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

D22.3 Demonstrator Set Up and Definition of Added Value Services: Part 2

Document identifier:	**APARSEN**-REP-D22_3-01-1_0
Due Date:	2013-12-31
Submission Date:	2014-01-17
Work package:	WP22
Partners:	FRD, UNITN,....
WP Lead Partner:	FRD
Document status	Released
URN	urn:nbn:de:101-20140516125

Abstract:

This deliverable presents the refinement of the WP22 Interoperability Framework for Persistent Identifiers on the basis of a two-stage evaluation provided by a group of experts. This report aims at describing the framework elements and functionalities. In particular, the ontology, the implementation strategy and two basic services are presented. In order to show the feasibility of the proposed approach a demonstrator has been implemented.

Delivery Type

Author(s) Emanuele Bellini (FRD), Maurizio Lunghi (FRD), Barbara Bazzanella (UNITN), Paolo Bouquet (UNITN), Salvatore Mele (CERN), Hervé L'Hours (UESSEX), , Rene van Horik (DANS), Sabine Schrimpf (DNB), Kirnn Kaur (BL), , Sunje Dallmeier-Tiessen (CERN), Patricia Sigrid Herterich (CERN), Samuele Carli (CERN), Artemis Lavasa (CERN), Laure Haak (ORCID), Juha Hakala (Finnish National Library)

Approval

Summary

Keyword List

Availability ☒ Public

Document Status Sheet

Issue	Date	Comment	Author
0.9	2013-12-19	Completion ready for internal review	Maurizio Lunghi, Barbara Bazzanella
1.0	2014-01-10	Taking into account internal and external review comments Final editorial checks	Maurizio Lunghi Simon Lambert, David Giaretta

Project information

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

Project Co-ordinator: 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

1.1 SUMMARY

The goal

This document presents the main results of the ongoing work within WP22 of the APARSEN project (www.aparsen.eu) for the definition of an Interoperability Framework (IF) for Persistent Identifier (PI) systems and related services. The main goal of WP22 is to propose a model for interoperability at service level among existing PI systems without interfering with their internal organisation and policies. Currently PI systems offer only a service to resolve the PI to the URL: for our point of view this is not enough, and is made worse by the possibility that PIs might vanish in the future.

The proposal

The IF is built on four main pillars: 1) PI systems or domains definition; 2) four main assumptions; 3) eight trust criteria; 4) an ontology model (see chapter 2). The main elements in the IF are the PI domains that are included in our definition. In the framework of this work we use the term “PI system” or “PI domain” as synonymous (see Glossary) to indicate the global combination of the user community which is interested in the PI services and in some cases provides the content to be identified, the bodies offering the PI services, the technology used, the roles & responsibilities architecture, and the policies for different parts of the appropriate long-term preservation plan. The IF model is suitable for any type of PI provided that it fulfils the four main assumptions and the eight trust criteria defined in the model. Currently we consider four different types of PI systems: PI for digital objects, PI for physical objects, PI for bodies and PI for actors.

The assumption and trust criteria have been largely endorsed by the experts and constitute a relevant result of the work carried out. In particular, the trust criteria are based on the basic criteria for a correct digital preservation policy and practice (see chapter 2.2 and 2.3).

The ontology schema describing the main entities, properties and relations, is based on FRBRoo as it is one of the most widely used schemas in the cultural heritage sector. Mapping with other schemas have been done and other can be extended in future (see chapter 2.4).

The community

To have a large expert validation of the IF model and to prepare the ground for future consensus building, the IF has been evaluated and reviewed by a High Level Expert Group (HLEG), a group of 44 experts in the domain of Persistent Identifiers (see chapter 1.6). The IF model was submitted to the expert group for comments, first in June 2012, and again in June 2013. The IF model has been improved thanks to the HLEG work, including establishing a common terminology and rising a wide consensus on the criteria required for a trusted PI system. Some relevant projects have been involved to establish cooperation around the PI issues and the IF model. Representatives of these projects attended the 2 workshops that we organised in Florence and Lisbon (see chapter 5.2, chapter 5.3). From the work of the HLEG and the workshops in Florence and Lisbon, many valuable contributions have been collected and integrated into the current version of the framework.

The demonstrator

A demonstrator with PI systems for digital objects and actors was developed in December 2012 and refined in December 2013 as described in this document. We intend to use this demonstrator for two main objectives: i) to test the feasibility of the IF model implementation; ii) to measure the user satisfaction about some services across different PI domains and get refinement.

For reasons of practicality in the current demonstrator we have implemented only PI for actors and for digital objects. The architecture is distributed over 3 SPARQL end-points collecting data from seven content providers and two services installed. In order to avoid going in too deeply with metadata describing the content identified by the PI, the demonstrator focuses only on PIs and related information. The work to set up these structures and to expose contents is described (see chapter 3.6). On top of this model we have developed some basic services.

1.2 OBJECTIVES

One of the main goals of the APARSEN project is to build a long-lived network of excellence, i.e. a Virtual Centre of Excellence (VCoE) for digital preservation, which aims to defragment the current complex ecosystem of digital preservation in Europe. The project activities within the WP22 address the current fragmentation between PI systems for digital resources and other entities, like physical objects, people, and institutions, aiming to propose an interoperability platform (IF) acceptable for all the PI systems. Since it is unlikely that a unique global identification solution will emerge in the future, the challenge is to establish an IF among the current PI solutions to enable persistent access, reuse and exchange of information through the use of existing identifiers and associated resources across different systems, locations and services.

Our concept of 'interoperability' is quite simple and is not used to indicate the ability of PI systems to interoperate between them in a direct way but it is conceived in terms of a common access method to data belonging to heterogeneous PI domains with different identification schemes. Our goal is to make accessible metadata from all the PI domains in the same format so that users can use them without worrying about different internal organisation and policies.

The IF model implementation and proof of effectiveness through a practical demonstrator will be useful and instrumental to help to pave the route for the APARSEN VCoE, even if it is still under preparation. On one side the IF will create a common open area for interoperability of PI systems with huge benefits for users in terms of usability and accessibility of data, and secondly it will offer a platform where any service tailored to user requirements can be implemented on all the data across PI domains which are currently isolated. After the end of the project, if the IF is widely implemented it can become a reference model for any future development for PI systems and it could create a 'Ring of trusted PI for Linked Open Data (LOD)' as explained in chapter 3 about future strategies. To ensure sustainability of the work done, the Fondazione Rinascimento Digitale and the University of Trento will maintain the framework implementation and some basic services after the end of the project at least to the end of 2015.

1.3 WORK PACKAGE & DELIVERABLES

The WP22 consists of three tasks. The relations among these tasks are described by the Figure 1 extracted from D22.1.

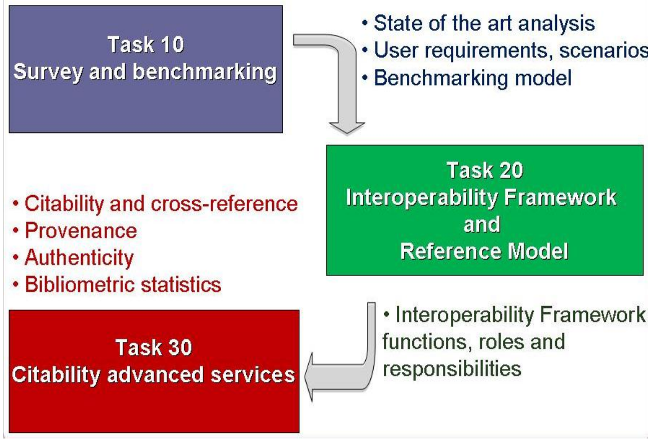

Task 10
Survey and benchmarking
• State of the art analysis
• User requirements, scenarios
• Benchmarking model
Task 20
Interoperability Framework
and
Reference Model
• Citability and cross-reference
• Provenance
• Authenticity
• Bibliometric statistics
• Interoperability Framework functions, roles and responsibilities
Task 30
Citability advanced services

Figure 1: WP22 tasks and their relationships

The **Task 2210** started with a wide state of art and user requirement analysis, through desk-research, questionnaire and analysis of past and current projects, definition of scenarios and use cases to identify current practices of PI systems and possible benefits or user requirements for future services across heterogeneous PI domains. Results from this task provided fundamental information about user requirements and needs to be addressed within the framework, as well as a number of critical gaps of the current PI landscape.

The **Task 2220** is focused on modelling an Interoperability Framework for Persistent Identifiers systems, which addresses functions, roles and responsibilities to allow interoperability among these systems. This task includes the definition of the first version of the Interoperability Framework and its revision by the HLEG.

The **Task 2230** aims at designing some advanced services for resources identified by different PI systems, such as services for citability, cross-referencing, quality assessment, citation metrics and evaluating the user satisfaction about these services. The definition of services is mainly based on the results of the Task 2210 (in particular those from the survey, scenarios and use cases) and the feedback provided by the expert group during the second round of evaluation Task 2220.

The results of the WP22 are described in 4 deliverables listed in Table 1.

Deliverable title	Deadline
D22.1 Persistent Identifiers Interoperability Framework http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D22_1-01-1_9.pdf NBN: urn:nbn:it:frd-6204	M 12
D22.2 Demonstrator set up and definition of added value services: Part 1 http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2013/02/APARSEN-REP-D22_2-01-1_7-M16.pdf	M 16
D22.3 Demonstrator set up and definition of added value services: Part 2 *... the current document ....*	M 36
D22.4 The Interoperability Framework implementation with added value services	M 48

Table 1: WP22 list of Deliverables

1.4 METHODOLOGY

The initial modelling work and design of the Interoperability Framework has been described in D22.1 together with the preliminary steps and results that supported the definition of the model (see Figure 1): state of the art analysis and benchmarking of the main current PIs systems for digital objects, people, and organizations, results from the survey on the use of PIs among a large sample of relevant users, description of scenarios, and use cases. All these steps allowed the definition of the user requirements implemented by the framework.

Following the recommendations of the EC, the initial model has been evaluated and improved by a group of experts in two rounds of refinement exercise (see Figure 2). Some temporal details of this work follows.

The WP approach was to start in the first year of the project from a wide analysis of the current practices of PI systems and user requirements about possible future services. Then, we made a proposal for an Interoperability Framework **(**IF) for PI systems, set up a High Level Expert Group of around 40 experts, both internal and external to the APARSEN consortium (see chapter 1.6) and carried out a revision of the model through a questionnaire (Jun-Sept 2012). A second version of the model has been presented at the International Workshop on Interoperability of Persistent Identifier

Systems in Florence on 13 Dec 2012 (see chapter 5.2). In the occasion of the workshop a demonstrator of the approach has also been presented. The demonstrator is necessary first of all to test the feasibility of the IF model and evaluate the implementation strategies. In a second phase, when the framework is populated by content providers, we will develop some basic services across PI domains to test users' satisfaction and refine the service definition accordingly.

A second round of the IF evaluation (Jun-July 2013) has been carried out again with the help of the HLEG. The revisited model is based on FRBRoo ontology[1], to arrive at a harmonized model to be implemented by PI providers and content providers in the demonstrator. However, the feedback collected during the second workshop on Interoperability of Persistent Identifiers (see chapter 5.3) at the 10th International Conference on Preservation of Digital Objects (iPRES2013, 2-6 September 2013, Lisbon) suggested the need to reduce the complexity of the ontological model focusing on the representation needs of PI providers (i.e. an elementary set of relationships to describe the identified entity).

After revising the IF, the WP22 activities has moved to the design and development of a demonstrator, which aims to show the potential benefits of using the proposed framework to identify digital data and related information. To this purpose, we have defined an action plan to set up a demonstrator for the IF and related services, with the limited resources for software development in WP22 and considering some external possible synergies with other projects like EUDAT, ODIN, SCAPE, SCIDIP-ES or other initiatives like RDA, LCC, ORCID, ISNI, ARK and DOI or NBN communities. In this phase we exploited the expertise on PI gained by the WP22 team during other projects and initiatives, namely 1) the OKKAM project[2]; 2) the DIGOIDUNA study[3], 3) the PersID project[4], 4) the Italian NBN initiative[5]. Based on that demonstrator, we aim to design some basic services to start to address the user requirements collected during the former work in the WP22 with the PI questionnaire and the use cases definition.

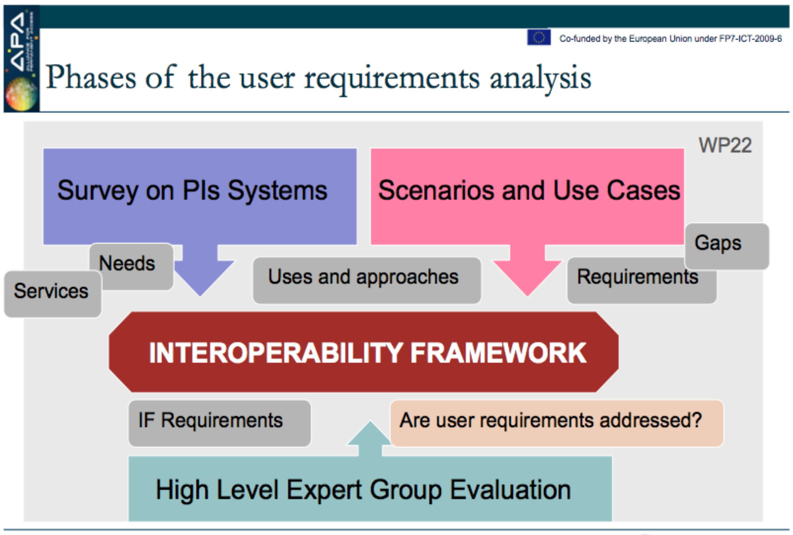

Co-funded by the European Union under FP7-ICT-2009-6
Phases of the user requirements analysis
WP22
Survey on PIs Systems
Scenarios and Use Cases
Gaps
Needs
Services
Uses and approaches
Requirements
INTEROPERABILITY FRAMEWORK
IF Requirements
Are user requirements addressed?
High Level Expert Group Evaluation

Figure 2: User requirements analysis within the WP22

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

[2] The OKKAM project (http://www.okkam.org/) led by the University of Trento was aimed at creating a scalable infrastructure, called the **Entity Name System** (ENS), for the systematic reuse of global and unique entity identifiers.

[3] The DIGOIDUNA study (http://digoiduna.wordpress.com/) led by the University of Trento, aimed at supporting policy makers at European and Member States level to understand the challenges of adopting solutions for managing identifiers in the context of *scientific data e-infrastructures* (SDIs).

[4] PersID initiative (http://www.persid.org/) aimed to provide Persistent Identifiers as well as a transparent policy and technical framework, for using scientific, cultural and other resources in the Internet. FRD was a partner of the project.

[5] Italian NBN initiative (http.://www.depositolegale.it) is linked to the national service for legal deposit in digital format of all the cultural and scientific content.

1.5 ACTION PLAN

The action plan consists of three main phases that cover the full duration of the APARSEN project up to the end of 2014. In some cases a phase has been repeated for refinement purposes.

I.	**Validation** – Validation of the model by a user group of experts on PIs; duration of the validation a couple of months; number of experts around 40; on-line working modalities and tools[6] (July-Aug 2012) & (June-July 2013).
ii.	**Demonstrator** – Definition and set up of a demonstrator with selected resources from the APARSEN community; it's important to have data from different PI domains and for digital objects and people; definition of the publication modality and tools development in support of publication of data; before the end of 2013 population of the demonstrator with data.
iii.	**Services** – Proposal of few services and development on a cross-PI domain basis using all the data available in the demonstrator; before Jan 2014 some basic services will be implemented. Monitoring of the complete functionalities & performance of the demonstrator (Jan-Dec 2014) to test the user satisfaction.

I. Validation

To evaluate the proposed IF model and approach, a High Level Expert Group (HLEG) on PIs has been set up: many of the participants are external to the APARSEN consortium to ensure a neutral judgement. A two-phase evaluation exercise was carried out, the first performed in 2012. This round of feedback greatly improved the model. The second round in June-July 2013 finalised the model for the development of the demonstrator with the basic services defined in the following phases of the project.

II. Demonstrator

To define a test-bed to check the practical implementation of the IF model, we must first of all select a significant amount of data from different PI domains, possibly for digital objects and people. Then, we must define the publishing/exposing modality for the selected data, coherent with the IF model and agreed by the HLEG, to populate the demonstrator and prepare the services development.

The IF implementation stage is divided in three steps or phases:

a) Publication modality definition & software design:

The experts define the workflow, interface, and data format for publishing/exposing the original data owned by PI domains on the IF model: we adopted a Semantic Web approach through the Ontology + RDF triples + SPARQL technology to implement the IF, that can be instrumental for a deep integration of the PIs technology and the Linked Open Data scenario. In the latter case, the ontology approved by the HLEG must drive the definition of the RDF syntax necessary to expose PI domains data on the semantic Web.

This phase is the follow up of the validation work done by the HLEG and it aims to transform the technical specifications of the model concepts and features; in particular the environment and strategy for publication of data and the basic functionalities of the system must be defined.

b) Software Developing

[6] E.g. EasyChair conference management software.

Once the publishing/exposing interface is designed, the software development phase can start. This phase can include a Web service development to implement the interface, the customization of present tools, the RDF style sheet transformation, and so on. Existing tools are selected and adapted to the project needs, and new tools can be developed. Cooperation with other projects is foreseen.

c) Software installation and IF Population:

Once the software tools are available, each involved institution can populate the IF with the PI information and other data according to the ontological schema of the IF. To test the generality of the approach, it would be useful to have not only PI data for digital objects but also for people. The content owners extract data from their database and make them available in line with a metadata schema which can be mapped to the IF ontology. Then data must be implemented in a ‘service provider point’ that can be a Web service (like an OAI-PMH point) or it can be a Semantic Web triples store or a SPARQL end point implementing the IF ontology. While the data are distributed over many archives, the data exchange services are foreseen to be centralised on a point.

III. Services

Developing data exchange services that use data from different PI domains supports cross-domain accessibility and citability of resources. These services can be tailored to specific user requirements. The pervasiveness of data exchange services is in fact one of the key factors for an extensive consensus building and drives the long-term sustainability of these services. Some of the most relevant services will be developed and tested through the demonstrator (see chapter 3):

1. Citability and Metrics Services
2. Global Resolution Services
3. Digital Object Certification

Monitoring of the complete functionalities & performance of the demonstrator is the final part of the project work, which aims also to test the user satisfaction.

1.6 HIGH LEVEL EXPERT GROUP

To validate and refine the IF a High Level Expert Group (HLEG) on PIs has been established, with qualified representatives from relevant initiatives, organizations, and centres. In line with the common vision underlying the VCoE of the APARSEN project, the involvement of experts in addition to the APARSEN community was necessary to prepare the ground for a wider consensus and implementation of the model beyond the project. Obtaining results from different stakeholder communities, we could start to address the defragmentation of activities on PIs in Europe, aggregating projects and exploiting joint results. In this respect, it is worth noting that the HLEG evaluation is the conclusive phase of the user requirement analysis conducted within the WP22 (see Figure 2), which includes three phases: 1) The survey on PI systems, 2) the collection and analysis of interoperability scenarios and use cases which provided an initial input about the needs, requirements, and services to design the first version of the IF and 3) the HLEG evaluation of the IF to refine the IF by specifying more concrete requirements to be addressed.

The HLEG members have also been involved in user requirements collections, as well as in dissemination activities[7]. The experts’ participation is ruled by a specific Cooperation Agreement**,** accepted in advance by the experts. It was the responsibility of HLEG members to read, participate, share, and stay informed of the current discussion with other members. A clear policy of benefits for experts has been defined and declared in the cooperation agreement (e.g., visibility of expert names, reduced fee for some events) and the contributions are made transparent in the produced documents. A common terminology has been agreed by the expert group (see the Annex), that will continue to be an essential tool for the evaluation exercise of the HLEG and is embedded in the APARSEN Glossary

[7] Some of the participants attended the International Workshop on Interoperability of Persistent Identifiers Systems held in Florence on 13 December 2012, and the iPRES workshop on PI held in Lisbon on 5 September 2013.

(http://www.alliancepermanentaccess.org/index.php/knowledge-base/dpglossary/). The members of the HLEG are listed in the following table.

Table 2: Members of the HLEG

1	Anila Angjeli	Bibliothèque nationale de France (BNF) ISNI	EXT
2	Sébastien Peyrard	Bibliothèque nationale de France (BNF) ARK	EXT
3	Ernesto Damiani	University of Milan (UNIMI)	EXT
4	Giovanni Bergamin	Central National Library of Florence (BNCF)	EXT
5	Laurents Sesink	Data Archiving and Networked Services (DANS)	EXT
6	Maarten Hoogerwerf	Data Archiving and Networked Services (DANS)	APARSEN
7	Martin Braaksma	Data Archiving and Networked Services (DANS)	APARSEN
8	Martin Dow	Acuity Unlimited	EXT
9	David Giaretta	Alliance Permanent Access (APA)	APARSEN
10	Alan Danskin	British Library (BL)	APARSEN
11	Marcus Enders	British Library (BL)	EXT
12	Oreste Signore	W3C Italy (W3C)	EXT
13	Gabriella Scipione	Consorzio interuniversitario per la gestione del centro di calcolo elettronico dell'Italia Nord-orientale (CINECA)	EXT
14	Heikki Helin	CSC - IT Center for Science (CSC)	APARSEN
15	Egbert Gramsbergen	TU Delft Library (TU)	EXT
16	Jeroen Rombouts	TU Delft Library (TU)	EXT
17	Hervé L'Hours	UK Data Archive (UESSEX)	APARSEN
18	Carlo Meghini	National Research Council (CNR)	EXT
19	Claudio Cortese	Consorzio Interuniversitario Lombardo per L'Elaborazione Automatica (CILEA)	EXT
20	Yannis Tzitzikas	University of Crete and Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH-ICS)	APARSEN
21	Martin Doerr	Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH – ICS)	EXT
22	Mariella Guercio	University of La Sapienza, Rome (UNIUR)	APARSEN
23	Maurizio Messina	Marciana National Library, NBN	EXT
24	Laurel Haak	ORCID	EXT
25	Norman Paskin	DOI foundation	EXT
26	Samuele Carli	European Organization for Nuclear Research (CERN)	APARSEN
27	Paolo Budroni	University of Wien	EXT
28	Sabine Schrimpf	German National Library (DNB)	APARSEN
29	Alexander Haffner	German National Library (DNB)	EXT
30	Julia Hauser	German National Library (DNB)	EXT
31	Jurgen Kett	German National Library (DNB)	EXT
32	Karaca Kocer	German National Library (DNB)	EXT
33	Lars Svensson	German National Library (DNB)	EXT
34	Juha Hakala	Finland National Library (FNL)	EXT
35	Piero Attanasio	multilingual European Registration Agency of DOI MEDRA	EXT
36	Roberto Delle Donne	University of Naples (UNINA)	EXT
37	Andrea D'Andrea	Università Orientale di Napoli (UNINA)	EXT
38	Aldo Gangemi	National Research Council (CNR)	EXT
39	Mark van de Sanden	SARA, EUDAT project	EXT
40	Antonie Isaac	Europeana technical coordinator	EXT
41	Jan Brase	DataCite	EXT
42	Tobias Weigel	Research Data Alliance (RDA), Deutsches Klimarechenzentrum (DKRZ)	EXT
43	Marcin Werla	Poznan Supercomputing and Networking Center (PSNC)	EXT
44	John Kunze	California Digital Library (CDL)	EXT

2 INTEROPERABILITY FRAMEWORK FOR PI SYSTEMS

2.1 INTRODUCTION

The Interoperability Framework for PI systems is based on four main pillars:

1. PI systems or domains definition
2. four main assumptions
3. eight trust criteria
4. the ontology model

The suggestions and feedback provided by the HLEG allowed us to revise and improve the IF for PIs. In this section we present the results of this refinement.

The main elements in the IF are the PI domains that are included in our definition. In the framework of this work we use the term "PI system" or "PI domain" as synonymous (see Glossary) to indicate the global combination of the user community interested in the PI services and which in some cases provides the content to be identified, the bodies offering the PI services, the technology used, the roles & responsibilities architecture, and the policies for different parts of the appropriate long-term preservation plan.

The IF model is suitable for any type of PI provided that it fulfils the 4 main assumptions and the 8 trust criteria defined in the model. As shown in Figure 3, currently we consider 4 different types of PI systems:

1. PI for digital objects → PI-do
2. PI for physical objects → PI-po
3. PI for bodies → PI-bd
4. PI for actors → PI-ac

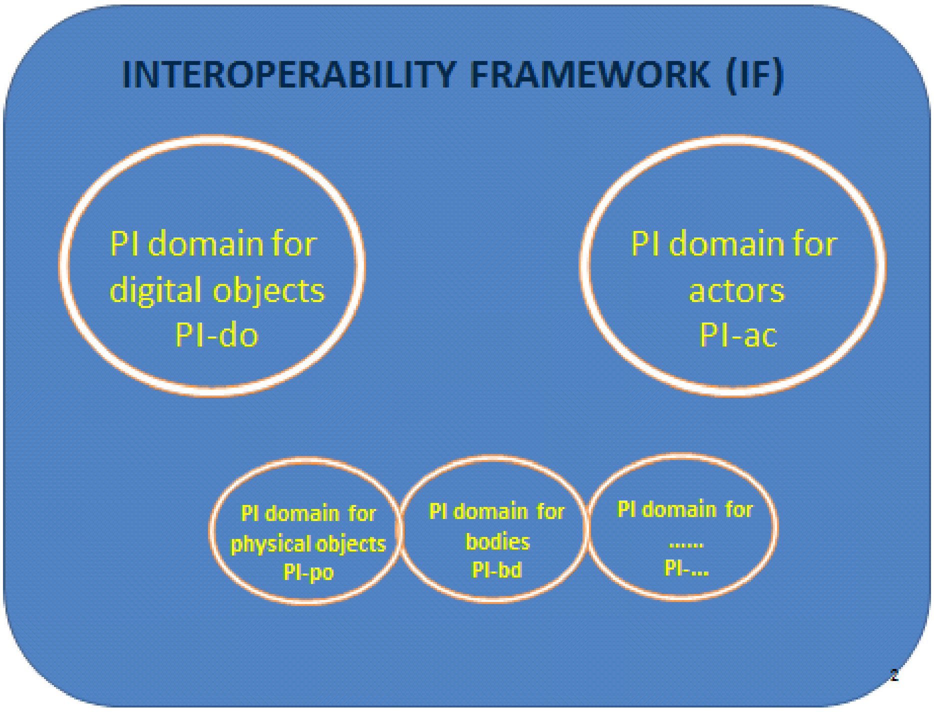

INTEROPERABILITY FRAMEWORK (IF)
PI domain for digital objects PI-do
PI domain for actors PI-ac
PI domain for physical objects PI-po
PI domain for bodies PI-bd
PI domain for ...... PI-...
2

Figure 3: PI domains

Other key actors in the IF scenario are the following.

- The PI domain managers or service providers generating PI on request of users and managing the updating of registries and basic services, like resolution of the PI.
- The Registration Authorities (RAs), that oversee and manage the identifier system;
- The Registration Agencies (RAGs), which manage the creation and registration of PIs according to the trust definition[8] and provide the necessary infrastructure to allow the registrants to declare and maintain the PI-entity relations. ;

- The Content Providers that make accessible the resources identified by the PIs. They can in some cases be the same that the Content Holders who own the content;
- User communities of the PI services, e.g. researchers, funding agencies, publishers.

The aim of the model is to capture significant entities and their relationships in the universe of PI systems with the goal of developing a concrete implementation of the model, ultimately to allow linkages between these entities and to support the implementation of interoperability services. The basic idea is that a common conceptual representation is the main prerequisite to design added-value interoperability services, which can exploit the value of a scheme of representation shared and agreed across trusted systems. The IF answers the general question "How to make PI systems interoperable to facilitate the exchange, re-use and integration of the resources identified in these systems by different PIs"?

The model proposes a high-level solution for representing digital resources and facilitating access and re-use of these representations beyond the borders of hosting systems, enabling a new generation of cross-systems interoperability services. To this purpose, the model standardizes the relationships between the identified entities (e.g., digital objects, authors, institutions) and their PIs, creating a common layer where meaningful information from independent systems can be exchanged.

[8] Emanuele Bellini, Chiara Cirinnà, Maurizio Lunghi, Barbara Bazzanella, Paolo Bouquet, David Giaretta and René van Horik (2012), "Interoperability Framework for Persistent Identifiers system" iPRES 2012, Toronto

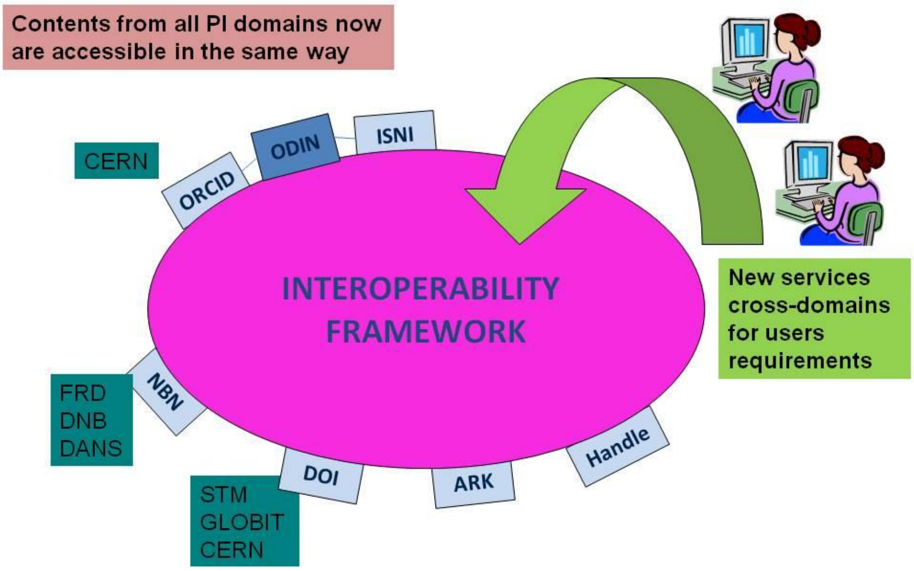

Contents from all PI domains now are accessible in the same way
CERN
ORCID
ODIN
ISNI
INTEROPERABILITY FRAMEWORK
New services cross-domains for users requirements
FRD
DNB
DANS
NBN
STM
GLOBIT
CERN
DOI
ARK
Handle

Figure 4: A graphical representation of the IF

The basic method for defining a PI IF reference model is setting the foundations and identifying the basic concepts within the universe of PIs systems, while creating the ground for the development of appropriate interoperability solutions and interactions with them. It is worth mentioning that the reference model should define a common semantics, not tied to any specific technological implementations, standards or systems. Since it is abstract in nature, the reference model can be used by system designers, as a template for designing different technical interoperability solutions and services based on it. These services are considered external to the framework in our model.

As we said, the IF is based on four main assumptions and eight trust criteria for PI systems to be eligible for the framework. An ontology based on FRBRoo schema describes the PI systems scenario with PI providers and PI domain managers, content providers and final users of PI information, the properties of each entity and their relations with other entities, as well as functions of the framework.

2.2 IF ASSUMPTIONS

The IF rests on four main assumptions described below.

1. **In the IF we consider only entities identified by at least one PI.** Only resources with a PI assigned by a trusted PI domain are eligible to enter the IF framework. Any resource without a PI assigned by a trusted PI domain cannot enter in our scenario as ‘entity’ but, in case, only as descriptive information with no property or relation to other entities.
2. **Only PI domains that meet some criteria are eligible to be considered in the IF: trusted PI domains.** We define some trust criteria as pre-requisites for PI systems to be eligible to the framework. These criteria represent some basic elements of a good policy for a PI domain management and digital preservation plan (see also assumption 4). Only PI domains that address the trustworthiness criteria can join the IF and populate it with their resources.
3. **We delegate the responsibility to define relations among resources and actors to the trusted PI domains.** The IF deals with at least 2 types of PI domains and provides a shared ontology to represent the identified resources properties and their relations. Then,

we forecast some possible relations between two or more objects, or between objects and persons and their PIs. These relations must be provided by the PI domains when they bring an entity into the IF, assuming this declared information as guaranteed by trusted PI domains. The IF does not validate what a PI refers to; any statement coming from a trusted PI domain is assumed to be correct.

4. **We don't address digital preservation policy but delegate that to the trusted PI domains.** PI domains managers are responsible for digital preservation. It is important to notice, that the user community board managing the PI domain is responsible for guaranteeing suitable policies for any aspect of the resource management, such as the content selection/granularity criteria, the Trusted Digital Repositories policies and certification, the PI assignment strategies, etc. The IF does not argue about the type of digital preservation policies for the resources adopted within a PI domain. Within each PI domain there can be different approaches and architectures to share roles and responsibilities among different components of the system, like the RA, the domain resolver, the digital repository, the digital preservation manager, and so on. The user community is free to choose the best solution for particular use cases.

2.3 TRUST CRITERIA FOR PI DOMAINS

To design a reliable IF among PI domains, we have to define the criteria that a PI domain should satisfy to be eligible for the framework. Thus, only those PI domains that match a definition of trust will be taken into account as a component of the framework. The criteria are distinguished between mandatory (M) and optional (O). The following criteria are adopted to decide if a PI domain is trusted and eligible for the IF. The definition of these criteria has been suggested by several studies such as, PIs for Cultural Heritage DPE briefing paper[9], NESTOR reports on trustworthiness of PI systems[10], A Policy Checklist for Enabling Persistence of Identifiers[11], and the ERPANET [12] and DCC [13] workshops.

1. Having at least one RAG.

PI registration should be regulated by well-defined registration policies committed by trustworthy registration authorities. Within a PI domain it is necessary that at least one RAG is established to assign and maintain the association between PI and digital resource. This criterion is considered mandatory in our trust assessment (M).

2. Having at least one Resolver accessible on the Internet.

The resolution of an identifier is the key mechanism enabling a system to locate the identified resource or information related to it within a digital network. Since the Web is the dominant network to access digital resources, in order to meet this criterion a resolver able to resolve a PI has to be accessible on the Web; in some cases the RAG acts also as Resolver for the PI domain. This criterion includes also the capability of a PI to be resolved to a single object such as Web page or file or to both object and metadata or to multiple objects, such as different formats of the same objects, or different content types, through the same PI (multiple resolution). We consider

[9] http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf

[10] http://files.d-nb.de/nestor/materialien/nestor_mat_13_en.pdf

[11] http://www.dlib.org/dlib/january09/nicholas/01nicholas.html

[12] ERPANET workshop Persistent Identifiers Thursday 17th - Friday 18th June 2004-University College Cork, Cork, Ireland www.erpanet.org/events/2004/cork/index.php

[13] DCC Workshop on Persistent Identifiers 30 June – 1 July 2005 Wolfson Medical Building, University of Glasgow http://www.dcc.ac.uk/events/pi-2005/

this criterion mandatory (M). Note that the resolvability of identifiers on the Web is one of the principles of identification stated in the Linked Content Coalition Framework[14].

3. **Uniqueness of the assigned PIs within the PI domain and globally**.

Every identified resource should be uniquely identified within a trusted namespace. The RA has to guarantee that a PI is univocally assigned to a resource within the PI domain through a clear assignment of roles. In fact, since a PI is essentially a string, the uniqueness can be assured only within a domain of reference served by a defined RA. This criterion is considered mandatory (M). As a matter of fact, the uniqueness within the PI domain brings also the global uniqueness for the PI. This principle also is delineated by the indecs project[15] for interoperability of data in e-commerce systems.

4. **Guaranteeing the persistence of the assigned PIs**.

Persistence of an identifier indicates that the identifier should support its intended function in the long term. The ID technology is important to guarantee identifier persistence (for example by excluding changeable or meaningful information from the ID string) but organizational commitments are more crucial for this purpose. This criterion is considered mandatory (M).

Each RA has to guarantee the persistence of the generated PI in terms of preventing the following possible actions:

a) *String modification*: indicates the PI string update, this kind of updating procedure is not allowed according to our definition of a trusted system.

b) *Change resource*: the original PI is reused and assigned to a new resource, this is not acceptable.

c) *Deletion*: indicates the possibility of deleting a PI once it has been created and assigned, this is another process that must be avoided to guarantee trust.

d) Lack of *sustainability*: indicates that a RA is not able to maintain a PI well beyond the lifecycle of the associated resource, the PI must survive also in case the original resource is not available anymore, as minimum some information must be provide as resolution of the PI. Managing identifiers in a sustainable way is another requisite for a trusted PI domain.

5. **User communities, which implement the PI domain, should implement policies for digital preservation for their resources (e.g. trusted digital repositories).**

It is well known that the main objective of a PI is to provide a reliable access to digital resources in the long term. Thus, if on the one side the RA has to guarantee the persistence of the PIs and their association with the identified digital resources, on the other side, PIs should be used to identify stable and preserved digital resources. The content providers should manage their contents with repositories compliant with standards and common criteria of trustworthiness[16] and implement digital preservation strategies for the resources identified by a PI. This criterion is considered mandatory (M) in principle even if we don't enter in details about the practical implementation, since content providers manage resources with different life cycles and they can also adopt different commitment to preserve their contents in respect to other institutions.

6. **Reliable resolution.**

[14] http://www.linkedcontentcoalition.org/#!lccframe/c4nz

[15] http://cordis.europa.eu/econtent/mmrcs/indecs.htm

[16] Examples of Trusted digital repository criteria are: *Date Seal of Approval*: http://www.datasealofapproval.org/, *Nestor Catalogue of Criteria for Trusted Digital Repositories*: http://files.d-nb.de/nestor/materialien/nestor_mat_08-eng.pdf, *Trusted Digital Repositories: Attributes and Responsibilities*, http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf - *Trustworthy Repositories Audit & Certification: Criteria and Checklist* (TRAC): http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/trac.pdf-ISO/DIS 16363: http://public.ccsds.org/publications/archive/652x0m1.pdf, ISO/DIS 16919 http://wiki.digitalrepositoryauditandcertification.org/pub/Main/WebHome/RequirementsForBodiesProvidingAuditAndCertification-SecRev1.doc

One of the crucial functionalities of a PI system is ensuring that the resolution results of a PI are always the same across time. The definition of the meaning of *the same* statement is critical, since different domains may manage digital resources at a different level of granularity and require that a PI is generated and assigned to different levels of abstraction of a digital resource. For instance, the PDF version of an article and the HTML version of the same article can be considered an "equivalent manifestations" of the same object within the DOI domain, while they would receive two different identifiers in the NBN domain. Again, if a digital resource is subjected to digital preservation strategies, such as migration, the results can be considered equivalent manifestations in a domain but not in another. In fact, in the CrossRef DOI service there is only a guideline, namely "Assign new CrossRef DOIs to content in a way that will ensure that a reader following the citation will see something as close to what the original author cited as is possible."[17] According to this, the reliability of resolution is referred to guarantee, provided by a PI domain, that the resolution of a PI points to *the same* resource along the time, according to the similarity definition adopted by a PI community. This criterion is considered mandatory (M).

7. Uncoupling the PIs from the resolver.

This criterion is crucial and it is referred to the PI generation rule defined by a PI system. To be eligible for the IF a PI system has to be based on identifiers whose syntax does not include the URL of the resolver or the content provider in the string. For instance, the NBN syntax definition does not include the URL of the associated NBN resolver. This feature is necessary because the URL of the resolver itself can change. Thus, if a part of the PI string specifies the URL of the resolver domain, all the PIs which contain the original URL will become invalid, in case the resolution service is moved to another domain. Once the PI and the resolver are decoupled, multiple resolution become possible. Different URLs may be associated to the same PI to point to other information about the object to which the identifier has been assigned. This criterion is considered mandatory (M) in the proposed IF.

8. Managing the relations between the PIs within the domain.

This criterion identifies the possibility to specify the linkage between resources within the PI domain through explicit relations between their identifiers. For example, a PI domain can make the part-of relation between resources explicit by embedding this linkage within the PI string, or using metadata. An example of this kind of relation is that which exists between a resource and the collection of which it is part, the resource and some actors, or finally, the relation when a resource has multiple PIs assigned. This criterion is considered optional (O) in our framework, but it represents an added value that can speed up the implementation of interoperability services. This is because based on the “trusted” relationships information can be integrated across systems and content providers.

2.4 THE ONTOLOGY

Introduction

This section presents the new version of the APARSEN ontology coming out as the result of an activity of refinement carried out after the evaluation of the first release which was based on a wide analysis of the user needs.

IF validation and refinement

The revision activity of the initial model was discussed from June to October 2012 by the HLEG (see chapter 1.6). A very important step forward in the evaluation process was the workshop in Florence on 13 Dec 2012 (see paragraph 3.2). After a deep analysis of the work carried out under the

[17] http://www.crossref.org/CrossTech/2010/02/does_a_crossref_doi_identify_a.html

WP22, most of the experts recommended a redefinition of the main entities and properties of the ontology and a tighter compliancy with already existing standards. In particular they suggested to define the notions of "Actor" and "Object" in a more rational way and to precisely identify the different aspects these concepts can assume among the different application contexts. In this sense, the definition of specific "Actors" by means of detailed "roles" to be assigned to them was also suggested.

Many suggestions also concerned the notion of "digital object" and its identity. The use of concepts inspired to the one provided by FRBR, CIDOC-CRM and PREMIS were recommended. The experts agreed that making extensive reference to well established and widely used ontologies makes the approach more stable, more sharable, and less intrusive in the existing systems. They also suggested the development of test cases adhering to Semantic Web Technologies to assure scalability and flexibility to the whole framework. Taking into serious account the various recommendations and the suggestions of the expert group, we have tried to sketch a new ontology more suitable for semantic interoperability of PI domains.

To better understand the various entities involved, we have asked both the service and content providers to provide us with snapshots of their archives. In this way we could combine a top-down approach by analysing the above-mentioned ontologies (in particular FRBR) with a bottom-up approach by extracting categories from real data and context of use. In particular, we got information from:

- JLIS (Italian Journal of Library and Information Science)
- DANS (Data Archiving and Networked Services, The Netherlands)
- Fondazione Rinascimento Digitale
- NBN Italian national register (National Library in Florence)
- INSPIRE (bibliographic archive from CERN, Switzerland)
- DNB, Deutsche Nationalbibliothek, Germany)

Work on the APARSEN Ontology

The first operation carried out was the redefinition of the main entities already present in the first release of the ontology by using the corresponding concepts provided by one of the suggested international standards. We have chosen to test the potentialities of the FRBRoo ontology and of CIDOC-CRM on top of which FRBRoo is built, which in our opinion offers the highest degree of similarities.

FRBRoo[18] is a formal ontology designed to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information. It looks very suitable to be integrated in the IF since it provides entities and relationships to describe concepts coming both from the service provider and the content provider domains.

The new APARSEN ontology release inherits the main concepts of its predecessor, while trying at the same time to achieve better compatibility with international standards for enhancing the semantic aspects and the interoperability among the different data sources. This approach is also compliant with the interoperability guidelines proposed in WP25 suggesting reusing or integrating existing standards and models (if possible) to reduce interoperability dependencies and increase semantic integration between services and applications.

Inherited Main Entities

Most of the original APARSEN classes and relationships have been rearranged and expressed using the CIDOC-CRM/FRBRoo classes and properties. Additional CIDOC-CRM/FRBRoo concepts have been inserted where needed. In particular:

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

- The original “Persistent Identifier” APARSEN class has been mapped on the “E42 Identifier” class of the CIDOC-CRM ontology
- The original “Digital Object” APARSEN class has been mapped on the “F5 Item” class of the FRBRoo ontology
- The original “Author”, “Institution” and “RegistrationAgency” APARSEN classes have been mapped on the “E39 Actor” class of the CIDOC-CRM ontology

These three classes and the related properties already provide the basic concepts and can be seen as a core set of entities that institutions, and especially service providers, can use to reach a minimum degree of interoperability among each other. They also constitute the cornerstone of every encoding activity since it will be very easy to add entities and relationships in a coherent way on top of this core for future model enrichments. We have drawn some “application scenarios” showing the usability of the model in different and realistic use cases. A brief description of the whole activity follows. In naming the various entities and properties, the following conventions have been adopted:

- Labels starting with “**E**” refers to **classes** of the **CIDOC-CRM** ontology
- Labels starting with “**F**” refers to **classes** of the **FRBRoo** ontology

- Labels starting with “**P**” refers to **properties** of the **CIDOC-CRM** ontology
- Labels starting with “**R**” to **properties** of the **FRBRoo** ontology

Scenario 1: Registration Authorities and Persistent Identifiers

The first important aim of the APARSEN ontology is to define the nature of the Persistent Identifier, an instance of the *E42 Identifier* CIDOC-CRM class and the Registration Authority (*E39 Actor*) which created it, and to describe their mutual relationships.

The relation between PIs and RAs is performed through the insertion of an Identifier Assignment event (an instance of the *E15 class*), which makes the relation explicit and offers the possibility to use the features of a temporal entity for future extensions (e.g., specification of time and place of the assignment event, and so on).

The tight relation between the PI and the RA is also the basis on which the PI Resolution Service will be built on. In particular, the system will always know where to look for the resolution of a given PI thanks to the *P76 has contact point* property and the specifications of the *E51 Contact Point* instances (see Figure 5).

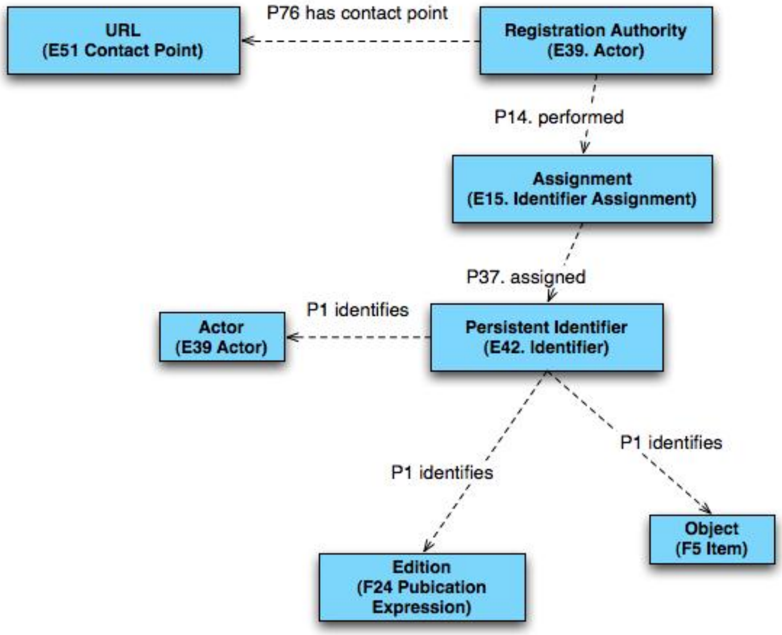


Figure 5: Registration Authorities and Persistent Identifiers

Scenario 2: Persistent Identifiers, Actors and Objects

We have also extended the description of the Persistent Identifier entity by giving specific information concerning its nature and use. In particular, in our vision a PI is an instance of the *E42 Identifier* class and is used to identify 3 kinds of entities in the scope of the APARSEN ontology:

1. **Actor** (instance of the E39 Actor): individual entities identified by a PI (for instance, by an ORCID identifier).
2. **Edition** (instance of *F24 Publication Expression*) represents the overall content of a work in terms of signs present in publications, reflecting the publishers' decisions as to both content and layout. Edition is identified for instance by an ISBN and DOI identifiers.
3. **Object** (instance of the *F5 Item* class): the physical or digital object that carries a Publication Expression. An Object can be identified for instance, by a CIDOC-ICOM code for museum objects.

The distinction between Object and Edition, not present in the preceding release, is crucial in capturing the essence of the different PI we deal with (e.g., the different entities identified by ISBN and DOI as we already pointed out) and to describe the resources identified by the PIs in an appropriate way. The Object is the carrier of the Publication Expression. Figure 6 shows the other relations interweaving these two entities and in particular the definition of the "Publisher" role, another instance of the *E39 Actor* class, by the CIDOC-CRM *P14* property.

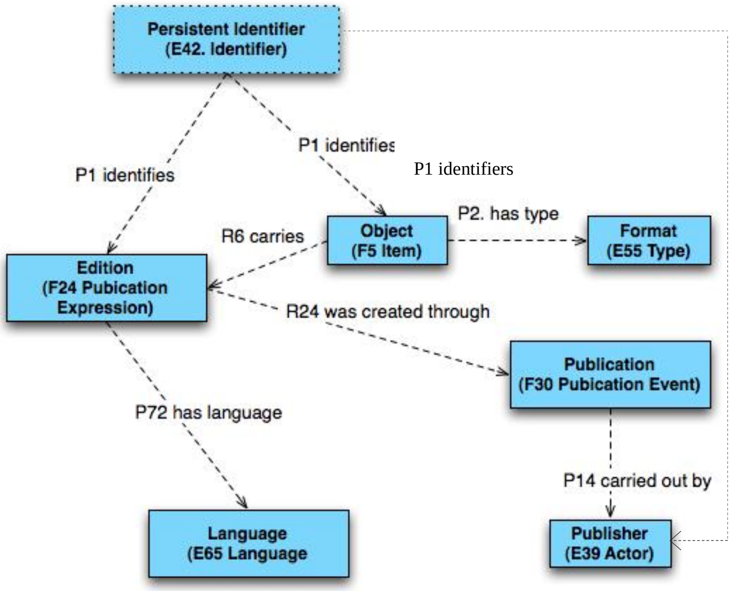


Figure 6: Persistent Identifiers and Actors and Objects

Scenario 3: Persistent Identifiers, Authors, and Works

A typical scenario to take into account for testing the ontology is the one in which an Actor becomes the "Author" of a work. FRBRoo provides the *F1 Work* class, which represents the conceptual object, a set of expressions evolving from an original idea, for instance the "Divina Commedia" by Dante Alighieri, without any regard to a specific edition. A Work may be created by one or more Actors, simultaneously or over time. A Work Conception event (*F27*) having a certain Time Span links the Actor(s) with the created Work. The Work itself, once created, has a Title, a Description and so on (see Figure 7).

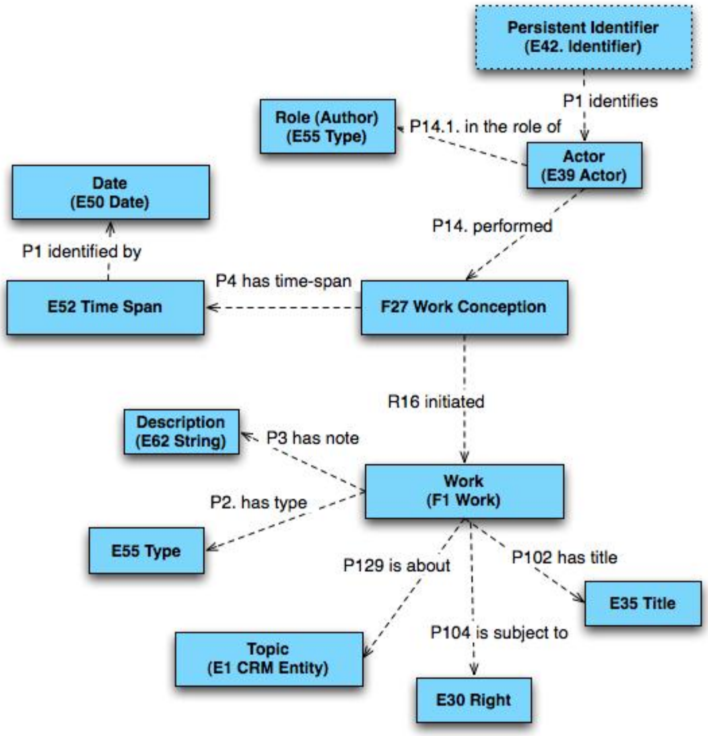


Figure 7: Actors and Works

Figure 8 summarises the interconnection of the three scenarios and exemplifies the role of the Edition (Publication Expression), which is on one side the realisation of the Work in terms of text, layout and graphic, and on the other side is carried by the Item which is the actual physical/digital object.

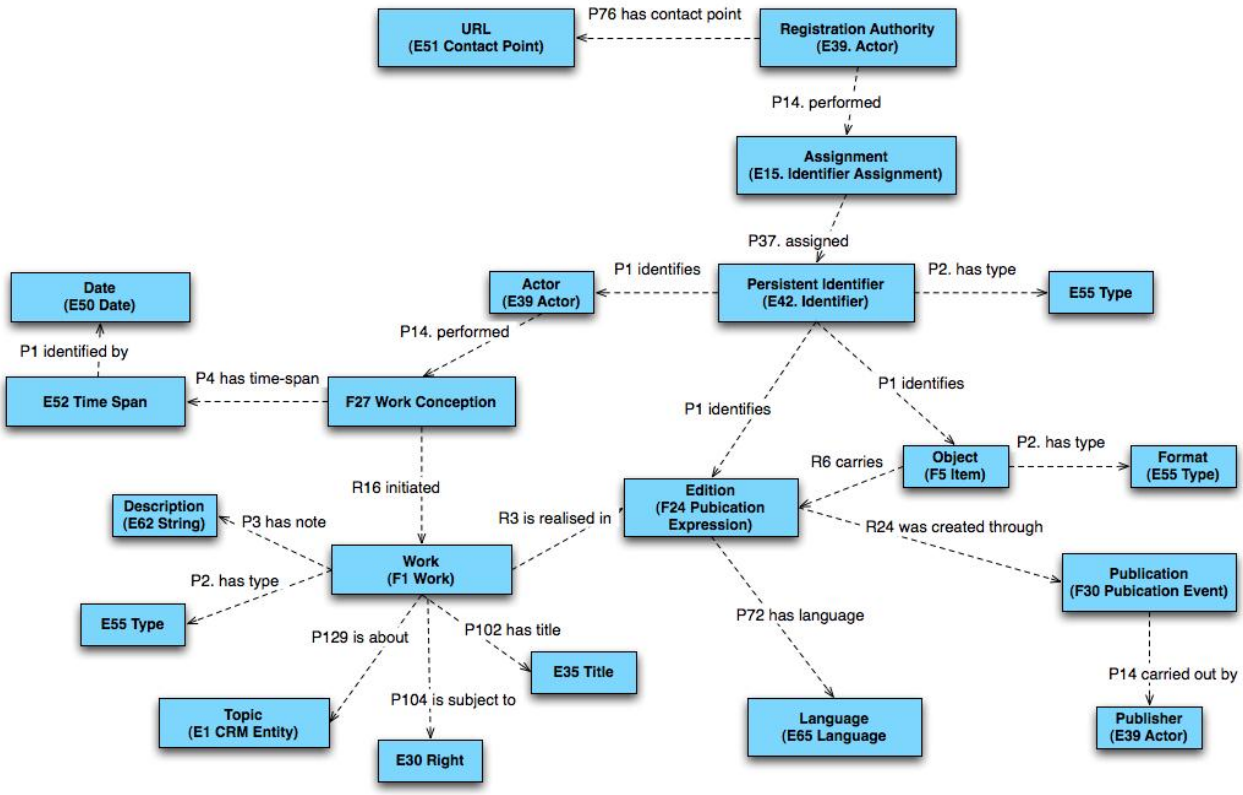


Figure 8: General overview and interoperability of the different scenarios.

Mapping Activities and the APARSEN Demonstrator

To test the capabilities of the new ontology we have created a prototype application able to demonstrate the capabilities to manage information coming from different archives and encoded using different metadata formats (in our case Dublin Core and MARC).

To create an interoperable context we have mapped the incoming DC and MARC information on the APARSEN entities. We were able to represent all the incoming metadata thanks to the flexibility of the FRBRoo/ CIDOC-CRM ontology (see Figure 9)

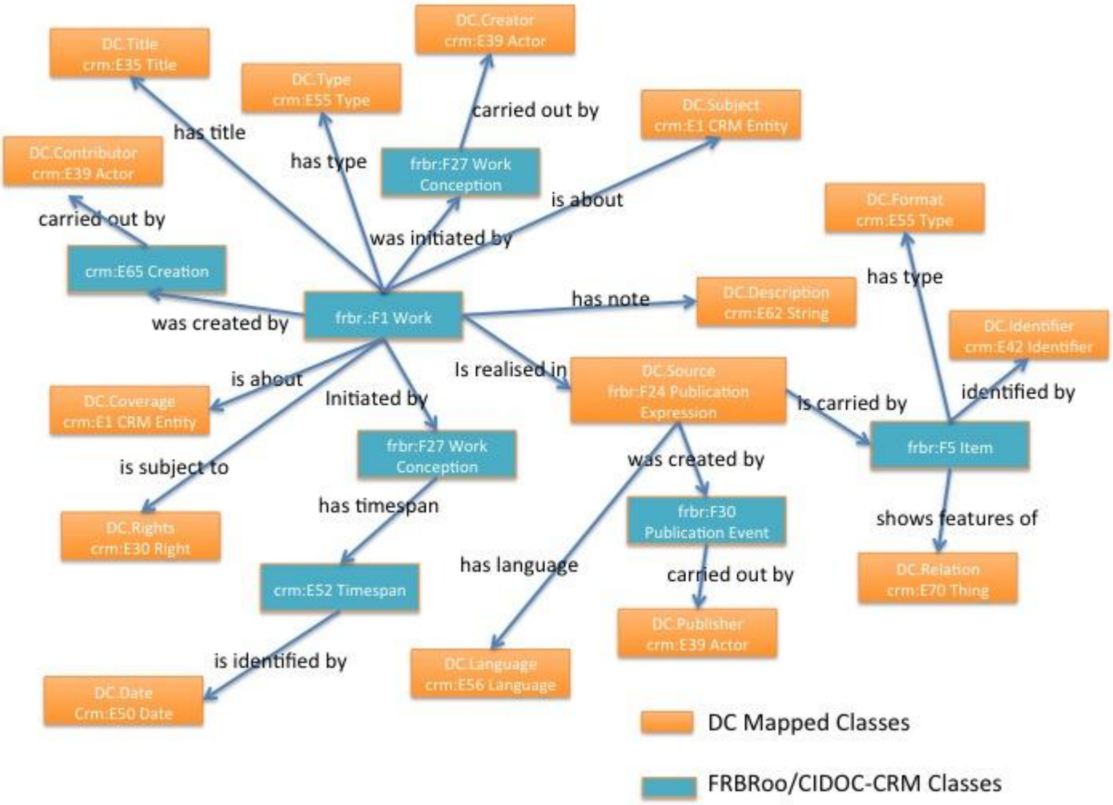


Figure 9: Dublin Core – FRBRoo Mapping

Dublin Core Element	FRBRoo/CIDOC Entity	FRBRoo/CIDOC Relations
DC.Title	crm:E35 Title	frbr:F1 Work crm:P102 has title crm:E35 Title
DC.Creator	crm:E39 Actor	frbr:F1 Work frbr:R16B was initiated by frbr:F27 Work Conception crm:P14 carried out by crm:E39 Actor
DC.Subject	crm:E1 CRM Entity	frbr:F1 Work crm:P129 is about crm:E1 CRM Entity
DC.Description	crm:E62 String	frbr:F1 Work crm:P3 has note
DC.Publisher	crm:E39 Actor	frbr:F24 Publication Expression crm:P94 was created by frbr:F30 Publication Event crm:P14 carried out by crm:E39 Actor
DC.Contributor	crm:E39 Actor	frbr:F1 Work crm:P94 was created crm:E65 Creation crm:P14 carried out by

		crm:E39 Actor
DC.Date	crm:E50 Date	frbr:F1 Work frbr:R16B was initiated by frbr:F27 Work Conception crm:P4 has timespan crm:E52 Timespan crm:P78 is identified by crm:E50 Date
DC.Type	crm:E55 Type	frbr:F1 Work crm:P2 has type crm:E55 Type
DC.Format	crm:E55 Type	frbr:F5 Item crm: P2 has type crm: E55 Type
DC.Identifier	crm:E42 Identifier	frbr:F1 Work crm:P1 is identified by crm:E42 Identifier / frbr:F5 Item crm:P1 is identified by crm:E42 Identifier
DC.Source	frbr:F24 Publication Expression	frbr:F1 Work frbr:R3 is realised in frbr:F24 Publication Expression
DC.Language	crm:E56 Language	frbr:F24 Publication Expression Frbr:P72 has language crm:E56 Language
DC.Relation	crm E70 Thing	frbr:F5 Item crm:P130 shows features of crm E70 Thing
DC.Coverage	crm:E1 CRM Entity	frbr:F1 Work frbr:P129 is about crm:E1 CRM Entity
DC.Rights	crm:E30 Right	frbr:F1 Work P104 is subject to crm:E30 Right

Table 3: Mapping DC- FRBRoo

Automatic Trust co-reference generation for PI interoperability

The stage after the metadata normalization through FRBRoo ontology is the co-reference generation among resources. The co-reference generated through <owl:sameAS> indicates that two URI references actually refer to the same thing: the individual objects have the same "identity".[19]

In the context of IF we assume that the PI assignment process in a trusted PI domain is reliable enough to guarantee that two different resources with the same PI associated could be considered copies of the same resource. This assumption is necessary to generate co-references automatically.

Using a trusted PI as a key field to generate co-reference instead of inferring it by applying semantic analysis on textual metadata fields like title, author, etc. is more reliable and reduces the risk of false-positive detection. Moreover, this process realizes the PI interoperability since it is possible to use one PI to retrieve all resources linked to the other PIs as shown in the Figure 10 below.

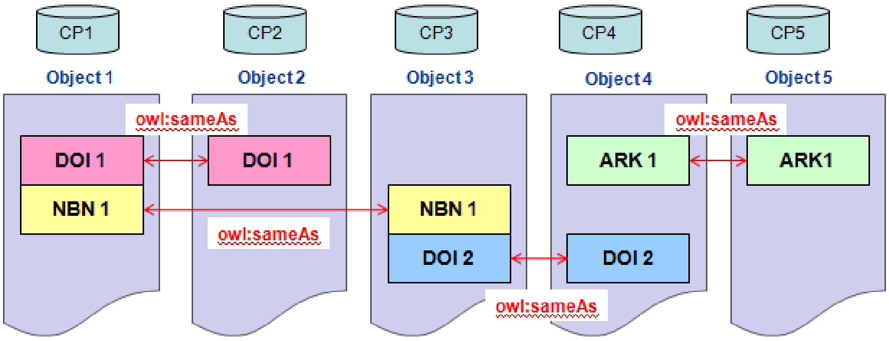


Figure 10: Trusted PI-based co-references

In this manner, PIs can be used **interchangeably**, and the resolution of each of them is able to retrieve the same list of linked resources. The Figure 11 below shows that given any PI (for example PI1) it is possible to get back the entire chain of resources that are linked together through co-reference relation.

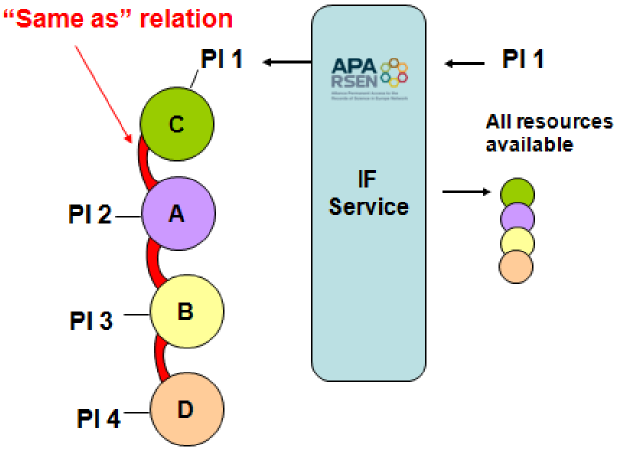


Figure 11: Trusted relations-based object retrieval

[19] http://www.w3.org/TR/owl-ref/#sameAs-def

This approach does not substitute other approaches like automatic co-reference generation through semantic analysis or manual generation. Instead it integrates the current practices exploiting PI interoperability as a source to enable reliable “same as” assertions.

3 DEMONSTRATOR

3.1 INTRODUCTION

In the framework of the WP22 work it is necessary to test and evaluate the proposed IF model to develop interoperability at the service level among independent PI domains. We decided to set up an operative demonstrator with two main goals:

I. To prove functionality and effectiveness of the IF, estimating also the amount of work for content providers to implement the model exposing their data in that format.

II. To provide a base of data entities and relations useful to demonstrate potential benefits for users through developing some services across different PI systems and in line with user needs and possible user expectations by PI systems

All the work on the IF model has been possible thanks to the HLEG. This group has been involved twice through a questionnaire to evaluate the IF model and make suggestions.

A first prototype was developed and submitted to the experts’ evaluation at the workshop on 12 Dec. 2012 in Florence. The second round of evaluation was carried out with the HLEG in summer 2013 and a refined version of the IF model was presented in Lisbon at the iPRES 2013 “PI interoperability framework workshop”. After the Lisbon workshop, a new strategy seemed necessary so we developed a second prototype and presented it at the All Hands Meeting (AHM) on 4 December 2013 in Den Haag (AHM is an annual event where all the APARSEN partners meet and exchange recent experience and results).

We briefly introduce the main aspects of the expert revision. First of all, there is a general consensus about the need to overcome the current fragmentation of PI landscape and to establish an interoperability strategy among the PI systems with clear benefits for end users foreseeable from new services across PI domains. Another relevant step forward is the agreement about the 4 main assumptions and the 8 trust criteria for PI systems. There are some doubts about the possibility that the IF is more a reference model than a practical schema ready for implementation.

The most serious criticism, however, is regarding the part of the model on metadata describing the content identified by the PI, because the FRBRoo schema is tailored to the library-specific objects and so it may not be easily adaptable to other user sectors such as archives, industries, museums, scientific research centres, public administrations and universities. Therefore, the demonstrator focuses on PIs and related information and avoids deep metadata descriptions.

3.2 WORKSHOP IN FLORENCE

http://www.rinascimento-digitale.it/workshopPI2012.phtml

I. Flyer + programme

TITLE Interoperability of Persistent Identifiers Systems Learning how to bring them together
LOCATION AND DATE Florence, 13 December 2012

Auditorium Ente Cassa di Risparmio di Firenze, via Folco Portinari 3/5
TOPIC Trusted, unique, certified and persistent identification of digital resources is a crucial component of a durable research infrastructure. A number of Persistent Identifiers (PI) technologies have been proposed and adopted by different user communities. The central topic of the workshop is to discuss how interoperability between different PI systems can be encouraged. The current situation where PI domains are isolated should be overcome and future scenarios and services stimulated. The current scenario is very fragmented and the PI domains work isolated de facto with serious problems and limits for the final users of Internet applications, so the main goal of the workshop is to stimulate cooperation among all PI user communities, the user needs and opportunities offered by a future scenario where each PI domain exposes data in a common format with all the other PI domains, without changing anything for internal organisation and policy. The new Interoperability Framework (IF) proposed by the APARSEN project and refined by a large group of independent experts is presented and a preliminary demonstrator is given, the IF model suitable to all the different user requirements and that is adoptable by all PI user communities serves to demonstrate potential benefits for final users. Representatives of different PI initiatives are invited to report on the current state of art and expose their position towards needs and opportunity of interoperability among PI systems. Participants will be invited to compare their requirements with the IF confronting on various aspects of the model, potential benefits and concrete terms for implementation of the framework in order to create consensus on a common platform to develop joint applications. During the workshop representatives of PI domains have the opportunity to bring their experience, plans and point of view as well as their requirements in respect to a model for interoperability with other PI domains.
REFERENCE The program and presentations of the workshop can be found at: http://www.rinascimento-digitale.it/workshopPI2012.phtml

II. Supporters + participants

Samuele Carli	CERN
Uta Ackermann	German National Library
Stina Degerstedt	National Library of Sweden
Roberto Delle Donne	CRUI – Datacite
René van Horik	DANS
Piero Attanasio	AIE/Medra
Nicole von der Hude	German National Library
Martin Braaksma	DANS - NBN Cluster
Mark van de Sanden	EuDAT – EPIC
Marco Scarbaci	ICCU
Laure Haak	ORCID
Ilara Fava	CINECA
Gabriella Scipione	CINECA
Ernesto Damiani	University of Milan
Claudio Prandoni	Promoter
Caterina Guiducci	University of Florece
Carlo Meghini	CNR – PRELIDA
Bonaria Biancu	University of Milan Bicocca

Bengt Neiss	National Library of Sweden
Barbara Bazzanella	University of Trento
Annelies Noelmans	University of Leuven
Anila Angjeli	ISNI – VIAF
Alex Siedlecki	Museo di Arte e Cultura Tibetana
Federica Tosini	University of Padova
Ye Cao	Max Planck Digital Library
Jan Molendijk	Europeana Foundation
Teresa Ko	Hong Kong Central Library
Vittore Casarosa	CNR
Veronika Praendl-Zika	ONB
Tommaso Agnoloni	ITTIG
Stefania Arabito	University of Trieste
Paolo Budroni	University of Wien
Oreste Signore	W3C
Mauro Guerrini	University of Florence
Maurizio Lunghi	FRD
Matteo Bertazzo	CINECA
Kakia Chatsiou	ELAR-SOAS- University of London
Juha Lehtonen	CSC - IT Center for Science Ltd.
Jordan Pi čanc	University of Trieste
Giulia Colombo	GAP s.r.l.
Esa-Pekka Kesk talo	National Library of Finland
Emanuel Bellini	FRD
David Giaretta	APA
Cin ia Luddi	FRD
Antonella Farsetti	University f Floren e
Achille Felicetti	FRD
Luisa G ggini	Cas lini Libri

III. Workshop conclusions

On December 13 2012 the international workshop “Interoperability of Persistent Identifiers” was held in Florence. There is a general agreement that trusted, unique, certified and persistent identification of digital resources is a crucial component of a durable research infrastructure. However, a number of Persistent Identifiers (PI) technologies have been proposed and adopted by different user communities and the current situation appears very fragmented resulting in aurgent need of interoperability among the available solutions. The central topic of the workshop is to discuss how interoperability between different PI systems can be encouraged.

The workshop consisted of three parts:

1. Presentation and discussion of an Interoperability Framework (IF) for Persistent Identifier (PI) systems developed by the APARSEN project

2. Round table on interoperability issues related to a number of PI systems

3. PI initiatives, state of art and future plans (see the presentations on the workshop website).

Persistent Identifiers (PI) are a crucial component of a durable research infrastructure. Different user communities have developed a number of PI systems, e.g. for the identification of physical and digital objects, for persons, publications or datasets. These PI systems are based on different principles to compile identifiers, to relate them to objects and to resolve the relation between the identifier and the object.

As stated above, the workshop was aimed to discuss potential benefits and conveniences to defragment the current situation and how interoperability between different PI systems can be encouraged. To this purpose, the APARSEN project has proposed an Interoperability Framework (IF). The rationale for the development on an Interoperability Framework is defined as follows by the APARSEN project: "To enable the persistent access, reuse and exchange of information through the use of existing identifiers and associated resources across different systems, locations and services".

Background information on the IF was distributed to the participants of the workshop in advance. An overview of the IF can be found in the article:

Barbara Bazzanella, Emanuele Bellini, Paolo Bouquet, and others, Interoperability Framework for Persistent Identifier Systems in: Proceedings of the 9th international conference on preservation of digital objects (iPRES 2012), see: https://ipres.ischool.utoronto.ca/. (page 29-36). A number of participants of the workshop participated in a survey on interoperability issues of PIs and have reviewed the IF documentation and principles in advance. Based on the theoretical foundations as described in the article a demonstrator was developed as a next step to test the applicability of the IF.

Important for the understanding of trusted PI systems are the following 8 criteria:

1. Having at least one Registration Agency.
2. Having one Resolver accessible on the Internet.
3. Uniqueness of the assigned PIs within the PI domain.
4. Guaranteeing the persistence of the assigned PIs.
5. User communities of the PI domain should implement policies for digital preservation (e.g. trusted digital repositories)
6. Reliable resolution.
7. Uncoupling the PIs from the resolver.
8. Managing the relations between the PIs within the domain.

The presentation of the PI Interoperability Framework (IF) started with an overview of the main issues related to PIs with an emphasis on the issues related to the realisation of interoperability between individual PI systems. The development of the IF is based on the following assumptions:

- Entities have to be defined by at least one PI. More than one PI can refer to the same entity.

- Only PI domains that meet some quality criteria are eligible to be considered in the IF. These quality requirements are related to trustworthiness.

- The responsibility to define relations among resources and actors is delegated to the trusted PI domains

- Digital preservation issues are not addressed directly by the IF.

Some of the PI domains for which the IF can be relevant are: Handle, ARK, DOI, NBN, ORCID, VIAF and ISNI.

The three main steps to develop the IF - that is

1. Design and validation of the IF model through a user group of about 30 experts. This is done.
2. Definition and setup of a demonstrator with data from different PI domains (objects, people, bodies). This is the main topic of the workshop.
3. Proposal for development of PI interoperability services. For this follow-up activities will be organised.

- were presented and discussed during the workshop.

The workshop was important for the second step in this process. Further cooperation and agreement is required to realise step 3. This is the focus of the second workshop in Lisbon described below.

The coverage of the IF demonstrator consisted of two parts. First the theoretical foundations were presented. Next the actual IF demonstrator was presented.

A requirement for the creation of a demonstrator to demonstrate the feasibility of an IF is the formulation of a shared ontology. This shared ontology represents the identified resources and their mutual relationships, relevant for the individual PI systems. For this existing standards were evaluated. Two candidates for this turned out to be:

- FRBR entity-relation model designed by IFLA (FRBR: Functional Requirements for Bibliographic Records) is a 1998 recommendation of the International Federation of Library Associations and Institutions (IFLA) to restructure catalog databases to reflect the conceptual structure of information resources).

The main ontologies of the library and museum community can be expressed in FRBRoo. This is a formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information. (See: http://www.cidoc-crm.org/frbr_inro.html).

The data providers that provided input for the demonstrator were: CERN, DANS, FRD and the JLIS Open Access Publisher. Most of the repositories use the Dublin Core Metadata Element Set. A number of scenarios were presented to facilitate the mapping of the local used ontologies (e.g. Dublin Core) with the FRBRoo ontology that is the backbone of the IF service. These scenarios are:

1. A central registration authority issues new identifiers to registered users and maintains the registry of identifiers.
2. PI, actors, works
3. Identifiers and objects

Further work is needed to complete the mapping of the ontologies. For this additional test cases and applications are required, as well as additional entities and properties. Also the vocabulary of types and roles has to be extended. A suggestion for further work is also to publish the bibliographic information as Linked Open Data.The theoretical foundations described above were followed by a presentation of the IF demonstrator.

http://www.rinascimento-digitale.it/workshopPI2012.phtml

3.3 WORKSHOP IN LISBON

http://www.alliancepermanentaccess.org/index.php/aparsen/aparsen-workshops/workshop-on-persistent-identifiers-ipres-2013/

I. Flyer + programme

TITLE APARSEN workshop at 10th International Conference on Preservation of Digital objects (http://ipres2013.ist.utl.pt), Lisbon, Portugal. Interoperability of Persistent Identifiers Systems – services across PI domains.
LOCATION AND DATE **Lisbon, Portugal** **CAMPUS ALAMEDA of the IST - "Instituto Superior Técnico"**

Thursday, September 5, 2013

Background

Trusted, unique, certified and persistent identification of digital resources is a crucial component of a durable research infrastructure and also of a future information society. A number of Persistent Identifiers (PI) technologies have been proposed for different types of objects and adopted by different user communities. Currently these systems are isolated and provide different levels of service, though awareness of the need for dialog between them is rising.

The current situation where PI domains are isolated should be overcome and future scenarios and services stimulated. The current PI landscape is very fragmented and the PI domains work isolated de facto with serious problems and limits for the final users of Internet applications. Therefore, a coordinated effort is needed to stimulate cooperation amongst all PI user communities and to identify user needs and opportunities offered by a future scenario where each PI domain exposes and shares data in a common format with all the other PI domains, without changing anything for internal organisations and policy.

Workshop goal

The central goal of this workshop is to bring together representatives from different Persistent Identifier communities to discuss potential benefits of PI interoperability for end users, as well as the challenges, requirements and technologies needed to implement an effective interoperability solution for different PI systems and related services. The workshop is a follow-up of a first workshop on this issue organised in December 2012 in Florence. (See: http://www.rinascimento-digitale.it/workshopPI2012). Supporters of this workshop proposal and the experts on the program committee represent large and significant PI user communities.

The workshop will discuss the report “Interoperability Framework for PI systems” (PDF document / 4,7 MB) Interoperability Framework for PI Systems: Evaluation of the Model by the HLEG (60).

The first part of the workshop is devoted to potential services and benefits for end users that could be built on such an interoperability framework. Participants are involved in the description of future user scenarios and potential applications of the PI systems, making evident user benefits and requirements.

The second part of the workshop is focused on technical aspects regarding the implementation of an interoperability solution and related services. As a starting point for the technical discussion, the Interoperability Framework (IF) proposed by the APARSEN project and refined by a large group of independent experts is described and a demonstrator is presented. The IF model is suitable to all the different user requirements and is adoptable by all PI user communities serves. Participants are invited to discuss their requirements as compared with the IF features and assumptions confronting various aspects of the model, potential benefits and concrete terms for a common roadmap for the implementation of the framework in order to create consensus in developing joint applications to achieve interoperability across domains.

Representatives of the most relevant PI initiatives and different PI user communities are invited to report on current activities and their vision, but also on possible approaches to define interoperability solutions and services and expose their position towards the needs and opportunities of moving toward the implementation of a comprehensive interoperability technological solution for all PI systems.

REFERENCE

http://www.alliancepermanentaccess.org/index.php/aparsen/aparsen-workshops/workshop-on-persistent-identifiers-ipres-2013/

II. Supporters + participants

Mark van de Sanden	SURF sara
Bob Bailey	Thomsom Reuters
Antoine Isaac	Europeana
Marco Kindt	Zuse Institut Berlin
Dimitar Dimitro	GESIS
Gildas Illien	BNF
John Kunze	CDL
Sean Martin	British Library
Stefan Proell	SBA Research
Sophie Derrot	BNF
Tobias Weigel	DKRZ
Barbara Bazzanella	University of Trento
Juha Hakala	National Library of Finland
Anila Angjeli	BNF
Laure Haak	ORCID
Maurizio Lunghi	FRD
René van Horik	DANS

III. Workshop conclusions

The aim of the workshop is to present and discuss the current state of art of the “Persistent Identifiers Interoperability Framework” as developed by the APARSEN project. The Framework is described in the report “Interoperability Framework for Persistent Identifiers”. Next to that a number of related initiatives report on the current state of affairs. A round of table was done as introduction of participants.

The participants represent PI initiatives, projects, infrastructures, publishers, researchers, content owners.

The workshop started with a presentation of the APARSEN WP22 work on Interoperability Framework (IF) of PI systems. The proposed approach has the main goal to defragment the current situation of PI systems and is based on an idea of interoperability only in terms of capacity to access the information from different domains in the same way and same level of service. This can be done – as it is proposed – by making data accessible from all the PI domains in the same format, i.e. expressing the relationships between the identified entities according to the same schema, and following some fundamental assumptions and criteria of trust.

Some preliminary results from the APARSEN survey about current practices of PIs and

user requirements and needs of PI services have been presented. First of all, users see the current fragmentation of PI systems as an obstacle to the cross-boundary information sharing and integration. Secondly, there is a large request of services across PI – domains, in particular 1) global resolution, 2) citability and metrics and 3) certification services (see BaB presentation for details). One evident result is that the PI systems cannot continue to simply offer the resolution to the URL of the resource, the PI systems must evolve towards a condition of ‘added - value service providers’. We listed also the most relevant criteria requested by users of PI systems. As an outcome of the APARSEN work we have proposed a draft model of an interoperability framework (IF) for PI systems including entities, properties and relations, as well as criteria for trusted PI systems.

This IF model has been evaluated in two different phases by a High Level Expert Group (HLEG) on PIs with many experts not belonging to the APARSEN consortium (see in Annex the list of names). The basic idea is not to request any changes in the PI domain internal management, but to ask each PI system to expose their data in a common format to make possible the access and meaningful use of information from any domain at the same level of service. On that base and in order to give evidence of the potential benefits for final users of PI systems, a demonstrator has been developed.

Maurizio Lunghi presented the Interoperability Framework and the demonstrator that provided a number of use cases (see the slides).

Comments after presentation

- Many participants confirmed that the basic approach and goals of the IF are desirable and useful for users of PI systems.

- There was also a general consensus about the need of cross-domain and added value services in respect to the current fragmented situation in some cases confirmed also by the standardisation - bodies activity.

- Multiple PIs for the same resource can also be accepted and maybe useful but a need of interoperability between different identification systems is a crucial need (see interoperability initiatives between ORCID and ISNI and between ISNI and VIAF).

- However, the FRBRoo model adopted by the IF has been evaluated too complex and too oriented to the library world (AI). A possible solution could be “to prune the ontology” focusing on the part describing PIs. But the side effect of this simplification, has been remarked, is to loose some key relationships.

- Lunghi remarked that the current situation is immature; most of the PI systems simply resolve the ID to a URL, that’s not enough. The APARSEN approach is that PI systems must evolve towards ‘added-value service providers’ exploiting the information they have. Most of the participants agreed on this vision.

- In order to make accessible and usable all the PI domains at the same level of service we should identify a ‘lowest common denominator’ or in other words the minimum list of data that each PI systems should expose in relation to each ID. After that, the discussion moved to metadata, some participants recommended not to deepen in the content providers area, for example keep in mind that only the owner of the data can adjust the documentation.

- A proposal of exploiting LOD strategies and representation schemes to extract meaningful relationships between data and relevant entities have been made. However, the lack of trust on LOD “same as” relationships casts some doubts about the feasibility of the approach. The IF aims at enabling a layer of trusted relationships, which can be exploited to integrate information across systems and domains through added-value services.

A general consensus was to focus on further development of the IF model on the PI systems and participants requested Lunghi to provide a scenario with definition of entities & roles (e.g. ‘PI domains’ and ‘PI managers’ and ‘content owners’) and a clear distinction among PI systems for people, digital objects, physical objects and bodies; after that it is necessary to define the minimum list of data that each PI must use to expose its identifiers.

Finally, a presentation of the APARSEN demonstrator was provided including two basic services.

The demonstrator is aimed to present the potential benefits of the IF implementation by the PI domains, like for example DOI, NBN, ARK, and some basic services are under development for testing user satisfaction and requirements during the last year of the project 2014 and beyond. In particular, in the current version of the demonstrator contents are from the following PI domains, DOI, NBN-IT,NBN-DE, NBN-NL, ORCID, ISNI, and the first 2 services have been implemented as explained here below.

Service 1: the user inputs a PI for a person, like an ORCID or an ISNI, and the system gives back the info profile of the person, if this person has multiple IDs for people the system presents all of them, and the demonstrator can retrieve e.g. publications related to this person from all the PI domains, like DOI, NBN-IT, NBN-DE, Handle, ARK.

From the list of items the user can access info related to the works from each PI domain, for example for a resource with a NBN-IT the system accesses the info through 'Magazzini Digitali' the legal deposit of the national library in Florence.

Service 2: the user inputs a PI for a digital object, like a DOI or a NBN, and the system gives back the info about this object, if this resource has multiple IDs for the digital object, the system presents all of them, and the demonstrator can retrieve a list of copies of this resource from all the PI domains, like DOI, NBN-IT, NBN-DE, Handle.

3.4 ADDED VALUE SERVICES SELECTION

In the Deliverable 22.2 (see Table 2) a number of concept services have been designed on the base of the given scenarios (D22.2) for exploiting the IF functionalities.

Using SOMF (Service-Oriented Modelling Framework) notation we have designed a number of Atomic, Composite and Cluster concept services. In particular we defined the three classes of services as follows: 1) **Atomic service** as a software component that is indivisible because of its high granularity and performs a smaller number of functionality; 2) **Composite services** can be seen as intermediate services that are built from the combination of basic services and can be envisioned as constituting an intermediate layer on top of the core of the IF which enables the advanced cluster services; 3) **Cluster service** is a collection of related services that are distributed and collected because of their mutual business or technological characteristics.

Among the services proposed in D22.2, we have selected a service representative of each of the three classes of services identified during the survey: 1) Citability and Metrics Services, 2) Global Resolution Service and 3) Digital Object Certification (see D22.1 table 2). In the present document we present the implementation of 2 of them related to the class 1 and 2, while the implementation of the class 3 service is planned for the year 4 of the project. The services implemented take into account the current implementation level of the systems in place at the institutions participating in the consortium and willing to join the demo. Hence, for the demo we have focused the **PI-alternative PIs** as an example of atomic service as representative of class 2 services, and the **Entity Relationship Service** as example of Composite service as representative of class 1 services.

- Global Resolution Service: the **PI-alternative PIs** service implemented in the demonstrator associates a PI to alternative PIs for the same resource. In the description of the IF in D22.1 and in the chapter 3 of this document, we stressed that a resource can be identified by more than one PI (e.g., a document can be identified by a DOI and by a URN) and it can be connected with other resources having the same PIs (at least one) and using them as trust linkage key for building <owl:sameAs> relationships. The functionality of discovering alternative identifiers for the same resource is a fundamental requisite for the IF because it guarantees multiple ways to access the resource and related information, making the resolution process even more persistent. Moreover, having an access point to alternative PIs is a prerequisite for building intermediate services which can exploit the alternative identifiers to extract new (i.e. implied) relationships between relevant entities and related information, and consequently integrate information across systems boundaries.

- Citability and Metrics Services: the **Entity Relationship service**, defined in D22.2, uses the PIs of a target entity (e.g. Actor) to retrieve all the entities related to the target entity (e.g. given a PI-do, you retrieve all actors associated; given a PI-ac you retrieve all PI-acs and objects associated). We consider this service a composite service because it exploits the alternative PI service, which is a basic service, and uses the alternative PIs to extract the relevant relationships addressing a unique interoperability task.

3.5 CURRENT PROTOTYPE

Live prototype on **http://93.63.166.138/demonstrator/demo7/**

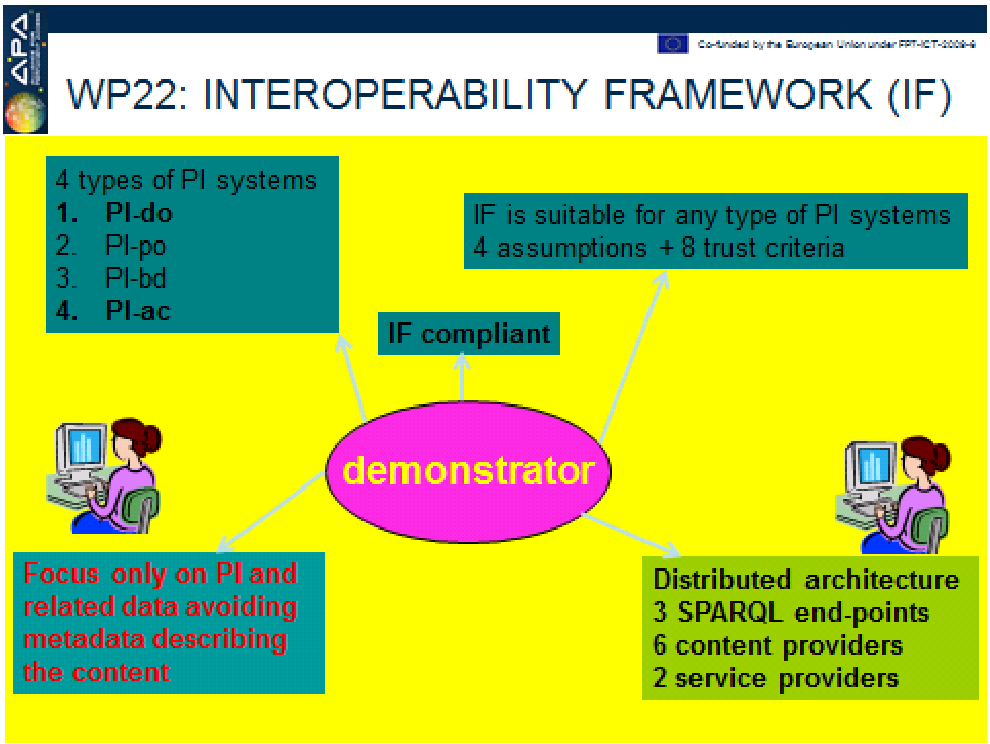

Co-funded by the European Union under FP7-ICT-2009-6
WP22: INTEROPERABILITY FRAMEWORK (IF)
4 types of PI systems
1. PI-do
2. PI-po
3. PI-bd
4. PI-ac
IF is suitable for any type of PI systems
4 assumptions + 8 trust criteria
IF compliant
demonstrator
Focus only on PI and related data avoiding metadata describing the content
Distributed architecture
3 SPARQL end-points
6 content providers
2 service providers

Figure 12: Demonstrator schema

Following the discussion at the iPRES workshop in Lisbon and the useful suggestions received by the experts we propose a light modification of the implementation strategy in respect to the former prototype. In particular, the current demonstrator will focus more on PI service providers instead of content providers.

Looking at the conclusions of the workshop in Lisbon it is evident that most of the experts agreed on the importance and convenience of creating conditions for interoperability of different PI domains, the relevance of having cross-domain and added value services with respect to the current fragmented situation. Most of the experts agreed also on the basic approach and criteria proposed in the IF. It was remarked that the current situation is immature; most of the PI systems simply resolve the ID to a URL, which is not enough. The APARSEN approach is that PI systems must evolve towards ‘added-value service providers’ exploiting the information they have. Most of the participants agreed on this vision.

Two main topics were discussed.

First, the FRBRoo ontology is very specifically oriented to the library world and it may be not suitable for other scientific sectors of content providers. A possible solution could be “to prune the ontology" focusing on the part describing PIs. But the side effect of this simplification is to lose some key relationships.

Second, when you enter in the field of metadata for content providers you must be careful to respect roles and responsibilities and to avoid duplicating data and functions. Following this last point, now the challenge is to identify the minimum set of data that PI managers must use to expose their content in a format common with all the other PI managers. In fact, to make accessible and usable all the PI domains at the same level of service we should identify a 'lowest common denominator' or in other words the minimum list of data that each PI system should expose in relation to each ID. The list of data will be specific for any type of PI domain; in particular it will be different for PI systems for actors and PI systems for digital objects.

The APARSEN WP22 team agreed on the following considerations for the next steps:

1. The IF system remains based on 4 elements, namely the PI systems definition, the 4 assumptions, the 8 trust criteria, and the model ontology. The first 3 elements have received a general consensus by the HLEG.
2. The model ontology proposed for the IF, based on the FRBRoo, remains as a reference model for long term development of the activity, even if some lacking features are already known.
3. While the demonstrator remains in line with the IF model and the proposed ontology, it should be 'reduced' in terms of entities, relations and functions, and should focus more on PI service providers instead of content providers. A minimum list of data for PI service providers is needed for the demonstrator development.

As we said, we consider different types of PI systems. The current list includes (see the Glossary for description):

1. PI for digital objects → PI-do
2. PI for physical objects → PI-po
3. PI for bodies → PI-bd
4. PI for actors → PI-ac

In the current demonstrator we implement only the PI systems for digital objects (PI-do) and the PI systems for actors (PI-ac).

To clarify what we consider the essential data that must be exposed to provide information about some PI practices, we make two examples of the PI-do and PI-ac that are implemented in the current demonstrator. This approach is focused on PI providers but can of course be also adopted by any content provider; in fact they can publish only data related to the PI or embed that in a wider schema presenting other information.

When a **PI-do** provider or a content provider decides to publish some contents on the IF framework, it must expose the following data:

1. Basic information about the resource associated to the PI-do and its simple description, as well as basic information about the PI-do generation and ownership.
2. Provide other PI-do associated with the same resource. This is the case when a PI-do is associated with a resource that has already a PI-do and so the reference is evident, or otherwise the PI-do managers are aware from a different source about the same-as relation.
3. Provide PI-do of other resources related to the resource associated to the original PI-do explaining the relation between the resources. For example, translation or new edition or series publications, etc.
4. Provide PI-ac of actors associated to the resource related to the PI-do and explaining the role, this is the relation with actors like author, editor, reviewer, curator, etc.

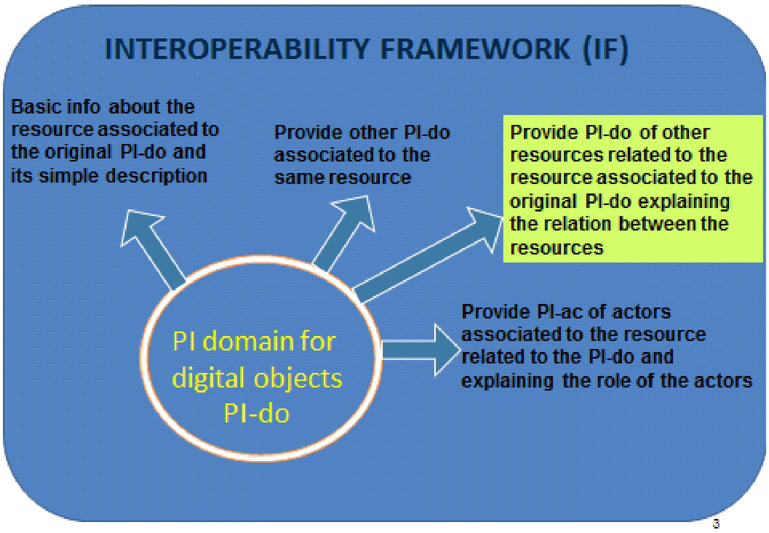

INTEROPERABILITY FRAMEWORK (IF)
Basic info about the resource associated to the original PI-do and its simple description
Provide other PI-do associated to the same resource
Provide PI-do of other resources related to the resource associated to the original PI-do explaining the relation between the resources
Provide PI-ac of actors associated to the resource related to the PI-do and explaining the role of the actors
PI domain for digital objects PI-do
3

Figure 13: PI-do in IF

When a **PI-ac** provider or a content provider decides to publish some content on the IF framework, it must expose the following data:

1. Basic information about the actor associated to the PI-ac.
2. Provide other PI-ac associated to the same actor. This is the case when an actor has multiple PI-do and so the reference is evident, or otherwise the PI-do managers are aware from different source about the same-as relation like in the case of ISNI and ORCID that are going to exchange mutual information about their databases; or the case when an actor has made the relationship with other AC explicit, a in the case of ORCID as Scoups Author Id or Researcher ID..
3. Provide PI-do of any resource associated to the original PI-ac explaining the relation between the resources and the role of the actor. In this case the system must provide also PI-ac related to other actors associated to the identified resources.

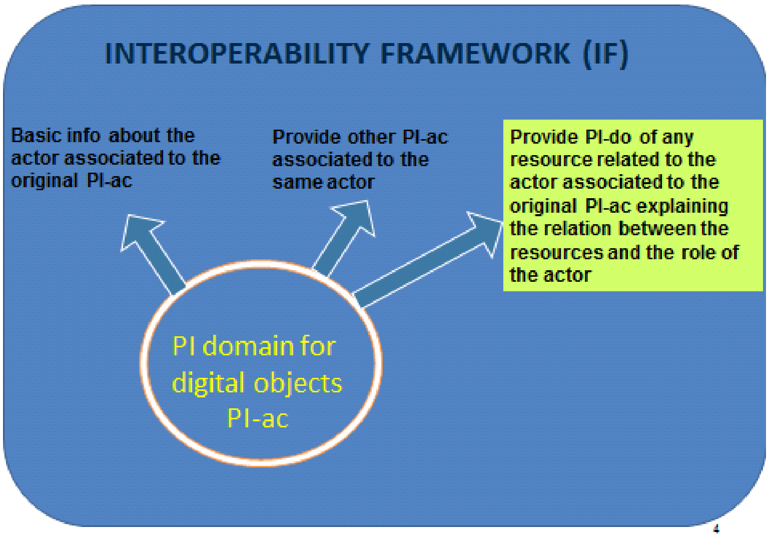

INTEROPERABILITY FRAMEWORK (IF)
Basic info about the actor associated to the original PI-ac
Provide other PI-ac associated to the same actor
Provide PI-do of any resource related to the actor associated to the original PI-ac explaining the relation between the resources and the role of the actor
PI domain for digital objects PI-ac
4

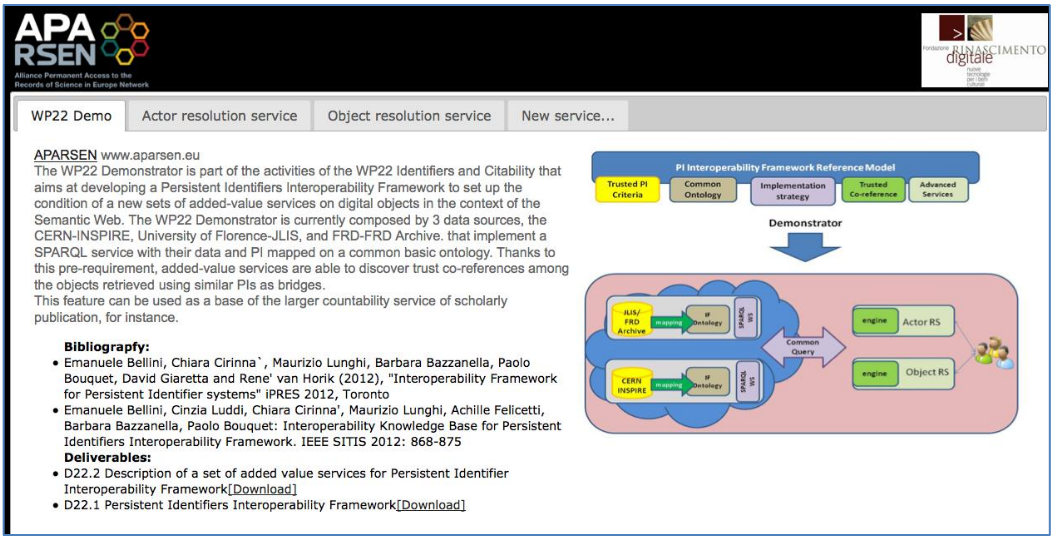
Figure 14: PI-ac in IF

The demonstrator implements the IF ontology in a RDF triple store mechanism and exposes them through a SPARQL end-point. The system exposes co-references among records in the knowledge base using information obtained by content providers. Then, two services were developed to test the framework functionality and demonstrate the IF potential application:

1. **PI alternative PIs of related objects (called Object Resolution Service in the online prototype)**. The service gets all the objects related to a given object identified by the PI in input. This functionality guarantees multiple ways to access the resources and related information, making the object retrieval process reliable. For example, as shown in Figure 16 entering a DOI for a given digital object, the service provides in output a brief description of the identified object and the list of alternative PIDs through which the object can be accessed.

2. **Entity Relationship service (called Actor Resolution Service in the online prototype)**. The service gets as input the PI of the author and provides as output the list of publications (metadata) taken just once. This service retrieves all the objects associated with an author, grouping the same objects in a unique view (discovering the *same as* relations based on trusted PIs). The number of items in the list corresponds to the number of publications of the identified author without repetitions. A screenshot of the output provided by the service for a given actor PI is shown in Figure 17.

Live prototype on **http://93.63.166.138/demonstrator/demo7/**



APARSEN www.aparsen.eu
The WP22 Demonstrator is part of the activities of the WP22 Identifiers and Citability that aims at developing a Persistent Identifiers Interoperability Framework to set up the condition of a new sets of added-value services on digital objects in the context of the Semantic Web. The WP22 Demonstrator is currently composed by 3 data sources, the CERN-INSPIRE, University of Florence-JLIS, and FRD-FRD Archive. that implement a SPARQL service with their data and PI mapped on a common basic ontology. Thanks to this pre-requirement, added-value services are able to discover trust co-references among the objects retrieved using similar PIs as bridges.
This feature can be used as a base of the larger countability service of scholarly publication, for instance.

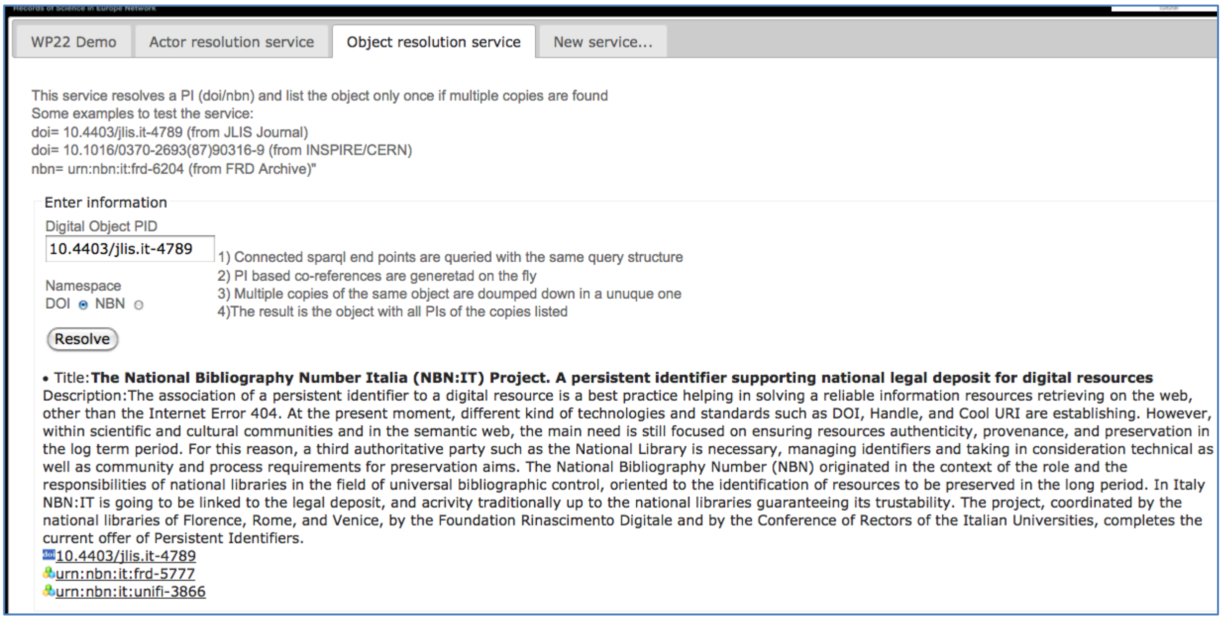
Bibliograpfy:

- Emanuele Bellini, Chiara Cirinna`, Maurizio Lunghi, Barbara Bazzanella, Paolo Bouquet, David Giaretta and Rene' van Horik (2012), "Interoperability Framework for Persistent Identifier systems" iPRES 2012, Toronto
- Emanuele Bellini, Cinzia Luddi, Chiara Cirinna', Maurizio Lunghi, Achille Felicetti, Barbara Bazzanella, Paolo Bouquet: Interoperability Knowledge Base for Persistent Identifiers Interoperability Framework. IEEE SITIS 2012: 868-875

Deliverables:

- D22.2 Description of a set of added value services for Persistent Identifier Interoperability Framework[Download]
- D22.1 Persistent Identifiers Interoperability Framework[Download]

Figure 15: Demonstrator home page



This service resolves a PI (doi/nbn) and list the object only once if multiple copies are found
Some examples to test the service:
doi= 10.4403/jlis.it-4789 (from JLIS Journal)
doi= 10.1016/0370-2693(87)90316-9 (from INSPIRE/CERN)
nbn= urn:nbn:it:frd-6204 (from FRD Archive)"

Enter information

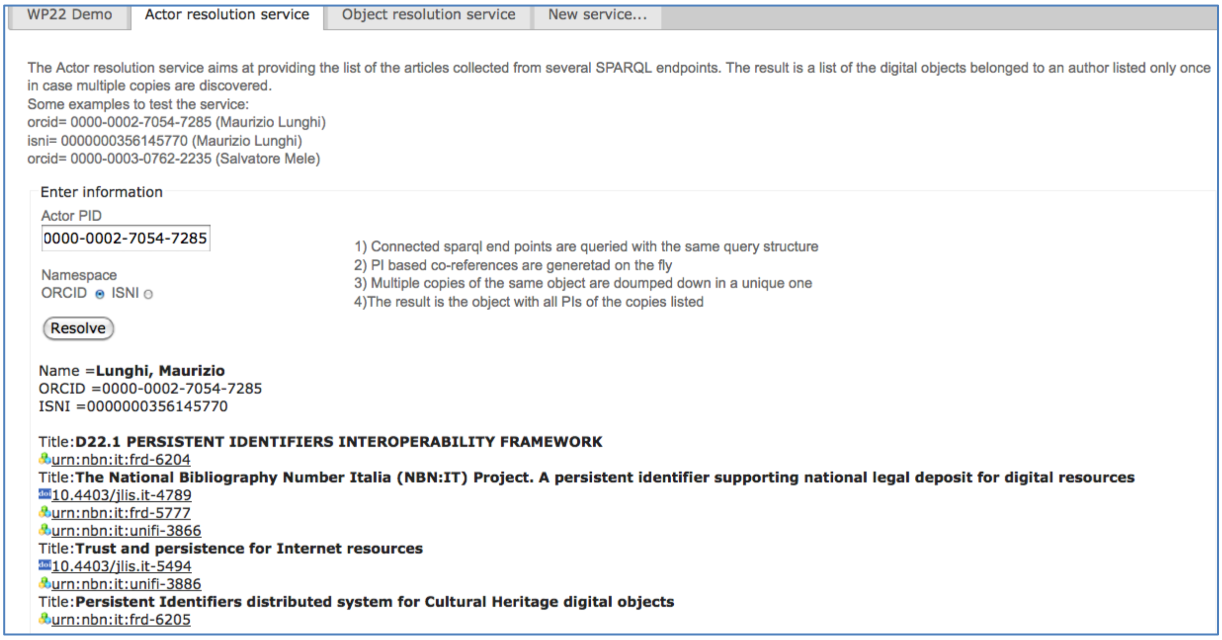
Digital Object PID: 10.4403/jlis.it-4789

Namespace
DOI ● NBN ○

1) Connected sparql end points are queried with the same query structure
2) PI based co-references are generetad on the fly
3) Multiple copies of the same object are doumped down in a unuque one
4)The result is the object with all PIs of the copies listed

- Title:**The National Bibliography Number Italia (NBN:IT) Project. A persistent identifier supporting national legal deposit for digital resources**
Description:The association of a persistent identifier to a digital resource is a best practice helping in solving a reliable information resources retrieving on the web, other than the Internet Error 404. At the present moment, different kind of technologies and standards such as DOI, Handle, and Cool URI are establishing. However, within scientific and cultural communities and in the semantic web, the main need is still focused on ensuring resources authenticity, provenance, and preservation in the log term period. For this reason, a third authoritative party such as the National Library is necessary, managing identifiers and taking in consideration technical as well as community and process requirements for preservation aims. The National Bibliography Number (NBN) originated in the context of the role and the responsibilities of national libraries in the field of universal bibliographic control, oriented to the identification of resources to be preserved in the long period. In Italy NBN:IT is going to be linked to the legal deposit, and acrivity traditionally up to the national libraries guaranteeing its trustability. The project, coordinated by the national libraries of Florence, Rome, and Venice, by the Foundation Rinascimento Digitale and by the Conference of Rectors of the Italian Universities, completes the current offer of Persistent Identifiers.
10.4403/jlis.it-4789
urn:nbn:it:frd-5777
urn:nbn:it:unifi-3866

Figure 16: A screenshot of the Object resolution service



The Actor resolution service aims at providing the list of the articles collected from several SPARQL endpoints. The result is a list of the digital objects belonged to an author listed only once in case multiple copies are discovered.
Some examples to test the service:
orcid= 0000-0002-7054-7285 (Maurizio Lunghi)
isni= 0000000356145770 (Maurizio Lunghi)
orcid= 0000-0003-0762-2235 (Salvatore Mele)

Enter information

Actor PID: 0000-0002-7054-7285

Namespace
ORCID ● ISNI ○

1) Connected sparql end points are queried with the same query structure
2) PI based co-references are generetad on the fly
3) Multiple copies of the same object are doumped down in a unique one
4)The result is the object with all PIs of the copies listed

Name =**Lunghi, Maurizio**
ORCID =0000-0002-7054-7285
ISNI =0000000356145770

Title:**D22.1 PERSISTENT IDENTIFIERS INTEROPERABILITY FRAMEWORK**
urn:nbn:it:frd-6204
Title:**The National Bibliography Number Italia (NBN:IT) Project. A persistent identifier supporting national legal deposit for digital resources**
10.4403/jlis.it-4789
urn:nbn:it:frd-5777
urn:nbn:it:unifi-3866
Title:**Trust and persistence for Internet resources**
10.4403/jlis.it-5494
urn:nbn:it:unifi-3886
Title:**Persistent Identifiers distributed system for Cultural Heritage digital objects**
urn:nbn:it:frd-6205

Figure 17: A screenshot of the Actor Resolution Service

3.6 SPARQL END-POINTS SET UP

Institutions manage their contents with a number of different archives and repository systems and to capture information and put it in Linked Data format, it is necessary to identify a technique for each of them. This chapter presents the IF demonstrator implementation strategy adopted according to the systems currently in use in the institutions.

The first task to be accomplished for Linked Data publication is the mapping. The mapping is a crucial activity where each institution should link their metadata fields with the target RDF-schema. The mechanism is based on a declarative mapping between the schemata of the database/in-house ontology and the target RDF-Schema. Such a mapping task and the publishing of the resulting RDF triples can be managed with several tools like:

- a local store with Virtuoso (which comes as package with Debian and Ubuntu).
- a hosting service, e.g. for Open Data offered by Science3.0, and Talis.
- an externals SPARQL service (like sparql.org) or more accurately used indirectly by using RDFaDev.
- a virtual RDF server like D2R Server for DBMS to RDF virtual mapping

Since D2R is considered a fast and cost-effective entry level for setting up, testing and providing a basic SPARQL service, to accomplish such a task, at FRD we have adopted the D2R Server to create Linked Data view of the databases. In fact, since we use a relational database for managing the digital objects, the strategy was to leave the information on the database and, with D2R to provide Linked Data views of it.

D2R Server uses a customizable D2RQ mapping to map database fields into a given RDF format, and allows the RDF triples to be browsed and searched. Requests from the Web are rewritten into SQL queries via the mapping. This on-the-fly translation allows publishing of RDF from large live databases and eliminates the need for replicating the data into a dedicated RDF triple store.

To connect D2R server with a DBMS like MySQL, it is necessary to express the connection triples in Turtle notation and stored them in the TTL mapping file. For each entities and properties defined in the ontology, a specific triple has to be written according to the DBMS schema. The examples provided below, are based on our database implementations and present the turtle based mapping triples written to link the columns of the database tables to the IF ontology.

In particular is reported how Digital Object and Actor classes are populated with the instances coming from our database.

```
#Class Digital Object
map:DigitalObject a d2rq:ClassMap;
  d2rq:dataStorage map:database;
  d2rq:uriPattern "apaif/digitalobject/@@records.itemID@@";
  d2rq:class apaif:DigitalObject;
  d2rq:classDefinitionLabel "DigitalObject";
  d2rq:containsDuplicates "true";
  .
#Class Actor
map:Actor a d2rq:ClassMap;
  d2rq:dataStorage map:database;
  d2rq:uriPattern "apaif/actor/@@authors.AuthorID@@";
  d2rq:class apaif:Actor;
  d2rq:classDefinitionLabel "Actor";
  d2rq:containsDuplicates "true";
  .
```

The fields within “@@” characters represent the name of the column of table in our database.

To populate the dc:title property of the class DigitalObject, for instance, the following mapping triples have been written.

```
map:title a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:DigitalObject;
  d2rq:property dc:title;
  d2rq:condition "records.field='dc:title'";
  d2rq:pattern "@@records.value@@";
  .
```

For each class and property in the ontology, a specific set of mapping triples has been written.

Once the mapping task is completed, the D2R server is lunched. At this point, a connection between the database and D2R Server is established and the SPARQL end-point integrated in the D2R server is active.

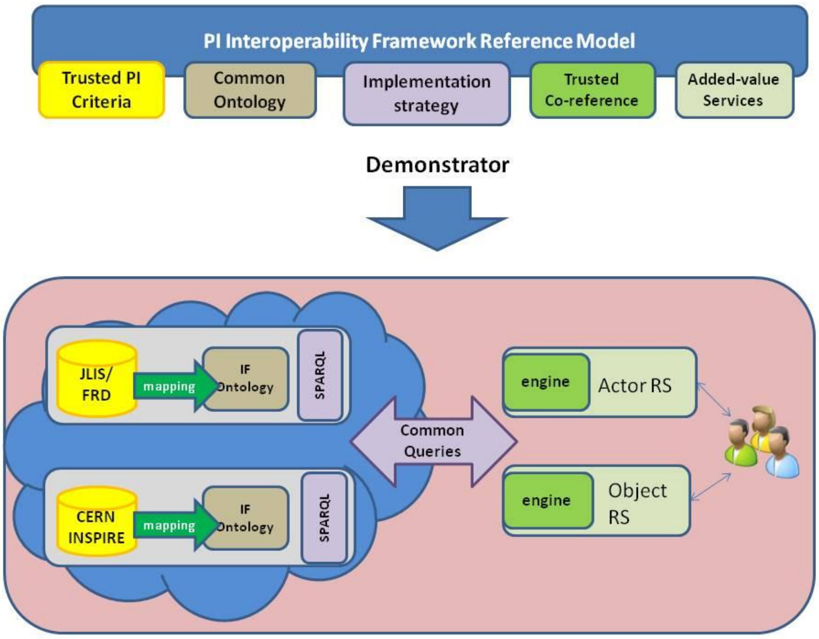


Figure 18: Demonstrator overview

Once the mapping task is completed, all institutions participating in the demonstrator expose the metadata in a shared and semantically reconciled format that is queryable through SPARQL end-point. The demonstrator is based on:

- 3 SPARQL end-points (1 CERN, 2 hosted on FRD servers)
- 7 content providers (DANS, JLIS, CERN,ORCID, NBN:IT, ISNI, NBN:DE)

In the current version of the demonstrator, only PI-do, namely DOI & NBN, and PI-ac, namely ORCID & ISNI, are implemented for a matter of practicality. As well as, only few contents populate the demonstrator, but these two limitations don't invalidate the IF model test results. The two services developed for the demo: Object Resolution Service and Actor Resolution Service, benefit of the IF common semantics and the SPARQL end-points availability by using the same semantic query for all SPARQL end –points and by avoiding further information reconciliation.

In this way, the processing (represented by the engine block in the Figure 18) is just focused on information retrieval, trust co-reference discovery and output packaging.

WP22 next steps

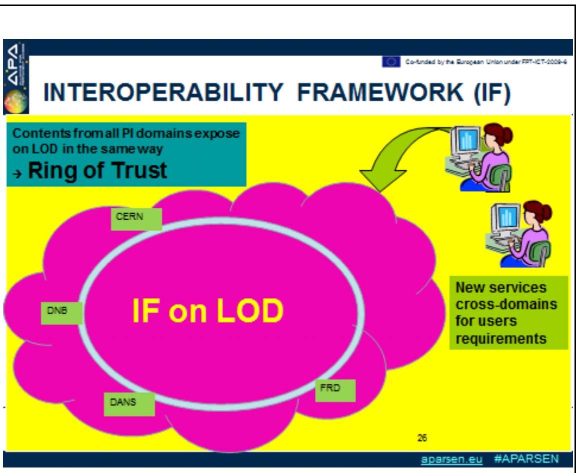
In the course of the last year of the project, the demonstrator will be used to evaluate user satisfaction about the potential benefits of the IF model and to refine the basic services.

Additional services will be implemented thanks to the collaboration with UNITN and OKKAM (a UNITN spin-off) by exploiting the ENS technology[20] for managing alternative PIs for different kinds of entities.

In conclusion we present our vision for future development.

The vision for the IF

The demonstrator development foresees to distribute contents on multiple nodes on LOD, likely some triple stores or SPARQL end points, and to populate the IF with contents from different PI domains for digital objects and for people. In future the framework can also be extended to physical objects like books with the ISBN domain or objects in museums through the system proposed by the CIDOC-CRM. The PI providers and CPs implementing IF will constitute a '**Ring of trust on LOD**' offering added value services to users both in terms of trust and in terms of common level of service and interface. On that framework some services across PI domains can be developed in line with the final users requirements.



Co-funded by the European Union under FP7-ICT-2009-6

INTEROPERABILITY FRAMEWORK (IF)

Contents from all PI domains expose on LOD in the same way
→ Ring of Trust

IF on LOD

New services cross-domains for users requirements

Figure 19: IF next step

Promising opportunities are envisaged from a potential cooperation with the ODIN project.

ODIN
ORCID and DataCite Interoperability Network

ODIN – *ORCID and DataCite Interoperability Network* - is a two-year project which started in September 2012, funded by the European Commission's 'Coordination and Support Action' under the FP7 programme.

Partners in ODIN are innovators in science, information science and the publishing industry: CERN, the British Library, ORCID, DataCite, Dryad, arXiv and the Australian National Data Service

ODIN will build on the ORCID and DataCite initiatives to uniquely identify scientists and data sets and connect this information across multiple services and infrastructures for scholarly communication. It will address some of the critical open questions in the area:

- Referencing a data object
- Tracking of use and re-use
- Links between a data object, subsets, articles, rights statements and every person involved in its life-cycle.

[20] http://www.okkam.org/

REFERENCES

DE22.1 Persistent Identifiers Interoperability Framework
http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D22_1-01-1_9.pdf

DE22.2 Demonstrator set up and definition of added value services: Part 1
http://aparsen.digitalpreservation.eu/pub/Main/ApanDeliverables/APARSEN-REP-D22_2-01-1_7-M16.pdf

CARROLL, J. M. (1995). Introduction: the scenario perspective on system development. In J. M. Carroll (Ed.) Scenario-based design: envisioning work and technology in system development (pp. 1-18). New York: John Wiley & Sons, Inc

SOA definition Barry &Associates, Inc. - http://www.service-architecture.com/webservices/articles/serviceoriented_architecture_soa_definition.html

Bell, Michael (2008). "Introduction to Service-Oriented Modelling". Service-Oriented Modelling: Service Analysis, Design, and Architecture. Wiley & Sons. ISBN 978-0-470-14111-3.

Emanuele Bellini, Cinzia Luddi, Chiara Cirinnà, Maurizio Lunghi, Achille Felicetti, Barbara Bazzanella, Paolo Bouquet: Interoperability Knowledge Base for Persistent Identifiers Interoperability Framework. IEEE SITIS 2012: 868-875

Emanuele Bellini, Chiara Cirinnà, Maurizio Lunghi, Barbara Bazzanella, Paolo Bouquet, David Giaretta and René van Horik (2012), "Interoperability Framework for Persistent Identifier systems" iPRES 2012, Toronto

Emanuele Bellini, Chiara Cirinnà, Maurizio Lunghi " Trust and Persistence for Internet Resources , Italian Journal of Library and Information Science. – Global Interoperability and Linked Data in Libraries workshop – Italian Journal of Library and Information Science - JLIS.it 2012

Bellini E, Cirinnà C. and Lunghi M. Persistent Identifiers for Cultural Heritage Digitalpreservationeurope (DPE) EU project – Briefing Paper
http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf

ANNEX I: Glossary for PI systems

Access system	is the mechanism that provides the ability to interact with a system, to retrieve relevant information (e.g., digital objects) and use this information
Archive (short for OAIS)	an organization that intends to preserve information for access and use by a designated community
Actor identifier	is a unique expression that makes it possible to disambiguate authors from each other. The use of these IDs has been recognized as a fundamental issue to establish the identity of authors and other contributors and reliably link them to their published works.
Centralized Naming authority	identifier management for a range of authorities is centralised if all authorities manage their identifiers through a common identifier management system, hosted on their behalf by a central party.
Citability	an entity is cited if its representation is communicated to an audience through some medium. The entity is citable if it can be cited. For example, citing the identifier (("Handle server 102.100.272", "XYZ"), "PILIN policy on citation") means coming up with an appropriate representation of the identifier (e.g. hdl:102.100.272/XYZ), and embedding that representation in a PDF(PILIN).
Content holder	who owns the rights of digital or physical contents that have been assigned a PI to, only in some cases content holders and content providers can be the same, e.g. the author of a paper is the holder and the library collecting and exposing it on the Web is the content provider. Both content holders or content providers can ask for a PI to be assigned to a resource.
Content provider	within a PI domain who makes accessible a resource, content providers ask for PI services both for actors and for objects. They request PI for themselves to be identifiable in a unique way and request PI for their contents to make them referable and usable. In most of the cases they also make their content accessible to all the other users.
Distributed Naming Authority	in a decentralized identifier management system, there is no single centralized authority that assigns and manages the naming service on behalf of all the parties. Instead each party, also called a peer, make a local autonomous management according to a minimum shared rules. Peers directly interact with each other and share information or provide service to other peers.
Digital object	an object composed of a set of bit sequences (OAIS). Pragmatically, it is a unit of information that can be identified, such as anything that might be stored in a digital repository. Examples of Digital Objects include documents, articles, books, images, web pages, applications, audio files, raw data, databases. A digital object is assumed here to belong to at least one digital repository.
Granularity	granularity refers to the level of detail at which PIs will need to be or may be assigned. In some situations, it may be necessary to cite a Web page which serves as access to a collection of Web files, or to cite a journal article, an item, or a chapter or a subset of a data file or perhaps a result of a database query. However, due to rights management, some finer details may be required. Each institution would need to evaluate whether a PI service provides the right level of granularity for their type of resources.
Identifier (ID)	it is an expression composed by one or more characters, digits or codes, that uniquely identifies an object. Identifiers can be local or global. Local identifiers

	uniquely identify entities in a given context or system (e.g. the employee IDs used by a company), whereas global identifiers identify entities across systems and contexts (e.g., ISBN).
Identifier scheme	is a scheme that defines the characteristics of an identifier, such as, for example, the syntax used to create the ID, the information and the kinds of metadata that can be associated to it, if the ID is resolvable, if it is language-dependent, how it is assigned and so on.
Identifier management system:	is a system that deals with identifying entities in a system by using identifiers. In the system IDs are used only as a way to make unambiguous reference to an entity and not as tokens to access to the system (this allows to distinguish ID management systems from authentication services described below).
Interoperability among PI systems	our concept of ‘interoperability’ is quite simple and is not used to indicate the ability of PI systems to interoperate between them in a direct way (DOI will not speak with NBN, it’s not required) but it is conceived in terms of a common way of access to data belonging to heterogeneous PI domains which are identified through different identification schemes. Our goal is to make accessible data from all the PI domains in the same format so that users can use them without worrying about different internal organization and policy.
Long Term Preservation	the act of maintaining information, independently understandable by a designated community, and with evidence supporting its authenticity, over the long term (OAIS).
Metadata	the term literally means “data about data”. Metadata provide additional information about a certain digital object, such as its author, creation data (time and date), Representation Information, Preservation Description Information (PDI), including possible access restrictions or the application used to create the file. XML is a standard to add metadata to documents and make them machine-readable.
Namespace	an abstract container providing context for the items it holds and allows disambiguation of items having the same name (residing in different namespaces). The namespace are registered by Internet Assigned Numbers Authority (IANA) and are defined by IETF-RFC where is identified also the naming authority. Examples is the URN namespace such as National Bibliography Number (RFC 3188-NBN) under the responsibilities of National Libraries.
Naming authority	independent authority that assigns names and guarantees their uniqueness and persistence. A naming resolution service corresponds to every naming authority and carries out the name resolution. In a Persistent Identifier distributed approach is foreseen that the responsibility of generation and resolution can be delegated to other institutions called sub-naming authorities who manage a portion of the name domain/space.
Opaque PI	a semantic PI is referred to the capability of extracting meaningfulness from the identifier. Examples are the mnemonic-based identifiers rather than those that contain a meaningless character sequence, although this has no relevance to machine processing.
Persistent	a component is persistent if it is managed and maintained for a defined timespan. Maintaining the component includes ensuring that its published content (such as its association data) is valid at all times. Normally when an identifier is called persistent, persistence of association is meant.
Persistent identifier (PI)	it is a maintainable identifier that allows to refer to and have reliable access to a resource or object over long periods. A PI has to be always resolvable through a resolution system.
PI system or	A system for generating and managing in long term some PIs assigned to some objects or an entity or other. The system is composed by some actors, namely, the

domain	content providers, the PI service providers and the user community. It uses a reliable technology implemented by some service providers in order to satisfy the requirements of a target user community. All the definitions and rules must be declared in a clear and public policy. Trust is a fundamental element of this infrastructure. Examples of the PI domain are the DOI community or the ARK community.
PI manager or service provider	Within a PI domain, depending on the internal architecture, some centres are devoted to generating PIs for content providers who will ask this service. Examples of the service providers are the Registration Agencies in different PI domains like Handle or NBN.
PI system for digital objects (PI-do)	A system for generating and managing in long term some PIs assigned to some digital objects. The system is composed by some actors, namely, the content providers, the PI service providers, the user community. It uses a reliable technology implemented by some service providers in order to satisfy the requirements of a target user community. All the definitions and rules must be declared in a clear and public policy. Trust is a fundamental element of this infrastructure. Examples of the service providers are the Registration Agencies in different PI domains like DOI or NBN.
PI system for actors (PI-ac)	A system for generating and managing in long term some PIs assigned to some actors physical or abstract. The system is composed by some actors, namely, the content providers, the PI service providers, the user community. It uses a reliable technology implemented by some service providers in order to satisfy the requirements of a target user community. All the definitions and rules must be declared in a clear and public policy. Trust is a fundamental element of this infrastructure. Examples of the service providers are the Registration Agencies in different PI domains like ORCID or ISNI.
PI system for physical objects (PI-po)	A system for generating and managing in long term some PIs assigned to some physical objects. The system is composed by some actors, namely, the content providers, the PI service providers, the user community. It uses a reliable technology implemented by some service providers in order to satisfy the requirements of a target user community. All the definitions and rules must be declared in a clear and public policy. Trust is a fundamental element of this infrastructure. Examples of the service providers are the Registration Agencies in different PI domains like ISBN or the CIDOC system for objects in museums.
Proprietary system	is a system which relies upon software and hardware which are licensed from a copyright holder.
Repository system	a system in which digital objects are stored for possible subsequent access, retrieval and management. Place where digital resources are held with or without a resource management system.
Registration Authority	Is the Authority that oversees and manages the identifier system
Registration Agency	Is the Agency that manages the registration process, which may be delegated further, to e.g. publishers
Resolution service (dereference):	an identifier is resolved by providing information on how to access the thing it identifies. This information is the resolution of the identifier: it is the output of the resolve action (PILIN) In other words it is the process in which an identifier is the input (a request) to a service to receive in return a specific output (resource, metadata, etc).
Semantic PI	a semantic PI is referred to the capability of extracting meaningfulness from the identifier. Examples are the mnemonic-based identifiers rather than those containing a meaningless character sequence, although this has no relevance to machine processing.

Trustworthy Digital Repository (TDR)	repository which has a current certification.(ISO 16919)
Versioning	A versioning of a **digital object** is an abstraction fixing the content but not the appearance of the digital object. Two instances belong to the same version if they have the same content; they belong to different version if they have different content, but are still seen to be underlying the same **thing**. Versions may include revisions, transformations, translations, and so forth. Expressions in the FRBR model are a type of version.
URI	A Uniform Resource Identifier is the generic set of all names/addresses that are short strings that refer to resources
URL	a Uniform Resource Locator is a URI that, in addition to identifying a resource, provides means of acting upon or obtaining a representation of the resource by describing its primary access mechanism or network "location"
URN	a Uniform Resource Name is a URI that uses the URN scheme, and does not imply availability of the identified resource. URNs are intended to serve as persistent, location-independent resource identifiers and are designed to make it easy to map other namespaces (that share the properties of URNs) into URN-space. Therefore, the URN syntax provides a means to encode character data in a form that can be sent in existing protocols, transcribed on most keyboards, etc. (IETF-RFC1737).
User community	Within a PI domain, the most important actor is the target user community meaning the group of users who decided to manage the contents and, as well as the possible use of those. The user community also defines the type of service about PI and the possible use or access. Therefore the user community is the origin of the system and its main user/supporter.