Interoperability Objectives and Approaches: Results from the APARSEN NoE

Barbara Bazzanella
Department of Information Engineering and Computer Science (DISI), University of Trento, Italy
barbara.bazzanella@unitn.it

Yannis Tzitzikas
Institute of Computer Science, FORTH-ICS, and Computer Science Department, University of Crete, Greece
tzitzik@ics.forth.gr

ABSTRACT

In this paper we report the main results of a study on interoperability objectives and approaches in digital preservation (DP), conducted within the APARSEN Network of Excellence (NoE)[1]. The aim of the investigation was to collect interoperability challenges and goals from various initiatives and project partners and to produce a matrix of solutions and guidelines that can guide the stakeholders in DP to the multi-dimensional and complex landscape of digital preservation interoperability. The paper describes the main findings of the research, including 1) an overview of the current projects and initiatives on interoperability in different areas of digital preservation, 2) an analysis of the main interoperability scenarios and challenges encountered by partners and other stakeholders in their daily life activity that served to drive the definition of the main common interoperability objectives for digital preservation, 3) a broad matrix of models, standards and services for interoperability that cover the main areas of digital preservation, which can be used as a working instrument to navigate the complex ecosystem of the current interoperability solutions, and 4) a list of recommendations and guidelines to create the ground for a coordinated and interoperable digital preservation ecosystem.

1. INTRODUCTION

Interoperability refers to the ability of two or more independent systems to exchange information and use the exchanged information in meaningful ways and without special effort to achieve common goals [4, 1]. Interoperability has become a critical imperative for digital preservation in recent years and several initiatives have started to focus on the definition of requirements, technological solutions and best practices in order to define digital preservation interoperability frameworks, services and standards for effectively and reliably access the preserved digital content between interoperating systems. This shows the general agreement within the DP community that an effective DP strategy or solution strictly relies on a broad international consensus on interoperability, as well as on appropriately designed technological infrastructures to enable it. Identifying the interoperability issues involved and the interoperability objectives to achieve is a first step to promote such a consensus. However this is not a trivial task due to a number of aspects to consider. On the one hand, digital preservation has started to be approached as a problem of "interoperability with the future" [10] or "temporal interoperability" [5], that is ensuring that current systems interoperate with future systems to guarantee that digital resources remain accessible and reusable over a long period of time maintaining their meaning and value. According to this definition, the techniques used for contemporaneous interoperability are applicable for temporal interoperability (i.e. digital preservation), indicating many potential commonalities and points of synergy between interoperability in real time and digital preservation even though temporal interoperability requires a specific focus on sustainability and applicability of the same strategies in the long term. On the other hand, the resources that need to be preserved are highly heterogeneous and increasingly distributed across different systems and organizations which should interoperate in real time, share responsibilities and rely on each other to provide integrated and cross-boundary DP services. The temporal dimension of interoperability is just one aspect of the complexity of the interoperability landscape in DP. First of all, interoperability is a very broad and complex concept, which is conceived on different levels of abstraction (as discussed in the next section) ranging from syntactic to semantic interoperability [7] passing through technical, political, organizational and legal perspectives [8] and dealing with many interoperability objects (e.g. metadata, persistent identifiers, policies). Secondly, several interoperability issues cut across different areas of digital preservation (e.g. Persistent Identifiers, Authenticity and Provenance, Preservation services) showing a very fragmented landscape where there is relatively little harmonization of models, standards and services used in the creation, management and preservation of digital cultural contents. Finally, different stakeholder communities deal with a broad range of interoperability challenges and barriers, which affect in many ways different local functionalities and approaches.

Diagnosing this complex ecosystem is a first fundamental step in order to reach a common awareness about the main interoperability challenges in DP and to define a core set of interoperability objectives for the future. The NoE of the APARSEN project should play a key role to coordinate the definition of this agenda due to its commitment in the creation of a common view and understanding about the preservation and interoperability requirements in different preservation domains, communities and research areas. This paper aims at providing a contribution in this context, summarizing the main results of an investigation on interoperability objectives and approaches conducted within the APARSEN project. First of all, it gives a broad overview

[1] http://www.alliancepermanentaccess.org/index.php/aparsen/

of ongoing and past projects and initiatives covering interoperability issues related to digital preservation. Secondly, the paper discusses interoperability scenarios and challenges encountered by partners and other stakeholders. Third, a broad matrix of interoperability models, standards and services is described as a working tool to navigate the complex interoperability ecosystem. The paper closes with an initial set of recommendations which should promote the realization of an interoperable long-term preservation ecosystem. More details and results can be found in the public deliverable, D25.1[2].

2. WHY INTEROPERABILITY IS IMPORTANT FOR DIGITAL PRESERVATION

A study conducted by the EC in 2011 mentions "interoperability" as one of the ten most important research topics for digital preservation research[3]. In this section we discuss why addressing digital preservation issues with a focus on interoperability may offer significant advantages over current practices for ensuring access, exchange and reuse of digital content in the long term.

First of all, digital preservation certainly requires preserving the bits of the digital objects, but this is probably the less difficult task. The preservation of their accessibility, intelligibility, provenance, authenticity, quality (and many others, e.g. citability, searchability, etc) is a more complex task. All these requirements can be considered as interoperability aspects, in the sense that they can be considered as abilities to apply (now and in the future) successfully in different objects the same operations for accessing them, understanding them, rendering them, getting their provenance information, etc. This is why digital preservation has been termed "interoperability with the future". Moreover, interoperability usually refers to the ability to "exchange and use information between independent systems in meaningful ways and without special effort". As a consequence, achieving interoperability (according to this definition), implies ability to exchange and use information without special effort, thus preservation of accessibility, intelligibility, etc, without special effort.

Secondly, expressing crucial digital preservation challenges as interoperability challenges has a beneficial impact not only for the design and implementation of scalable technical solutions, but also for the definition of a common research agenda agreed by stakeholders, which are concerned with long-term preservation and stakeholders that are focused on building interoperable digital environments. By recognizing that common needs and issues are in play, it should be easier to adopt integrated solutions and expand the applicability of standards and models developed within a certain context to data created and used by other communities and across technical, organizational, political and social boundaries.

Third, DP can be conceived as an interoperability exercise along the entire spectrum of steps that form the lifecycle of a digital object, from its creation to its re-use through the process of preservation. A fundamental aspect of this exercise is the adherence to digital preservation standards, as pointed out by (National Information Standards Organization, 2004) "An institution must ensure that its standards are in line with those used across the digital library community to enable interoperability where possible". To this purpose, digital preservation standards should not be conceived from a repository-centric point of view but should be defined as a set of functional requirements which can be implemented by multiple systems with different hardware and software platforms, data structures, and interfaces to manage and exchange data in the medium and long term with minimal loss of content and functionality.

Finally, in the global context of digital information, DP has more and more to deal with data (e.g. cultural heritage data, scientific data) that are syntactically and semantically heterogeneous, multilingual, multicultural, semantically rich, distributed and highly interlinked. Making this content mutually interoperable so that it can be searched, accessed and reused in the long time is a big challenge for DP involving different levels of interoperability. On a syntactic level, it is needed to harmonize different character sets, data formats, identification syntaxes, notations and collection records adopted in different collections but also to agree on communication protocols for information exchange between content providers. At the level of semantic interoperability, different metadata standards are in use by different institutions to describe the same type of content, metadata formats may be interpreted differently, data is encoded at different levels of precision, vocabularies and ontologies used in describing the content are different and ontology alignment and mapping is hard to completely automate. The multi-organizational and multidisciplinary nature in which content is collected, maintained and published poses new issue of organizational interoperability for DP dealing non only with formats and technical standards but also with different policies, rights and restrictions management, mandates, roles and responsibilities. Interoperability appears a complex and multi-layered concept and a crosscutting concern [9], which encompasses a multidimensional spectrum of aspects ranging from more technological aspects to include several dimensions of the digital preservation universe (e.g. users, policies, legal issues, disciplines). Moreover, different communities and disciplines may have very heterogeneous interoperability requirements since their needs with regard to data management and curation vary considerably. It follows that devising an appropriate solution to the digital preservation interoperability challenges is far from being a merely technical problem and the diversity of the community requirements makes it impossible to aim for a single strategy or system for economical, political, organizational and disciplinary reasons. Interoperability is crucial to address issues like access, provenance, citability, data quality assessment and many others, going far beyond the technical level to embrace a much wider horizon where organizational, social and business strategies must be taken into account in considering effective solutions. If an all-encompassing perspective is taken, including technical, social, organizational and many other factors, a comprehensive picture of this complex landscape can be provided, enhancing the understanding of its faces and orienting strategies for finding specific solutions.

[2]available at `http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2013/03/APARSEN-REP-D25_1-01-1_7.pdf`

[3]`http://cordis.europa.eu/fp7/ict/telearn-digicult/report-research-digital-preservation_en.pdf`

3. INTEROPERABILITY INITIATIVES

As a first step of our diagnosis of the ecosystem of interoperability initiatives and solutions in DP, we performed an analysis of ongoing and past projects and initiatives covering interoperability issues related to (or relevant for) digital preservation. The aim of the investigation was to produce a database of projects and initiatives to be made publicly accessible within the APARSEN NoE and maintained updated in the long term as a collaborative tool to raise awareness and understanding within the DP community. We collected information about 64 projects and initiatives, clustered around eight macro-areas:

1. **Digital Preservation Conceptual Models and Interoperability Frameworks**: in this category we included the main digital preservation projects, which addressed interoperability issues by defining shared conceptual models or developing interoperability framework architectures. This group contains 1) early research projects in the field of DP (e.g. DELOS) focused on the definition of basic concepts and shared conceptual models as fundamental ways to enable interoperability of the various content holders (mainly digital libraries and archives) and rising awareness about the theoretic basis for the key preservation concepts and entities, 2) later-stage projects which addressed interoperability by developing solutions to integrate digital preservation modules into framework architectures to enable the interoperation with other systems. Examples of this kind of architectures are the PLANETS Interoperability Framework for preservation actions, the CASPAR Integrated Framework based on the OAIS reference model, and the integrated preservation framework using grid-technologies of SHAMAN.

2. **Data Infrastructures for E-Science**: E-science infrastructures represent a key strategic area for digital preservation and a rich source of interoperability challenges. First of all, they are of crucial importance to significantly enhance science in many areas, promoting research, innovation and enabling new ways of collaboration and resource sharing. However, the realization of the innovation potential of these infrastructures, strongly depends on the creation of an interoperable data sharing, re-use and preservation layer. Secondly, these infrastructures may represent robust components to support digital preservation services for science data in general (see the PARSE.Insight project) or in specific domains (see for example the SCIDIP-ES project in the earth science domain). This macro-area clusters existing initiatives that aim to promote interoperability in specific e-science domains through the implementation of e-science infrastructures (e.g. INSPIRE, SCIDIP-ES, CLARIN, DASISH) and describe also some relevant initiatives committed to promote and develop reference models and architectures to enable infrastructure interoperability across systems (e.g. iCORDI, EUDAT GEANT, D4Science-II).

3. **Digital Libraries**: In this category, we included some of the most relevant initiatives to address the interoperability challenges in the domain of digital library. Some of these initiatives focused on the development of a common conceptual framework for enabling interoperability between digital libraries (e.g. DL.ORG) or for exchanging specific types of content (e.g. IIIF), others addressed the issue of creating a unique point of entry to distributed content and heterogeneous resources (e.g. EUROPEANA, EUROPEANA GROUP).

4. **Open Repositories**: Open repositories represent another important domain for developing interoperability solutions related to DP purposes. In the recent years, Open Access repositories and their associated services have become an increasingly important component of e-Science Infrastructures. It has been widely recognized that the real potential of open access repositories for e-Science infrastructures lies on the creation of a network of interconnected repositories providing unified access to distributed scientific resources and scholarly content. The creation of this decentralized infrastructures and the development of added-value services on top of it are entirely reliant on interoperability. In this category we included projects and initiatives addressing three main issues: 1) Metadata harvesting and exchange (CRIS/OAR Interoperability Project) ; 2) Infrastructures for digital repositories (DRIVER and DRIVER II); 3) Repository deposit and access (OpenAIRE, Open Access Repository Junction, Open Archives Initiative).

5. **Persistent Identifiers**: Interoperability between persistent identifiers (PIDs) is one of the key challenge for guaranteeing persistent discoverability, accessibility and reuse of digital resources and therefore is of central importance for enabling effective digital preservation solutions [2]. This category includes a remarkable number of initiatives that in the last years focused on persistent identifiers interoperability, for digital objects (PersID, RIDIR, PILIN), for authors (ORCID), for scientific data and related resources (DIGOIDUNA, EPIC) and for entities in general (OKKAM).

6. **Semantic Interoperability and Linked Data**: this category groups some relevant initiatives, which have adopted the Linked data framework to face problems of interoperability related to digital preservation issues in the library context, such as data interoperability, unified data access and interconnecting data silos. It includes library initiatives aiming at exposing their records as Linked Data (LOCAH, CEDAR, LUCERO), promoting the use of Linked data as a Web standard within the library community (W3C Library Linked Data Incubator Group, BIBFRAME) and using semantic web technologies for enabling semantic interoperability of metadata vocabularies (STITCH).

7. **Semantic Access to Earth Sciences resources**: Exploiting the experience of one of the partners of the project (ESA), we included in the analysis also projects and initiatives in the specific domain of Earth Science since its relevance for DP research (see for example SCIDIP-ES project). The analysis focused mainly on the problem of interoperability issues concerning semantic access to Earth Science resources based on ontologies (OTE, OTEG), semantic discovery tools and frameworks (SMAAD), data and metadata sharing (like GEOSS).

8. **Other**: the last category was introduced to include those projects and initiatives which could not fit into one of the previous categories or domains (i.e. EpSOS in the domain of e-Helth, ISA in e-Government)

Each initiative has been described according to the following categories: 1) Name: the name of the initiative or project, 2) Domain: indicates a specific area to which the project or initiative belongs; 3) Timescale: indicates the duration of the project or initiative; 4) Description: provides information about the project or initiative, its objectives and the issues addressed by it; 5) Interoperability objectives: provides a list of the specific interoperability goals addressed by the project or initiative. 6) Link: is the URL of a Website where more information and documents can be found. An overview of the analyzed initiatives is shown in Figure 1. We refer to the project deliverable for more details about each initiative.

4. CHALLENGES

To frame the discussion around interoperability and start to identify interoperability objectives, gaps and recommendations, we collected from partners and other stakeholders a set of interoperability scenarios and challenges. Each scenario has been evaluated (using Likert-type scales) according to three dimensions, i.e. 1) the current situation about the raised issue, 2) the importance/impact of the issue, 3) the level of difficulty to address the problem. Some scenarios have been directly extracted from other deliverables of the APARSEN project (and we will refer to them for more details) and other sources (e.g. the DIGOIDUNA study[4]). The 13 collected scenarios have been organized into the following clusters pertaining different areas of the digital preservation landscape or specific domains (e.g. Earth Science): 1) Persistent Identifiers (PIDs) Interoperability, knowledge discovery and citability; 2) Semantic metadata Interoperability and lifecycle management; 3) Semantic Interoperability in the EO Domain; 4) Provenance and Authenticity Interoperability. For space reason we can not include a full description of all the collected scenarios which are reported in the project deliverable mentioned above. An example scenario about provenance interoperability is illustrated in the following box to give an idea of the adopted approach.

SCENARIO: Exchange and Aggregation of Provenance Information

A sensor e.g. at a satellite, makes some measurements. The measurements are then transferred to a ground station. The data are then processed by a group of researchers, say group A, to produce an image, say img1. The image is then processed by group B to produce a second image, say img2. To produce the complete provenance of the img2 (which may be important for assessing the credibility/authenticity of img2) we have to aggregate the provenance information of each data object and link them appropriately. This aggregation requires having a common model for representing provenance or mappings between the adopted models.

Challenge: Ability to exchange and aggregate provenance information of various processing tasks or transfer/archiving events.

Evaluation: Current situation (bad); importance (high); level of difficulty (fair).

Relevance for DP: Provenance information is of crucial importance for e-Science (e.g. for checking and validating results, for reproducing them, etc). However, even though several solutions for modeling and recording provenance information have been proposed and various mapping between these models have been defined (see for example [11]), their adoption by the various organizations is still scarce. In short, interoperable solutions for enabling exchange and aggregation of provenance information, like methods that can aid the ingestion and management of provenance information, are available but there is a lack of awareness and understanding by the e-Science stakeholder communities of the importance of adopting these solutions. The interoperability issue is more at the organizational and inter-community level than at technical level.

In this section we describe the main interoperability challenges derived from the analysis of the scenarios for each domain of investigation. In the first cluster, called Persistent Identifiers (PIDs) Interoperability, Knowledge Discovery and Citability, the scenarios covered the following aspects: 1) knowledge discovery and data integration through PIDs; 2) author identifiers interoperability; 3) impact and quality assessment; 4) citability of scientific datasets. From these scenarios we derived the set of challenges reported in the following box.

Challenges from scenarios about Persistent Identifiers (PIDs) Interoperability, knowledge discovery and citability:

1. To provide a global resolution mechanism, which ensures that given an identifier of any kind the correspondent resource can be persistently retrieved and accessed. If the resource is not available any more, a matching resource if available (also from a different provider) should be linked.
2. To provide a unique interface to find integrated information across different systems about an identified entity (e.g. a paper) and related entities (related publications, authors, datasets).
3. To create a collection from resources, that belong together (e.g. enhanced publications).
4. To associate multiple identifiers with the same entity (e.g. author) to enable the long term access to the entity or a description of it.
5. To locate all versions of a resource.
6. To find information about authenticity and availability of a resource.
7. To integrate metadata referring to the same resource from multiple sources.
8. To make citation and their relationships more explicit

[4] http://digoiduna.wordpress.com/about/

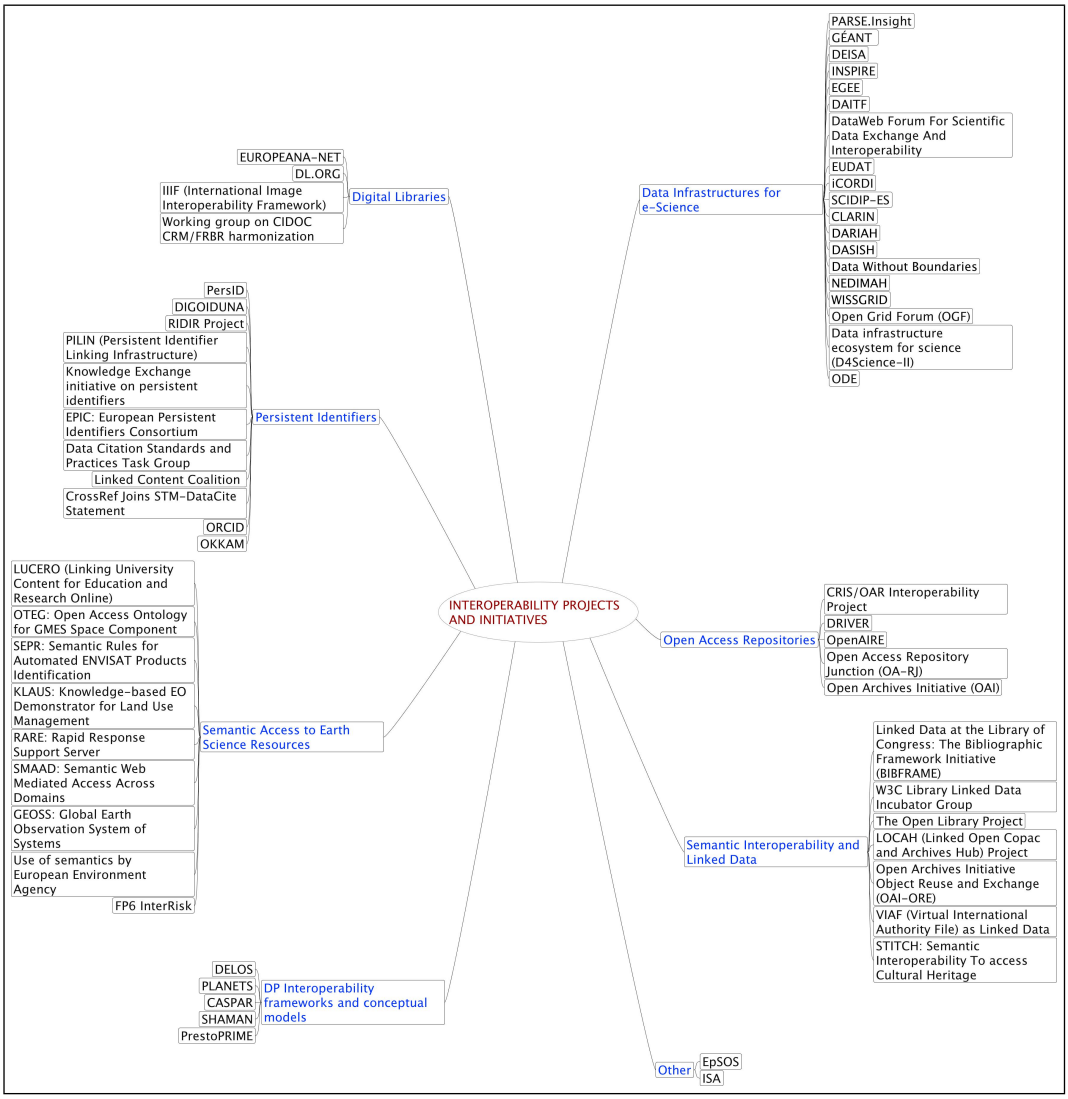

Figure 1: Interoperability projects and initiatives relevant for DP

so that data can be accessed more easily, supporting reuse and verification and strengthen the link between the contributor and data.
9. To define a standard to uniquely identify datasets and manage them as separately citable items.

The second set of challenges pertains the domain named semantic metadata interoperability and lifecycle management. In this domain we included scenarios describing narratives about vocabulary alignment, creation of semantic links between archival collections and other (Web) sources, use of integrated metadata search interfaces across several providers for accessing digitized objects. The following challenges have been derived from the analysis of this second set of scenarios.

Challenges from scenarios about Semantic metadata Interoperability and lifecycle management:

1. To provide mapping between vocabularies, thesauri and categorization systems to facilitate browsing and searching in several library catalogues in parallel with the keywords from any of the used thesauri.
2. Aggregating diverse data sources and performing vocabulary alignment to a common ontology in order to facilitate searching and finding structured results also across multiple languages.
3. To provide metadata mapping between domain-specific metadata models used by different sources.
4. To interlink metadata relevant for digital preservation actions (e.g. metadata about digital objects, their formats, versions, events and agents involved in the events).
5. To aggregate metadata from different data providers

and provide a common way to search for their content using these metadata.
6. To create semantic links between heterogeneous materials from different sources including web resources.
7. To provide identification mechanisms for accessing provenance (metadata) of digital objects and intellectual entities.
8. To develop a common standard for exchange information between institutions adopting different archival systems.
9. To define a framework to relate library publications to datasets that are held by other institutions.

As mentioned above, we investigated the specific domain of Earth Science as an important testbed for DP practices and solutions. From the analysis of the proposed scenarios in this domain, the following challenges emerged:

Challenges from scenarios about Semantic Access to Earth Science resources:

Allowing application domain experts to access the needed EO resources through an interoperable and pluggable architecture, permitting:
1. Data discovery via controlled vocabulary, which would permit the user to search resources through familiar terminology;
2. Direct access to the needed resource, independently where the resources are physically hosted (e.g.: federation of smaller and remote catalogues).

Finally, the analysis of the scenarios provided some interoperability challenges concerning Provenance and Authenticity. In particular the collected scenarios addressed 3 main issues: 1) exchanging and aggregating provenance information of various processing tasks or transfer/archiving events; 2) Querying provenance records of any digital object through services that can fetch and integrate the required provenance information from heterogeneous and distributed sources; 3) finding information about resource authenticity and availability. The analysis of the provenance scenarios produced the following set of challenges.

Challenges from scenarios on Provenance, Authenticity and Rights:

1. To develop a common model for representing provenance information or a mapping solution between different models to aggregate provenance information from different sources.
2. To provide query and retrieving systems and user interfaces to give access to heterogeneous and distributed provenance information.
3. To develop a trusted PIDs infrastructure which guarantees access to authentic digital objects and related provenance information.
4. To define a standard way to expose rights expressions with metadata.

5. SOLUTIONS

The challenges described in Section 4 provide a partial view on the complex ecosystem of interoperability problems in DP. Since the goal of our investigation was to identify the interoperability issues encountered by the APARSEN partners as part of their daily activities and gather the conceptual models, services and standards used by them to address these issues, we deepened the analysis by identifying concrete interoperability barriers, needs and related solutions (i.e. models, standards, frameworks, services) adopted by the partners in relation to the key digital preservation areas investigated within the APARSEN project. The final aim was to describe which are the critical interoperability aspects pertaining a certain area of digital preservation, which main layers of interoperability are mainly involved, which are the interoperability objects that are implicated and finally which concrete solutions (e.g. models, standards) have been adopted to address these issues. The result of the analysis led to define a sort of matrix, which combines different layers of interoperability (e.g. syntactic, semantic, organizational) with the areas of digital preservation (e.g. persistent identifiers, metadata, provenance) and the related interoperability objects and models, providing an interoperability conceptual framework for digital preservation that can be used as a starting point to facilitate practical interoperability solutions and design concrete interoperability services for long-term preservation. To this purpose, we organized the collected information on the basis of a common framework that aims to characterize the problem facets as well as the existing and forthcoming solutions and models. In this way the specific challenges of interoperability within a specific area could be directly linked with the current available solutions, providing a useful working instrument to address concrete issues of interoperability encountered by relevant stakeholders in their daily work activities. The proposed framework includes the following categories: 1) **Digital Preservation area**: indicates the area of digital preservation where interoperability takes place. Examples are preservation services, persistent identifiers, authenticity and provenance. 2) **Interoperability issue/challenge**: a problem of interoperability which hinders a certain task or process in an interoperability context. 3) **Interoperability objects**: are the entities that actually need to be processed in interoperability scenarios. They can include for example the full content of digital resources or mere representations of such resources (i.e. metadata, identifiers). 4) **Adopted solutions/ models/ standards**: are those approaches, which are adopted to address specific interoperability issues/challenges at different levels. An example of a described solution for enabling interoperability for PIDs for authors is shown in Figure 2. The Figure 3 provides a mind map that summarizes the contents of the collected material which has been organized in a matrix containing 58 interoperability solutions. The solutions have been clustered around eight categories identified by colours in the figure: 1) Persistent Identifiers, 2) Provenance, 3) Data Quality 4) Metadata 5) Metadata Harvesting and Information Exchange, 6) Authentication, Authorization, Rights, 7) Preservation Models and Services, 8) Research data deposit, discovery, access, reuse and citation.

6. RECOMMENDATIONS

In order to put theory into practice we have devised four sets of recommendations, which should promote the realization

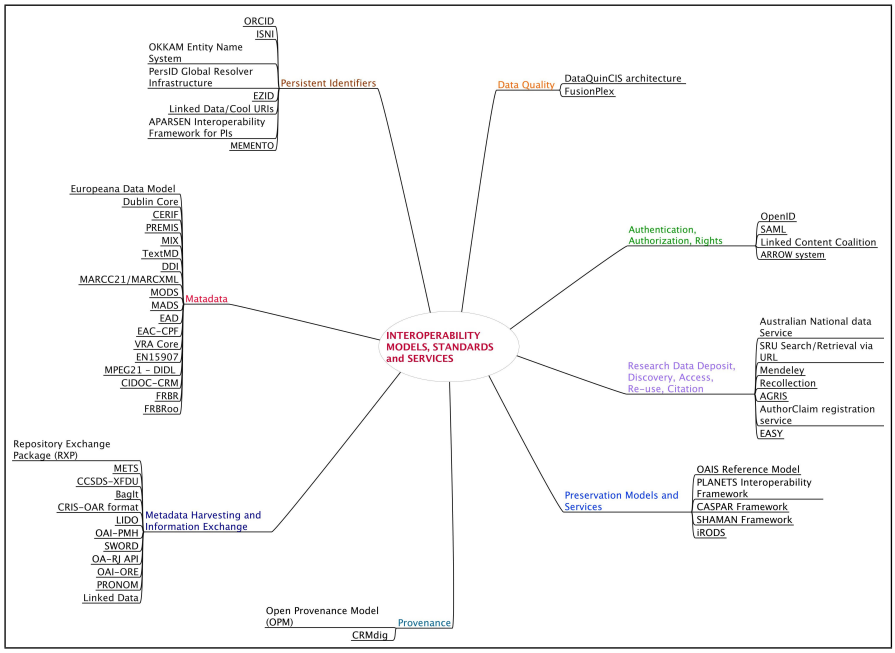

Figure 3: Interoperability solutions for DP

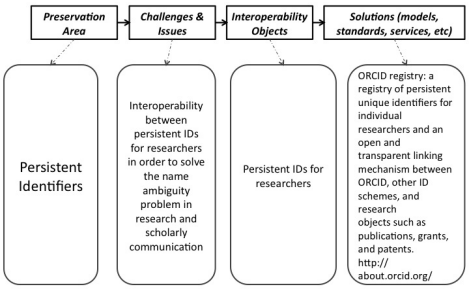

Figure 2: An example extracted from the matrix of solutions.

of interoperability in DP. These are recommendations that are applicable to all the categories of stakeholders and aim at: 1) Fostering the broad adoption of common standards and specifications reducing dependencies, facilitating the interoperation between systems for the entire digital object lifecycle management process and enabling higher-level services on top of standard compliant systems. 2) Promoting the use of appropriate identification systems and their interoperability. 3) Promote the convergence towards agreed common policies and governance models, which foster the adoption of interoperability solutions and trust on them. 4) Ensuring the necessary long-term financial support and the efficient use of economic resources.

6.1 Standards

Standards are considered essential elements of interoperability. The first set of recommendations concerns the definition and adoption of standards as starting point for DP interoperability.

> **1. Standards are Good: rely on standards in case there are appropriate standards for the digital objects at hand.**

The first recommendation about standards states that if compliance to one standard guarantees the achievement of one or more interoperability objectives, then the adoption of the standard is certainly beneficial and recommended. From a dependency point of view, we can say that the standardization essentially makes the dependencies more clear and resolvable.

> **2. Standards are not a Panacea: be aware that standardization does not vanish the dependencies of the digital objects.**

The second recommendation mitigates the first one, specifying that not all the interoperability objectives which can occur in the DP landscape can be addressed through the use of common standards. The obsolescence of a standard, for example, may represent a potential threat for interoperability, e.g. if a standard Y becomes obsolete and there are no longer tools that support it, then a digital object represented through Y could be not reusable any more. The open issue is therefore whether we could tackle the interoperability problem without having to necessarily rely on several and possibly discrepant standards, and whether we can exploit solutions to reduce dependencies and tackle the problems of vanishing or evolving standards. One approach to address

these issues will be briefly discussed in the conclusions to this paper.

3. Define Interoperability Standards through the entire lifecycle of a digital object.

According to the third recommendation, standards should regulate the entire chain of digital preservation steps that form the lifecycle of a digital object from its creation to its re-use through the process of digital preservation.

4. Content providers should adopt standards to ensure that their digital content interoperates with other services and collections allowing the development of a common access point to distributed resources.

The forth recommendation remarks the importance of the use of standards in global information spaces. In these contexts where a huge amount of resources from heterogeneous sources is integrated and made accessible, it is important the adoption of common standards that enable interoperability. Some level of interoperability, for example, is assured at data ingestion by requesting data providers to expose their metadata according to a common standard model for metadata[5].

5. Involve stakeholders in the definitions of standards.

The last recommendation states that since it is difficult to mandate standards, it is easier to work on community accepted standards. Community evolution of standards should be encouraged. A concrete example of a successful coordinated effort between two communities to define a common interoperability standard is the joint effort of the CIDOC Conceptual Reference Model and Functional Requirements for Bibliographic Records international working groups to establish a formal ontology, called FRBRoo[6], intended to solve the problem of semantic interoperability between bibliographic and museum resources, facilitating information integration and exchange.

6.2 Identification

The second set of recommendations deals with two aspects of identification of digital resources: 1) the use of Persistent Identifiers (PIDs) to identify digital resources and other related entities (e.g. authors) and 2) the identity of content.

[5]This approach is used for example by the digital library Europeana (`http://www.europeana.eu/portal/`) which has introduced a cross-domain semantic framework to accommodate the range of metadata standards adopted by the different cultural heritage sectors from which the data are collected.

[6]`http://www.cidoc-crm.org/frbr_inro.html`

Bootstrap an interoperability solution for Persistent Identifiers.

The persistent identification of digital objects (e.g. articles, datasets, images, stream of data) and non-digital objects (namely real-world entities, like authors, institutions but also teams, geographic locations and so on) is becoming a crucial issue for the whole information society and for the development of e-Science infrastructure in particular. However the proliferation of several PIDs systems within different communities and the resulting fragmentation of the PIDs ecosystem has led to an urgent demand for establishing an interoperability solution among the current PID systems to enable the persistent access, reuse and exchange of information across different systems, locations and services. Therefore, actions are needed to bootstrap the convergence towards an interoperability solution for PIDs which open new prospects for advanced value added information integration services. However, since any identifier system is always used within cultural, organizational, geographical and disciplinary boundaries through a technical system, it follows that designing an appropriate solution to the problem of identifiers interoperability involves a number of non-technical issues. This means that any action to bootstrap an interoperability solution needs to work towards systematic implementation of those organizational, political, social and economical factors that foster trust and agreement among the relevant stakeholders.

Elaborate on Information Identity.

Apart from the problem of identifiers, another critical point is the identity of the content. Even though library and archival practice, as well as Digital Preservation, have a long tradition in identifying information objects, the question of their precise identity under change of carrier or migration is still a riddle to science. One theory, developed in the context of APARSEN, that tries to give some light to this aspect is described at [3]. The objective is to provide criteria for the unique identification of some important kinds of information objects, independent from the kind of carrier or specific encoding. The approach is based on the idea that the substance of some kinds of information objects can completely be described in terms of discrete arrangements of finite numbers of known kinds of symbols, such as those implied by style guides for scientific journal submissions.

6.3 Organization, governance and trust

The third set of recommendations concerns the organizational dimension of the interoperability exercise. Since DP is currently conceived as a responsibility to share between different organizations, it has become clear that in such cooperative context, interoperability issues at technical level cannot be solved without promoting an agreement and improving communication at an organizational level.

Raise agreement, increase awareness and social support towards a common interoperability agenda.

Given the complexity of the interoperability exercise in many areas of digital preservation and the variety of stakeholders involved, a common direction must be defined. The involved parties should work together to define a common agenda ensuring a coordinated and interoperable digital preservation ecosystem. The forthcoming VCoE (Virtual Centre of Excellence) of the APARSEN project should play a key role to coordinate the definition of this agenda due to its role in the creation of a common view and understanding about the preservation and interoperability requirements in different preservation domains and research areas. The agenda will define a clear conceptual framework, which will be a pre-requisite for dialogue and achieving consensus across the communities impacted, and serving as the basis for promoting awareness and mobilisation of skills and resources. The common agenda should include at least the following points: 1) Raising awareness about digital preservation interoperability objectives, challenges and available solutions. 2) Promote a cross-boundary view on challenging issues and opportunities. 3) Planning interventions to promote awareness, dissemination and education programs in order to reinforce knowledge and skills on interoperability strategies and solutions.

Foster good interoperability practices.

Spreading good practices for interoperability digital preservation needs to include a more deliberate exchange of lessons learned and case studies documenting the use of emerging solutions, workflows, and techniques across national, organizational and disciplinary boundaries. The analysis and evaluation of scenarios, as well as the identification of prioritized interoperability challenges described in the present document, can be used to benchmark available approaches and systems and identify best practices according to certain identified interoperability objectives. Moreover the use of specific variables of performance (e.g. sustainability of the solution, scalability) can be adopted to develop plans on how to make improvements and adapt specific best practices to specific contexts.

Promote and encourage coordination and collaboration among stakeholder communities around policies and governance addressing interoperability objectives.

The different needs and goals of the stakeholders involved in different areas of digital preservation may hinder the adoption of available interoperability solutions. Therefore, actions are needed to favor the convergence towards common policies and governance, which can help to achieve consensus across the communities. The APARSEN NoE is actively working to promote such a consensus (in particular within the WP35) by defining a methodology for implementing governance structures and data policy management mechanisms to enhance interoperability for permanent access to the records of science.

Work towards global trustable solutions.

Trust is a fundamental issue for DP [7], but it also critical for interoperability solutions working effectively. Actions are needed to promote international agreement on global standards and policies. In this way, users can have evidence of authenticity for world-wide data (e.g. scientific) and resources. The creation of an European Framework for Audit and Certification of Digital Repositories is an example of the actions promoted within APARSEN to build global trust by enabling interoperability between increasingly challenging audit processes in digital preservation.

6.4 Economic

DP poses not only technical, social and organizational interoperability issues but raises also interoperability issues which deals with the economic imperatives of DP which is required to guarantee sustainable results against limited resources. In this section we discuss recommendations which consider the economic aspects of interoperability strategies for DP.

Devise sustainable interoperability solutions.

Securing long-term sustainability of an interoperability solution or service is a key factor for promoting its trust, adoption and success. This can be ensured only if the organization behind it is sustainable and can guarantee the longevity of the solution. This is not simply a matter of finding sufficient funds but concerns many different aspects.

Build a robust community behind the interoperability solution or service.

The first step to establish a sustainable interoperability solution is to gain the support of (possibly) all the involved actors. Interoperability solutions are only possible if cultural heritage institutions, governments, public administrations, research institutions and private organizations work in close cooperation in supporting them, sharing responsibilities and finding adequate business strategies. This strategy has been pursued, for example, by the ORCID initiative (see Section 2.4.4) which worked to gain the support of a broad community including many different stakeholders (like individual researchers, universities, national libraries, commercial research organizations, research funders, publishers, national science agencies, data repositories and international

[7] see `http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/09/APARSEN-Trust-Brochure-Low-Res-Web-Version.pdf` for a discussion about this topic.

professional societies) before working on devising a technical solution to the problem of interoperability between author identifiers.

Align the interests, roles and responsibilities of the involved stakeholder communities into a sustainable economic strategy and operationalise them in a business model.

The stakeholder participation is also crucial in the definition of sustainable business strategies. To this purpose a business working group including the representatives of all the communities can be created to review membership policies, budget models and investigating funding options to ensure the long term sustainability of the solution.

Provide clear incentives to adopt the interoperability solution.

The lack of clear incentives to adopt a given interoperability solution may threat its use and long-term sustainability. For example, the adoption of shared methods and services by independent organizations may bring costs. Sometimes the costs are financial due to the purchase of hardware or software or for hiring and training staff. In other cases costs are organizational. Introducing a new standard requires inter-related changes to existing systems, altered workflow, changed relationships with suppliers and so on. It is important to make clear the added value of adopting the solution and its beneficial impact in the long-term.

7. CONCLUSIONS AND FUTURE WORK

In this paper we have discussed interoperability challenges and approaches in DP and we have proposed an initial set of recommendations to foster the creation of an interoperable DP ecosystem. The results of this investigation have shown the importance of a coordination among the actors of this ecosystem which goes beyond the technical aspects of implementing a valuable solution, to embrace a much wider horizon including organizational, social, political and economical aspects and implications of adopting it. Raising awareness and increasing a common understanding about the current initiatives and available solutions is a first important step towards this coordinated effort. Therefore, a first future work activity within the APARSEN NoE will be dedicated to make the collected information publicly available, hopefully implementing searching and filtering tool to facilitate the query formulation and navigation of the information space. A second activity will be dedicated to the topic of managing interoperability dependencies. We could say that each interoperability objective/challenge, like those described in the current paper (and deliverable D25.1), is a kind of demand for the performability of a particular task (or tasks). The next step (which will be done in the context of APARSEN) is to identify such tasks, and reflect on their dependencies and on how these can be modelled. The ultimate objective is to propose a modelling approach that enables the desired reasoning, e.g. task performability checking, which in turn could greatly 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 could also assist preservation planning, especially if converters and emulators can be modeled and exploited by the dependency services. The plan is to follow the general approach described at [6], in particular the approach that supports also modeling converters and emulators described at [12].

Acknowledgements

Work done in the context of NoE *APARSEN* (Alliance Permanent Access to the Records of Science in Europe, FP7, Proj. No 269977). Many thanks to all who have contributed to this report, and in particular to David Giaretta (APA Director, STFC), Simon Lambert (STFC) Veronica Guidetti (ESA), Andrea Della Vecchia (ESA), Emanuele Bellini (FRD), Barbara Sierman (KB), Juha Lehtonen (CSC), René van Horik (DANS).

8. REFERENCES

[1] Tylor A.G. *The Organization of Information*. Westport, CN: Libraries Unlimited, 2004.

[2] P. Bouquet, Bazzanella, R. B. Riestra, and M. Dow. Digoiduna: Digitla object identifiers and unique author identifiers to enable services for data quality assesment, provenance and access. `http://digoiduna.files.wordpress.com/2011/12/digoiduna_final_report_expert_feedback1.pdf`, 2011.

[3] Martin Doerr and Yannis Tzitzikas. Information carriers and identification of information objects: An ontological approach. *arXiv preprint arXiv:1201.0385*, 2012.

[4] Anne Geraci. *IEEE Standard Computer Dictionary: Compilation of IEEE Standard Computer Glossaries*. IEEE Press, Piscataway, NJ, USA, 1991.

[5] Margaret Hedstrom. Exploring the concept of temporal interoperability as a framework for digital preservation.

[6] Y. Marketakis and Y. Tzitzikas. Dependency Management for Digital Preservation using Semantic Web technologies. *International Journal on Digital Libraries*, 10(4), 2009.

[7] Jinsoo Park and Sudha Ram. Information systems interoperability: What lies beneath? *ACM Trans. Inf. Syst.*, 22(4):595–632, October 2004.

[8] Miller Paul. Interoperability: What is it and why should i want it? *ARIADNE Web Magazine for Information Professionals*, 2000.

[9] Jeff Rothenberg. Interoperability as a semantic cross-cutting concern. *Interoperabiliteit: Eerlijk zullen we alles delen*, 2008.

[10] C. Rusbridge, P. Burnhill, S. Ross, P. Buneman, D. Giaretta, L. Lyon, and M. Atkinson. The digital curation centre: a vision for digital curation. In *Proceedings of the 2005 IEEE International Symposium on Mass Storage Systems and Technology*, LGDI '05, pages 31–41, Washington, DC, USA, 2005. IEEE Computer Society.

[11] Christos Strubulis, Yannis Tzitzikas, Martin Doerr, and Giorgos Flouris. Evolution of workflow provenance information in the presence of custom inference rules. In *3rd Intern. Workshop on the Role of Semantic Web in Provenance Management (SWPM 2012), co-located with ESWC*, 2012.

[12] Y. Tzitzikas, Y. Marketakis, and Y. Kargakis. Conversion and Emulation-aware Dependency Reasoning for Curation Services . In *Proceedings of the 9th Annual International Conference on Digital Preservation (iPres2012)*, 2012.