 AUTHENTICITY

This term is meant to include all the Preservation Description Information defined by OAIS, plus aspects of data quality and non-technical provenance such as reputation. It implies the early automated collection of types of information recognizable as PDI and their standardized organization over time according to well defined policies.

Asset base		
Issue	**WP/Project/Tools/Services**	**Asset**
Authenticity	APARSEN WP24	Deliverable D24.1
	SCIDIP-ES authenticity toolkit	Software
Policies for authenticity evidence management	APARSEN WP35: recommendations	D35 (under development)
Audit	APARSEN WP33	D33b
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.	APARSEN WP22 DIGOIDUNA ODIN	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
Modelling and exchanging provenance information	APARSEN WP24	Core ontology for provenance, mappings with other schemas that can capture provenance. Inference rules for propagating provenance

Gaps
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
Additional kinds of annotation are available, ranging from publications in refereed journals to grey

literature. An automated way of evaluating evidence about authenticity is not available.
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).

Evidence	
Source of evidence	**Summary of evidence**
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.	One of the main results of the WP22 has been the definition of an IF for PIs. The trust criteria of eligibility to the framework have been recognized and agreed by many experts (in particular those who have evaluated the framework in two distinct phases, HLEG) as the crucial aspect to adopt the framework as the core interoperability layer on which to build authenticity and provenance services.
	APARSEN Internal Deliverable: ID2401 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

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 but 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[1], 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.

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**
Creation of RepInfo to match knowledge base of a community	APARSEN WP25	Deliverable D25.1
	SCIDIP-ES RepInfo Toolkit, Preservation Strategy Toolkit, Registry, Gap Identification service	CASPAR evidence SCIDIP-ES software and User feedback
Virtualisation techniques	SCIDIP-ES Virtualisation toolkit	Experience of application of virtualisation techniques
	SHAMAN	Multivalent software
Discovery techniques	APARSEN WP22	Ontology and tools for PI
	SCIDIP-ES Finding Aid Preservica	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 ODIN	PI Interoperability Framework PI resolvers

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

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 modeled 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 a structured way enabling complex searches	PANGEA	

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).

Evidence	
Source of evidence	**Summary of evidence**
Prototype of IF services for accessing integrated information across systems	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.
ODIN Proof of concepts in two domains: High-Energy Physics and Humanities and Social Science	Preliminary models on data exchange and workflows
Interoperation initiatives presented at the two workshops on interoperability between PIs	ORCID and ISNI joint statement on interoperation VIAF and ISNI interoperability initiative
Research Publications. System Epimenides that proves the feasibility of the approach.	

4.7 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 Preservation 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**
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
Ingest workflow to check if the digital objects are protected	APARSEN WP31, tools for generating technical metadata and file analysing like FITS, JHOVE, ExifTool, WP21 (Characterization service)	Overview of appropriate best practices from the workflow perspective, i. e. the 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 TIMBUS	Overview of metadata standards that allow the including of digital rights information TIMBUS DRM Manager and License Ontology
Use of persistent identifiers in rights management	LCC project	

Gaps

Evidence	
Source of evidence	**Summary of evidence**

OAIS: http://public.ccsds.org/publications/archive/650x0m2.pdf, page 3-1, 3-2.	
Mets Rights: http://www.loc.gov/standards/rights/	Rights Declaration Schema
PREMIS http://www.loc.gov/standards/premis/	PREMIS Rights entity

4.8 VALUE

Value is usually created by the usefulness of the 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 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 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 objects, perhaps arising from the value or potential value of the object.

Asset base		
Issue	**WP/Project/Tools/Services**	**Asset**
How to assign value to digital objects	BRTF report [6]	Revised 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
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.
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	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

services deal with many aspects of the e-science landscape including data and information access, knowledge discovery, Citability, quality assessment and provenance.		

Gaps
The major gaps are related to the identification, classification and quantification of the benefits and impacts the usage of preserved 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) Comment by RR: Already described in the beginning of 4.5.2. in current version of the master doc 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, economical, 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.

Evidence	
Source of evidence	**Summary of evidence**
Deliverables D36.1 & D36.2 DPImpact chapter 4 & 5	Description of currently in place policies and practices in pioneering DP practices
WP32 deliverable D32.2 – see benefits section	See paper
ODIN Proof of concepts in two domains: High-Energy Physics and Humanities and Social Science	Preliminary models and results on data exchange improvements and workflows Results will be published soon at http://odin-project.eu/project-outputs/
Tutorials and resources explaining Risk Calculation	http://preservica.com/resource/using-the-dvar-calculator/

4.9 GOVERNANCE

Asset base		
Issue	**WP/Project/Tools/Services**	**Asset**
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

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

Evidence	
Source of evidence	**Summary of evidence**
Deliverable D36.2	Strong presence of DP implementation based on projects, not as a general policy

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**
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
Escrow services	TIMBUS	
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
So far there exists no commercial escrow service for custom software that protects investments in tailored software from bankruptcy by depositing source code at third parties.

Evidence	
Source of evidence	**Summary of evidence**
Deliverable D36.2	Lack of business cases based on quantitative evidence of usage of preserved assets, actual quantifiable benefits and thus value measurements

4.11 COSTS

Asset base		
Issue	**WP/Project/Tools/Services**	**Asset**
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	See the work in the 4C project (Neil Grindley, JISC) And in RDA (Data Publishing - cost models)	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 projects a coordination action on a ‘Collaboration to clarify the costs of curation’. The Curation Costs Exchange tool may be of relevance – due Jan 2015. Reliable cost models for support of persistent identifiers.

Evidence	
Source of evidence	**Summary of evidence**
Deliverables D32.1 and D32.2	

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**
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
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

Evidence	
Source of evidence	**Summary of evidence**
D36.1 & D36.2	DP being primarily funded by short term budgets

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**
	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

Evidence	
Source of evidence	**Summary of evidence**

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**
How to obtain some level of audit and certification for trustworthiness in terms of preservation.	APARSEN WP33 test audits	D33.1B
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
Security issues in audit	TIMBUS	Security ontology

Gaps
Additionally to audit policies, technical tools that allow to store audit data in a secure way are still missing.

Evidence	
Source of evidence	**Summary of evidence**

4.15 GLOSSARY

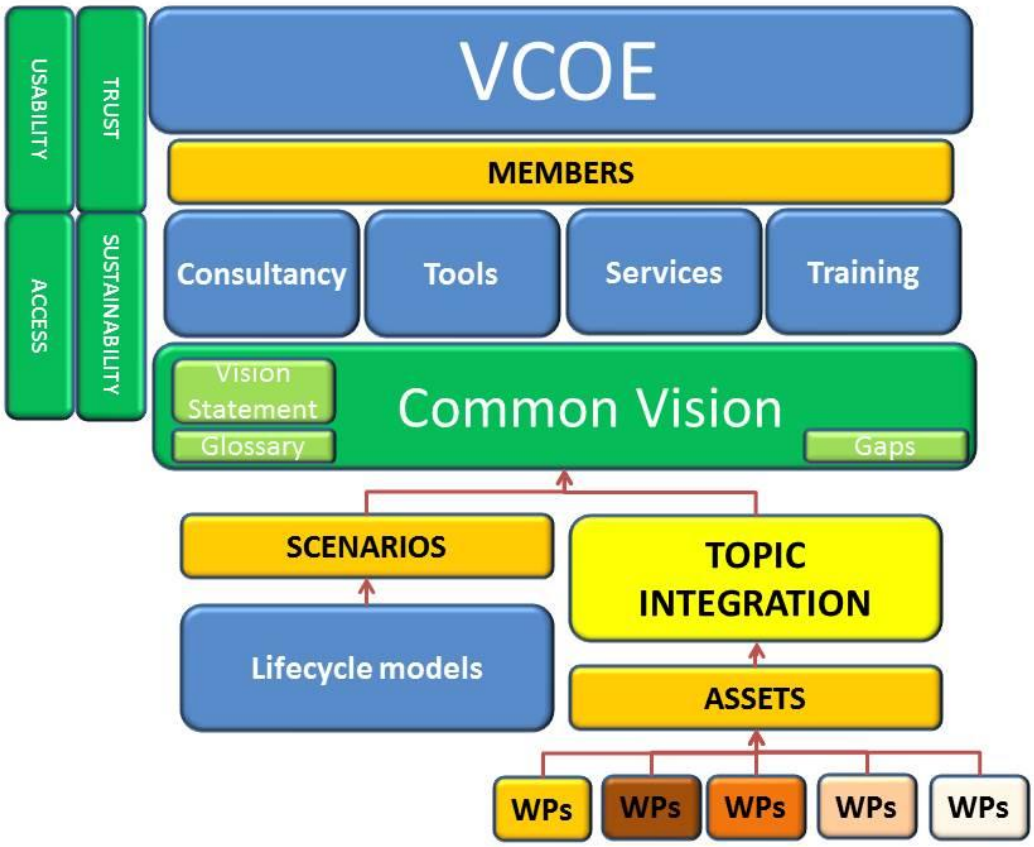
A common vocabulary is very important to avoid confusion, especially when communicating with and between a number of different disciplines/cultures.

There are a (growing) number of candidate vocabularies.

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, a SKOS approach has been used to indicate relationships (broader, narrower, related etc) between the terms [4].

5 ROLE 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 Knowledgebase) underpins all of these, as described next.

5.1 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.

5.2 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 Representation Information Toolkit.

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.

5.3 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)

5.4 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.

6 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

7 ANNEX 1: SCENARIOS

7.1 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	WP21 and SCIDIP-ES
-- 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

7.2 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

7.3 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 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 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 an 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[2] and a selection will be appended here.

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