Interoperability Framework for Persistent Identifiers systems

Barbara Bazzanella
DISI, University of Trento
via Sommarive, 5
38123 Povo Trento, Italy
+39 0461 28 3383
barbara.bazzanella@unitn.it

Emanuele Bellini
FRD, Fondazione Rinascimento Digitale
Via Bufalini 6
50100 Florence, Italy
+39 0555384925
bellini@rinascimento-digitale.it

Paolo Bouquet
DISI, University of Trento
via Sommarive, 5
38123 Povo Trento, Italy
+39 0461 28 3383
bouquet@disi.unitn.it

Chiara Cirinnà
FRD, Fondazione Rinascimento Digitale
Via Bufalini 6
50100 Florence Italy
+39 0555384925
cirinna@rinascimento-digitale.it

Maurizio Lunghi
FRD, Fondazione Rinascimento Digitale
Via Bufalini 6
50100 Florence Italy
+39 0555384925
lunghi@rinascimento-digitale.it

David Giaretta
APA, Alliance for Permanent Access
2 High Street
Yetminster
Dorset DT9 6LF, UK
director@alliancepermanentaccess.org

René van Horik
DANS – Data Archiving and Networked Services
Anna van Saksenlaan 10
2593 HT The Hague, the Netherlands
+31 70 3446484
rene.van.horik@dans.knaw.nl

ABSTRACT

In this paper, we propose an Interoperability Framework (IF) for Persistent Identifiers (PI) systems that addresses functions, roles and responsibilities needed to make heterogeneous PI systems interoperable. The fundamental steps, which provided the main inputs for the design of the model have been: 1) a survey on the use of PI among different stakeholder communities and 2) the definition of interoperability use cases and requirements. The IF is presented as a solution addressing the PI interoperability issues, which have been identified in the survey and have been translated into concrete use cases to serve as requirements for designing the model. Conclusions and intended future work close the paper.

Keywords

Persistent Identifiers (PI), PI Domain (PID), Digital Preservation (DP), Interoperability Framework (IF), reference model, trust.

1. INTRODUCTION

The main goal of this work is to present an Interoperability Framework (IF) for Persistent Identifiers (PI) systems able to overcome the current limits in the use of PI technologies in the actual isolated application domains. When the IF is implemented, the current fragmentation will be reduced, with many additional benefits for the users, provided by some new cross-domain and cross-technology services.

The research work has been carried out through a scientific study and a desk research analysis on the state-of-art of technologies and projects. A questionnaire and some interviews helped to understand the user requirements. The survey investigated current uses and approaches by different user communities of identification systems for digital objects, people, institutions, and few examples of projects trying to implement interoperability among systems. This survey confirmed the absolute lack of such interoperability and showed that the current systems usually work isolated or in competition, hindering the use of PI across systems and creating complications for the final users. This investigation has been crucial also in order to understand the potential interest by the user communities and the most relevant use cases for our scenario and objectives.

Global and standardized identification systems for people and institutions are not very common. In the digital arena many different systems or methods for objects identification are in use: some of them are valid only locally or for specific types of content, others are used for the identification of physical objects, some are not freely resolvable, others are dynamic and can change over time, and only some of them are really persistent over time and can be considered part of a Digital Preservation (DP) policy. A key concept in this work is the Persistent Identifiers Domain (PID) meaning the system of policy and technology implemented by a user community interested in preserving/using digital contents and managing a PI system for them.

To overcome this fragmented situation, in the framework of the APARSEN Network of Excellence, a reference model has been developed that can be adopted and implemented by any current PI application domain to expose data in a format agreed in the IF, common to all the systems. In this work we ignore all the identification systems not in line with digital preservation criteria and, moreover, we define a benchmark, which specifies the criteria requested to the PI systems to be eligible for our reference model.

2. THE RESEARCH CONTEXT

In order to understand the present work, it is important to contextualize the research within the APARSEN community. Alliance for Permanent Access to the Records of Science in Europe Network (APARSEN), see: http://www.aparsen.eu is a Network of Excellence (NoE) co-funded by the European Commission at the call 6 of the FP7, started on the first of January 2011, a consortium of experts on digital preservation with 34 partners in Europe. A NoE is a very specific instrument with the main goal to fight fragmentation of initiatives and research in Europe, a NoE must be thematic and cover a specific topic in line with the FP7 objectives. In Europe even on specific area, like digital preservation, we have a dramatic fragmentation at any level, countries, research centers, professional associations, projects and this causes a waste of resource, investments, impact and competitiveness of our institutions and companies.

APARSEN large consortium brings together a diverse set of practitioner organizations and researchers in order to bring coherence, cohesion and continuity on long-term accessibility and usability of digital information and data researches. The project aims to exploit also this diversity of the partners by building a Virtual Centre of Digital Preservation Excellence. The objective of this project may be simply stated, namely to look across the excellent work in digital preservation which is been carried out in Europe and to try to bring it together under a common vision. The success of the project will be seen in the subsequent coherence and general direction of practices and researches in digital preservation, with an agreed way of evaluating it and the existence of an internationally recognized Virtual Centre of Excellence.

3. PI SURVEY

The main goal of Work Package 22 (WP22) of the APARSEN project is to propose an Interoperability Framework (IF) among different Persistent Identifiers (PI) systems in line with the user communities' requirements. The first year of the WP22 includes two tasks: Task 2210: Survey and benchmarking led by the University of Trento and Task 2220: PI evaluation and integration into an Interoperability Framework and Reference Model led by FRD. The outcome of the Task 2210 and Task 2220 are included in the public deliverable (DE22.1) available at http://www.aparsen.eu/index.php/aparsen/aparsen-deliverables/

In order to gain a clearer understanding of the current state of the use of PI systems by different user communities, a questionnaire has been disseminated to the partners belonging to the APARSEN network of excellence and beyond this community (see complete results in Annex I of the DE22_1). The intent of this questionnaire was to explore existing practices**,** requirements and resources for persistent identification as well as to identify real needs, gaps and challenges related to the use of PI systems. The questionnaire was spread among several mailing lists such as those hosted by JISC, DPC, APA, DANS, project communities such as Nestor, CASPAR, PLANETS, DPE, PersID, DataCite, etc. and association communities such as AIB, LIBER, CRUI, etc.

Desk research was conducted to identify relevant features, which characterize the main current PI systems and may have an impact on interoperability. This analysis was also useful to understand weaknesses and strengths of each PI system in relation to the user expectations about digital preservation. The results of the desk research activity and the correspondent feature analysis are reported in the Annex II on the DE22_1.

Several APARSEN partners are involved directly in PI projects or services such as STM (DOI), CERN (ORCID), DNB (NBN:DE), DANS (NBN:NL), FRD (NBN:IT), where DOI and NBN are PI systems for digital objects and ORCID is an initiative for PI for authors, or are users of these services, since they manage institutional repositories, usually universities and research institutions, or scientific datasets. Other key players such as DataCite, SURF Foundation, National Library of Sweden, National Library of Australia, National Library of Finland, CrossRef, IETF NBN Working Group have been interviewed during workshops and meetings such as the meeting organized by Knowledge Exchange on "Exploring interoperability of Persistent Object Identifier systems" which produced an important contribution to the identifier interoperability issue through the Den Hague Manifesto http://www.knowledge-exchange.info/Default.aspx?ID=440
The point of view and the suggestions of these stakeholders have been taken into account throughout our work.

3.1 Survey structure and Method

In the questionnaire we considered three kinds of persistent identifier systems: 1) PI for digital objects; 2) PI for authors and creators and 3) PI for organizations. The survey was composed of five sections: 1) PI for digital objects; 2) PI for authors/information creators; 3) PI for organizations; 4) Criteria for the adoption of a PI system for digital objects; 5) Digital preservation strategies and practices. In the first three sections we focused on identification practices, limits and requirements for PI for digital objects, authors and institutions. The fourth section contains the criteria adopted by the users for the adoption of PI systems for digital objects, focusing on aspects related to technology, organization of the service, scope, naming rules and expected services. Finally, we addressed issues concerning digital preservation strategies and practices with a special focus on the use of written guidelines, time span, funding and financial sustainability.

3.2 Results

The questionnaire received 103 full responses from participants of three main represented organizations: libraries (47%), universities (27%) and archives (22%)

mainly from academic/research, government and public sectors, 85% of participants were from European countries.

We report here only the results which are more relevant for the design of the IF. The complete analysis of the results is available in the Annex I of the DE22.1.

1) A first analysis was conducted to investigate the **current use of PI systems for digital objects, authors and institutions** among different stakeholder communities. The results show that the DOI (32%), Handle System (28%) and URN (25%) are the most popular PI systems for digital objects even though local identifier systems are still widely adopted (24%). In particular, referring to the stakeholder communities, DOI is the most common system used by universities, libraries, archives and publishers, Handle is mainly adopted by libraries and archives and URN is almost exclusively adopted by libraries. Other systems, like PURL and ARK, are used by a minority of participants (<10%). This scenario shows that PI systems are becoming increasingly oriented towards a specific community, indicating that an IF that allows a cross-community and cross-system communication is clearly needed.

From this result we gained a first indication on which systems have to be considered to be included into the IF. The survey results show also that PI systems for identifying authors are scarcely adopted (52% of participants claimed that they do not use PI for authors). In any case, the IF has to assume the existence of Author ID systems, but avoiding a focus on specific implementations.

A very similar result to the previous one has been found for the persistent identification of organizations. The answers of the participants indicate that there are no specific PI initiatives for organizations. In fact, the majority of the respondents (39%) reported that no system is adopted to identify their organizations. Globally, a fragmentary picture emerges where PI systems adopted for digital objects are slowly adopted for institution. Following the same approach held for author PI systems, the IF assumes the existence of PI systems for organizations avoiding a focus on specific implementations.

2) About the **types of digital objects**, the results of the questionnaire show that textual documents (reported by 98% of participants) and images (selected by 86% of participants) are the most commonly held digital objects. These results suggest that the IF has to address these two types of objects first.

Two other relevant issues deal with **granularity and versioning**. Concerning granularity the survey results show that a finer capability of a PI system to identify and access parts of digital objects is required. Concerning versioning the survey results indicate also that the most common approach for content versioning is linking a new version to the original version through metadata, followed by the practise of considering the new version as an autonomous object. The use of naming rules is less common among the participants.

Thus the IF should include those PI systems that support the scalability, granularity and versioning issues working mainly at metadata level.

3) One of the objectives of the survey was to investigate the **limits** experienced in using PI systems for digital objects. Some expected results have been reported, such as “locally defined” and “no standard associated” referred to internal identifiers solutions. It is worth mentioning that one of the limits reported regarding DOI and URN is “low adoption” even though these systems are the most widely used systems within our user sample. Finally, “ongoing costs” is one of the most frequently mentioned limits for DOI system.

In general, users perceive a certain level of immaturity for author identification systems which concerns services, trust and authority.

If we compare the obstacles that the respondents reported about the use of PI systems for authors with those about the use of PI systems for organizations, we can notice that the two most frequently selected obstacles are the same: the lack of awareness and the fact that the use of PI systems is not considered a key issue for the organization. This result confirms that one of the main actions of intervention to promote agreement across the different stakeholder communities about the adoption of PI systems should start from increasing the level of awareness about the available systems and their potential positive effects.

4) About **user requirements**, we investigated four domains: technology, organization of the service, scope and naming rules. In terms of technology, our results indicate that users prefer to adopt a system that represents a de facto standard (53%), widely adopted (56%) and based on an open source infrastructure (88%). This was an interesting input in defining the criteria to evaluate as eligible for the IF the PI system (Trusted PI). In terms of the organization of the service, distributed naming authority (48%) and supported by an institution with a mandate (55%) were the preferred options. In terms of scope, the respondents reported to prefer systems open to any digital objects (81%) and cross-community (76%). Finally, concerning naming rules opaque identifiers (55%) (supporting deep granularity (57%)) are preferred above semantic identifiers supporting low-level granularity. No relevant differences were found between the stakeholder groups in the requirements for adopting a PI system for digital objects.

5) The last relevant aspect for the design of the IF deals with **services**. Citability is the most important service associated to the use of PI, followed by services, which support resolution (i.e. global resolution services,

resolution to the resource or to metadata). More than half of the participants reported services for digital object certification among the required services. According to the stakeholders analysis it seems that if citability is a desired service for all the stakeholder groups in long term vision, aspects related to the resolution mechanisms are more relevant for libraries, archives and publishers, while aspects related to certification (and metrics) are more important for universities and research organizations.

Moreover, against our expectations, the PI basic services are those most required. The so-called "advanced services" that were considered most important for the IF received less votes[1]. According to this result, the framework design took into account also the objective to empower the basic PI services in addition to set up the conditions for developing new advanced services. This result was crucial in the distinction between different levels of service within the IF infrastructure.

4. USE CASES

Some user scenarios have been defined to introduce and concretize the interoperability concepts and requirements, by providing a number of use cases for IF following the Scenario Based Design technique [6]. We asked the partners to provide one or more scenarios from their experiences about PI use in a long term vision. Since the APARSEN partners are from different domains, the aim was to cover a wide variety of requirements for different stakeholders communities. We have collected 13 scenarios divided in three groups: 1) Scenarios on Citability and Metrics services, 2) Scenarios on Global Resolution Services (GRS) and 3) Scenarios on Digital Object Certification.

These scenarios have been translated into more simple use cases, a schematic framework useful for identifying entities, their relations, functionalities and so forth. The results of this phase have been used as input for the modeling phase.

5. THE PI INTEROPERABILITY FRAMEWORK (IF)

5.1 PI interoperability: related initiatives

Recently, several initiatives and projects have started to address the problem of PI interoperability and solutions have been proposed in different contexts facing some issues at identifier or metadata levels. A first distinction can be made between national and international initiatives. Some initiatives have been emerged within a national context (e.g PILIN[2] in Australia and RIDIR [3]in United Kingdom) and some of these started as a funded project on a broader geographical level (e.g. PersID[4]). Other initiatives show their presence at an international level (such as ORCID[5]) and aim at introducing global standards for identification, creating a consortium of participating organizations. We can also distinguish between initiatives limited to a specific discipline (e.g. for linguistic resources) or more generic initiatives dealing with a broader range of resources (e.g. OKKAM[6]). Some projects focus exclusively on the problem of PI interoperability for digital objects (e.g. PILIN), while other initiatives address the interoperability issue for author identifiers (e.g. ORCID). The diffusion of a given initiative can also be determined by the way in which the identifiers are assigned by the underlying ID management systems. Some governmental initiatives limit the assignment to people, that embark on an academic career, while other systems allow the registration of any kind of entity (e.g. OKKAM).

5.2 IF definition

Interoperability is an essential feature for federated information architectures which operate in heterogeneous settings also over time. However, the use of the concept is very heterogeneous: interoperability is conceived in an object-related or in a functional perspective, from a user's or an institutional perspective, in terms of multilingualism or of technical means and protocols. Moreover, interoperability is conceived at different levels of abstraction: from the bitstream level up to the semantic interoperability level [1] [2].

In this paper we describe a conceptual framework addressing the identifier interoperability issues, which have been identified in the survey phases and have been translated into concrete scenarios and use cases to serve as requirements for designing the reference model. The IF describes the entities of our domain, their relations and dependencies, the main functionalities and a minimal set of concepts in order to enable the development of specific implementations (i.e. interoperability services).

When the contents from different PIDs (which are currently not interoperable and are completely isolated) are visible through a common interface provided by the IF, users can access and use any content or relation available in the scenario. In particular, we can create any type of service accessing all the contents across the domains and using them even if they are from different PIDs, overcoming in this way a relevant limit in the current situation. The survey on current practices of PI and the

[1] Although the relatively small size of the survey is a concern, there are practical advantages in starting with the basic services.

[2] PIs Linking Infrastructure (PILIN) project - http://www.pilin.net.au/

[3] Resourcing Identifier Interoperability for Repositories (RIDIR) project http://www.jisc.ac.uk/whatwedo/programmes/reppres/ridir.aspx

[4] PersID project – http://www.persid.org/

[5] ORCID (Open Researcher and Contributor ID) www.orcid.org

[6] OKKAM project http://www.okkam.org/

description of the use cases have been crucial in order to understand the user potential interest and access modalities or specific required functionalities.

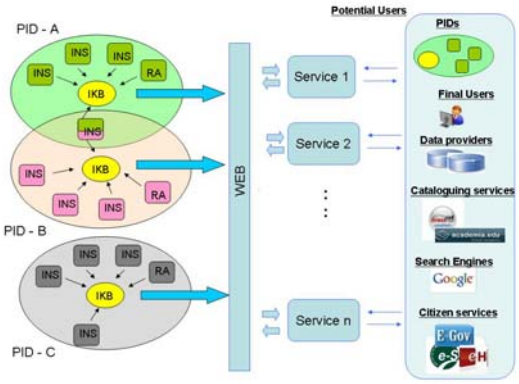

Figure 1 - Interoperability Framework Architecture

5.3 Main assumptions

The IF definition starts from the following main assumptions:

a) In the IF we consider only entities identified by at least one PI.
b) Only PIDs that meet criteria of Trustworthiness are included in the IF.
c) We delegate the responsibility to define relations among the identified entities to the Trusted PIDs.
d) We don't address the digital preservation (DP) issues directly but the DP strategy is demanded from the Trusted PIDs. However the IF allows spreading the preservation risk.

According to the main assumptions stated above, only trusted PIDs can join the framework and populate the scenario with their entities. It is important to notice, for the purposes of the present work, that the user community board managing the PID is responsible for guaranteeing suitable policies for any aspect of the DP plan underpinning that system, like for example, the content selection/granularity criteria (included the FRBR[7] levels), the Trusted Digital Repositories policies and certification, the trustworthiness of the PI management, and so on.

Moreover, within each PID there can be different approaches and architectures to share roles and responsibilities among different components of the system, like the Registration Authority (RA), the Certification Authority (CA), the domain resolver, the digital repository curator and content holders, the DP manager, and so on. The user community is free to choose the best solution and we trust them for the correctness of this choice.

5.4 The reference model

The key actors in the IF are the PI Domains (PIDs) that include in our definition:

1) The Registration Agencies (RAs), which manage the allocation and registration of PI according to the trust definition and provide the necessary infrastructure to allow the registrants to declare and maintain the PI-entity relations.
   We limit to only 3 types of PIDs based on the three different types of identified entities: a) PID for digital objects, b) PID for authors and c) PID for institutions
2) The content providers (INS in Figure 1 and 2) that are the institutions responsible for storing, managing and preserving the access to digital contents through the use of PI.
3) The resolver is a service able to provide information on the object, its current location and how to get it.

The framework provides a shared conceptual infrastructure to represent the identified entities and their relations within what we call an Interoperability Knowledge Base (IKB), assuming this declared information as guaranteed by trusted PIDs. These relations must be provided by the PIDs when they bring an entity into the interoperability knowledge base. In particular, some trusted PIDs will populate the IKB with their entities presenting these contents following an API so providing specific info requested by the IF. For any digital object the PID, in addition to some descriptive metadata, should declare existing PI (e.g., DOI, NBN), any relation with other objects within the domain and any PI for persons or institutions known by the PID.

In this way, the IKB defines the fundamental relations between the entities in play in the domain (e.g. between objects and PI), creating a layer of accessible knowledge on which interoperability services can be built thanks to the explicit representation of these relations (see Figure 2). Indeed, the knowledge generated independently by the trusted PIDs using the framework, will be exposed on the Web with a common semantics and interface enabling user to access to all the domains and using all the contents even if they are from different PIDs. **Figure 2** shows also that institutions that adopt more then one PI system for their resources, for instance DOI and NBN, contribute to the IKB of the DOI PID and NBN PID with the same relation statements. Thus, IKBs present some overlapping (in Figure 2 is represented by overlapping area between PID-A and PID-B) that can be exploited as a bridge to walk across PIDs and enabling new services to discover new relationships and make inferences on digital resources.

[7] IFLA- FRBR http://www.ifla.org/publications/functional-requirements-for-bibliographic-records

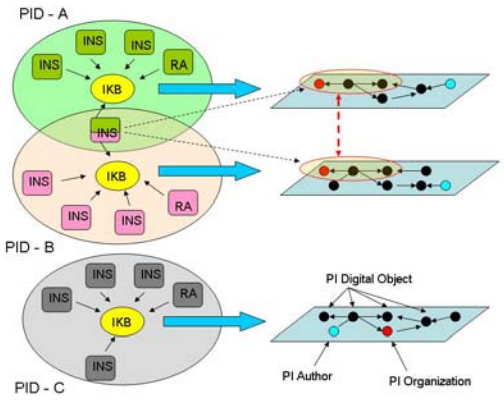

Figure 2 - Significant relations established through the IF across the PID boundaries.

5.5 IF main concepts

Resource: A resource is one of the most primitive concepts in the IF reference model and covers any entity that can be identified by at least one PI. Entities, which are not assigned to a PI, are not eligible for the IF. A resource is a representation of a physical or an abstract entity. Since the concept of resource can be very different in different PIDs, we propose a very general definition, which encompasses the diverse range of digital resources, including resources such as objects, annotations, and metadata. We consider three main kinds of resources in the framework: a) Digital Objects, b) Authors, c) Institutions.

Other kinds of resources can be included in the future with the development of PI systems dealing with other types of entities, such as events, locations and so on.

Digital Object: A digital object is any kind of digital resource, which is identified by at least one PI assigned by a trusted PID. We don't provide a more specific definition because we rely on the definition provided by the trusted PID which has assigned the PI to the resource. A digital object with no PI is not eligible for the IF.

Author: An author is a physical entity, which is the creator of a digital object and is identified by at least one PI assigned by a trusted PID. Whereas digital objects are digital in nature, authors are physical entities which are represented through descriptions (i.e. profiles) in the digital world. Therefore, while a PI for a digital object can point directly to the object, a PI for an author does not point to the author but always to a description of him/her. Moreover the resource, which describes an author, is expected to change as the referent inherently changes across time. Therefore, "the sameness" property of a PI for an author means referring to the "same physical entity" (i.e. the same author and not the same unchanged digital resource), while that of a PI for a digital object means referring to the "same digital entity" (i.e. the same digital resource, in some cases migrated or not, it depends by the PID policy).

Institution: An institution is a physical entity, which affiliates authors and other human agents and is identified by at least one PI assigned by a trusted PID for institutions.

Persistent Identifiers: a PI is a character string used to uniquely identify a resource within a PID regardless of where the resource is located. In the framework we distinguish between 3 kinds of PI.

PID: a PI Domain is a system of users and service providers, which manages the assignment of PI for any type of relevant entities (e.g. digital objects, authors, institutions). Typically, these types of systems are different for different communities and specific for types of objects. PIDs must be trustable in a very long-term vision. We trust PIDs for the implementation of adequate DP rules and strategies.

Policy: the concept represents the set of conditions, rules, restrictions, terms and regulations governing the entire life cycle of a digital resource and its management within a trusted system. This domain is very broad and dynamic by nature. The concept of policy captures the minimal relationships connecting it to the other relevant entities in the framework. The model is extensible and other subclasses of policies could be easily added in future

Resolver: A resolver is a system that provides the link between a PI and information about the object and its current location on Internet, and if available relations with other entities.

User/Actor: An actor is an entity that is external to the interoperability system and interacts with it and uses the related services. Both humans and machine can be users.

5.6 PI trust criteria

In order to design a reliable IF among PI systems, we have to define the criteria that should be met by a PI system. A PI framework has to be reliable to enable the development of advanced services. Thus, only those PIDs that match our criteria for trust will be taken into account as potential component of the framework.

In order to define the trusted PIDs we introduced a small set of criteria distinguishing between mandatory (M) and optional (O) criteria. The 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, "PI for Cultural Heritage DPE briefing paper" [3], NESTOR reports on trustworthiness of PI systems [4], A Policy Checklist for Enabling Persistence of Identifiers [5], the results of the ERPANET [8] and DCC [9]workshops.

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

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

1. **Having at least one Registration Agency (RA).**

Within a PI domain it is necessary that a RA is established to assign and maintain the association PI- digital resource. This criterion is considered mandatory in the IF trust assessment.

2. **Having one Resolver accessible on the Internet.**

To meet this criterion, a resolver able to resolve a PI has to be accessible on the Web. This criterion includes also the capability of a PI to be resolved to an entity represented by a 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. We consider this criterion mandatory in our framework.

3. **Uniqueness of the assigned PI within the PID**.

The RA has to guarantee that a PI is univocally assigned to a digital resource within the PI domain. In fact, since a PI is essentially a string, the uniqueness can be guaranteed only within a domain of reference served by a defined RA. This criterion is considered mandatory in our framework.

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

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) *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.

c) Lack of *sustainability*: indicates that a RA is not able to guarantee its commitment to maintain a PI as far as the identified resource exists. Managing identifiers in a sustainable way is another requisite for a trusted PID.

The point a) and b) can be addressed at a functional level of the PI service but they depend by the PID policies; point c) is related to the sustainability of the PI service and the PID business model. This criterion is considered mandatory.

5. **User communities, which implement the PID should implement policies for digital preservation (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 PI and their association with the identified digital resources (even if they are moved), on the other side, PI 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[10] and implement digital preservation strategies for the resources identified by a PI. This criterion does not require an unlimited guarantee from an organisation but a hand-over procedure should be in place, 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.**

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* 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 "equivalent manifestations" of the same object within the DOI domain (see CrossRef guidelines[11]), while they would receive two different identifiers in the NBN domain. According to this, the resolution within a PI domain is reliable if 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.

7. **Uncoupling the PI from the resolver.**

This criterion is crucial and refers 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 PI which contain the original URL will become invalid, in case the resolution service is moved to another domain. This criterion is considered mandatory in the proposed IF.

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

This criterion identifies the possibility to specify the linkage between resources within the PIDs through explicit relations between their identifiers. For example, a PID can make explicit the part-of relation between resources 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

---

[10] 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

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

part. This criterion is considered optional in our framework, but it represents an added value that can speed up the implementation of interoperability services.
We are aware that there are other features and criteria which can be considered in a Trusted PI definition. A critical example is scalability. A PI system that aims to identify an increasing number of objects on Internet (i.e. a global distributed system) must also handle scalability to be considered Trusted. In fact, scalability is one of the basic requirements for the long-term sustainability of every PI service. The main reason why we have not included the scalability as a criterion is due to the variability of the possible technical implementations of a system, and the difficulties in obtaining sufficient information about the technical implementation for making an accurate assessment. The difficulties of obtaining definitive results on such a criterion represent an ongoing concern that has been taken into account in the present work.

6. CONCLUSIONS

In the 2nd year of the APARSEN project the WP22 team will implement a validation mechanism in order to evaluate the Interoperability Framework for PI by around 30 experts, part of them external to the APARSEN consortium. So an action plan to set up a demonstrator for WP22 IF and related services, is under preparation with some external possible synergies with other projects like SCIDIP-ES[12] or other initiatives like ORCID and DOI or NBN large communities. In that demonstrator, some basic services will be tested and refined in order to implement the user requirements collected during the former work in the WP22 with the questionnaire and the use cases definition.
The validation of the model through a user group with experts, including ones external to APARSEN, will be a key strategy to reach consensus and make the model suitable for all the user communities' requirements. Thanks to this consensus building strategy, other user communities beyond the APARSEN consortium will be invited to join the framework and make their content public on the demonstrator, because it is very important to have data from different PIDs and for objects, people and bodies for the potential application spectrum of the user services. By the end of the 2nd year a first prototype with some cross-domains basic services will be set up and become available for the further development of the IF.

7. References

1] Stefan Gradmann, INTEROPERABILITY. A key concept for large scale, persistent digital libraries. Digitalpreservationeurope (DPE) project - Briefing Paper - September 2008 http://www.digitalpreservationeurope.eu/publications/briefs/interoperability.pdf

[2] Norman Paskin - Interoperability Identifiers - Briefing Paper - Digitalpreservationeurope.eu

http://www.digitalpreservationeurope.eu/publications/briefs/identifier-interoperability.pdf

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

[4] Catalogue of criteria for assessing the trustworthiness of PI systems http://files.d-nb.de/nestor/materialien/nestor_mat_13_en.pdf

[5] N. Nicholas, N. Ward and K. Blinco A Policy Checklist for Enabling Persistence of identifiers; D-Lib Magazine Jan/Feb 2009 http://www.dlib.org/dlib/january09/nicholas/01nicholas.html

[6] 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

[12] http://www.scidip-es.eu