persid

PersID
III.a – Current State and State of the Art
&
III.b – User Requirements

Overview and studies on
persistent identifier
infrastructure
commissioned by
Knowledge Exchange

and

Prototype development of
Meta Resolver Solution
commissioned by
SURFfoundation

SURF FOUNDATION

KE Knowledge Exchange

About this publication

PersID – III.a & III.b
Current State and State of the Art
User Requirements

A Knowledge Exchange and SURF initiative: Studies on Persistent Identifier Infrastructure and development of a URN-NBN based Global Resolution Service

SURFfoundation
PO Box 2290
NL-3500 GG Utrecht
Netherlands
T + 31 30 234 66 00
F + 31 30 233 29 60
info@surf.nl
www.surf.nl

Knowledge Exchange
Danish Agency for Libraries and Media
H. C. Andersens Boulevard 2
DK-1553 Copenhagen V
Denmark
T +45 3373 3315
office@knowledge-exchange.info
www.knowledge-exchange.info

Authors

Laurents Sesink	- DANS
Maarten Hoogerwerf	- DANS
Stina Degerstedt	- Swedish National Library
Adrian Price	- DEFF
Christa Schöning-Walter	- German National Library
Maurice Vanderfeesten	- SURF

Editor

Bas Cordewener	- SURFfoundation

Knowledge Exchange is a co-operative effort that supports the use and development of Information and Communications Technologies (ICT) infrastructure for higher education and research.

SURF is the collaborative organisation for higher education institutions and research institutes aimed at breakthrough innovations in ICT (www.surf.nl/en)

This publication is online available through www.knowledge-exchange.info

April 2011

urn:nbn:nl:ui:13-9g4-i1s

Contents

1 III.a – Current State and State of the Art

This chapter describes a framework that allows investigating the current state of the different persistent identifier solutions. It takes elementary functions as a basis and describes how policy, workflow and implementation are related to them. Preservation will be seen as one (or more) of these functions.

1.1 Framework

1.1.1 Functions

This paragraph defines general functions within the life cycle of persistent identifiers: from becoming a registrar to resolving an identifier and removing an identified object. The functions are abstract enough to be valid for every resolver.

#	Scenario	Description
01	Become a registrar	Organizations need to be authorized to create and register valid persistent identifiers.
02	Create unique and durable identifier	Organizations need to ensure that they generate unique and durable identifiers for objects.
03	Assigning identifiers	Organizations need to assign the identifier to an object by assigning the location of a resource.
04	Registering identifiers	Identifiers and their locations need to be registered in order to be valid and/or functional.
05	Updating locations	The location registered with an identifier needs to be updated when the identified object moves.
06	Updating identified objects	How updates on identified objects are reflected in their identifiers.
07	Transferring responsibility	Ownership and/or responsibility can be transferred to another organization. This can be a complete collection, parts of a collection or individual objects.
08	Removing identified objects	Identified objects can no longer be maintained because of financial or legal issues.
09	Fetching identified objects	Clients want to retrieve the identified objects to view / use them.
10	Fetching metadata	Clients want to retrieve information about the identified object.
11	Preserving contents	The data is stored in a safe place, its state is monitored and the contents are migrated when necessary

1.1.2 Policy

Policy is an essential component of persistent identifiers because it defines how functions should behave in order to ensure a trusted identity and retrieval. This trust depends on the maintenance of the identifiers. It is essential that every identifier that gets registered is maintained by a trusted organization or can fall back on one. Policy defines how these organizations should deal with their identifiers and their identified objects. It requires a trusted organization to commit to such a policy.

1.1.3 Workflow

The workflow describes how the different functions are dealt with. Workflow is to support the given policy: It determines under what conditions certain operations can be executed and what the outcome of these should be.

1.1.4 Implementation

Finally, the workflow can be implemented using procedures, responsibilities and technology: Who does what, how is it done and what systems or interfaces are used to achieve that?

1.2 Current situation within PersID

This chapter describes the current situation on persistent identifiers over the different participating countries. It follows the functions identified above.

Most organizations have a policy, but this is often not formalized and it is often outdated and in need of revision. The demand for persistent identifiers by other organizations (than the National Libraries themselves) to identify more heterogeneous content is growing and there is an increasing amount of applications and infrastructures that require persistent identifiers for access to the resources. This demands for explicit policy about who is allowed to assign identifiers, how to assign them and to what kind of resources they should be assigned.

1.2.1 Becoming a registrar

All participants allow external organizations to register identifiers. This is usually initiated by a request via email of phone. Continuing steps are: filling in an application, validation of this application, explanation of the procedures, signing an agreement and the assignment of a namespace. The differences are in the validation process: what type of organizations can apply to become a registrar and under what conditions. Usually the organizations must be official bodies from a specific domain (scientific, cultural heritage, education, etc.), and they must be willing to / able to keep the identified material available. As such, individuals usually don't qualify as registrars.

1.2.2 Creating identifiers

The created identifiers need to be unique, valid and durable. Uniqueness is usually ensured by providing (decentralized) repositories with a namespace that enables them to generate their own unique identifiers. A character-set and/or a template are given to take care of the validity of the identifiers. It is recommended to use opaque identifiers or at least to minimize the use of semantics in order to increase the durability of the identifiers. In addition to the above, Sweden and Finland also provide a service to generate identifiers that are unique, valid and durable. Italy is an exception with regard to uniqueness: organizations are free to generate identifiers within the national NBN namespace and their validity will be checked upon registration. Germany adds a checksum of the identifier at the end of the identifier to enable detection errors when transferring identifiers

1.2.3 Assigning identifiers

All providers have a policy on what can be assigned an identifier, but there is variation over the different providers on the contents: Usually these differ in terms of scope (academic, cultural heritage or educational) and in type (books, publications, research data, audio/video, etc.). In addition, multiple providers do not yet know how to deal with representation (conceptual, physical, digital, online). Finally, there is not always clarity on the level of identification: a work, a specific version or even a representation[1]? Given that it will not be able to reach agreement on these issues, it is important that the chosen policy for each repository or object will be transparent.

[1]Functional Requirements for Bibliographic Records (FRBR) http://ifla.org/VII/s13/frbr/frbr.pdf

1.2.4 Registering identifiers

The registration of identifiers is needed to make the resolver aware of its existence and to allow it to redirect requests to one of the locations* that need to be registered with the identifier. Another goal of registration can be to validate the identifiers: whether they have a valid format, whether there is a valid location attached and whether the identifier is unique. It is recommended to mark the identifier on the identified object, to allow users to verify the identity of that object. This makes the workflow rather difficult: the identifiers need to be known and validated before ingesting the identified object, but not yet official and/or resolvable. Otherwise the identifier that is registered on the object might be rejected by the resolver, in which case the identified object needs to be updated with another identifier.

All resolvers have a harvester in place that harvests repositories for identifier/location pairs, both new ones and updated ones. Such a (pull-) mechanism introduces a delay between publishing an object with identifier in the repository and the registration of the identifier in the resolver, which results in temporarily irresolvable identifiers. Some resolvers provide an additional (push) web service interface and sometimes an email interface (Germany), which have no or minimal delays. Some resolvers require additional metadata to be registered: the owner, a modification date, a checksum or even complete metadata schemes (Dublin Core, METS) (See the Appendix for an overview).

* The Italian resolver does not register locations, but local identifiers. Using these identifiers, the resolver can delegate the resolution request to the repository that can retrieve the corresponding URL.

1.2.5 Updating locations

Updating the registered locations is an essential feature of persistent identifiers: it allows objects to move to new locations while keeping existing references valid. The repositories are responsible for updating the location as soon as these move, and this can always be done via the same interfaces as for registration of new identifiers. In Italy such moves can be dealt with locally because these are only registered locally. The validity of the registered locations is verified after registration. Only in case of failing resolution, some providers notify the responsible repository and request them to update the location. There is no (known) policy to avoid temporary failures between moving the object and updating the identifier. Such failures can be avoided by having temporary redundant objects on old and new locations, or by having temporary redundant registrations, where the resolver can poll both old and new locations before redirecting.

It is important that changed locations are registered to allow users to fetch the identified object. If these are not registered, it is important to notify the responsible repository and/or notify the user about the cause of the failure. Such functionality requires a service to check all identifiers for valid locations, either as a background process or at the moment of resolution. The German National Library will implement such a service within their local resolver.

1.2.6 Updating identified objects

All organizations require the repositories to assign a new identifier to a new version of the object. This rephrases the problem to: "What is seen as a new version?". There is no explicit policy on this and it cannot be validated. Germany provides registration of relations between versions, allowing users to see the identifiers of newer versions and/or the original version. Most providers are planning a more strict policy.

1.2.7 Transferring objects to new owners

The persistent identifiers are not to change, not even if the object moves to another owner. The creation of a second identifier (to the same object) is also not desirable. The existing identifier should thus be resolved to a location at the new repository and, more important, the responsibility for the object should be transferred. Transfers between repositories happen on a basis of single records, collections or complete repositories. The last case is easiest to deal with: the new repository takes over the complete namespace and responsibilities for all identified objects. The first two cases are more difficult with regard to ownership: this cannot any longer be determined on the basis of the namespace, but should be registered at the central resolver.

All providers have different workflows to deal with these issues: In Italy new repositories can register the existing objects in the new repository and can remove the old registrations via informal contact. Germany can register the new locations as newer versions of the object and the Dutch system determines ownership on individual basis, based on the primary location that is registered (if the primary location is removed, then the secondary location becomes the primary).

1.2.8 Removing identified objects

The removal of identified objects is inevitable and usually happens because of legal regulations. Examples: the repository appears not to be the owner of an object, or the object contains privacy sensitive information. In such cases most organizations consider that the best thing to do is to present a page (sometimes called a tombstone) that explains why the object is no longer there instead of redirecting to a 404 page. Other organizations consider it not their responsibility and show the 404 page.

In other cases of removal, a fallback to a copy in a Long Term Preservation (LTP) archive is desirable, but such archives/copies are not always available.

1.2.9 Fetching identified object

All organizations provide a web application that takes a persistent identifier, via either a GET- or a POST-command, does a lookup of the identifier and responds with a HTTP 302 redirect to the corresponding URL. The differences are in error handling: some organizations consider it their responsibility to avoid 404-pages, others don't. Those who do feel responsible need to test the registered location(s) and provide a redirect to a secondary URL (possibly a copy at an LTP archive) or a custom message explaining why the link broke. The way to submit the identifiers also differs: some use the path on the webserver (http://example.com/urn:example), others a query parameter (http://example.com?identifier=urn:example).

In most cases resolution is handled directly by the resolver that redirects to the registered location. In Italy, the resolver resolves via the repository. The latter is aware of the exact position of the identified object and returns the final location of the identified object.

1.2.10 Fetching metadata of identified objects

Most resolvers have a feature to do a lookup of the identifier and respond with the metadata that is available. This information is either fetched from the resolver itself or from the repository of the identified object. The feature is available either via a parameter or as a separate service with a separate URL. See the Appendix for the available metadata.

1.2.11 LTP facilities

Different countries have different long-term preservation facilities. Usually the National Libraries are responsible for LTP of publications that are created and/or published in their country. This is often a legal obligation. Besides the National Libraries there can be (or will be) other organizations that cater for LTP of e.g. research data, cultural heritage, educational resources and governmental resources.

So far none of the repositories or LTP archives have any kind of certification. In order to create a trusted environment, where contents can be safely referred to, it is essential that the trust can be verified using assessment programs like TRAC[2], DRAMBORA[3] or DSA[4], [5]. These will require an explicit statement about the mission of the organization and the policy and workflows they operate. See Appendix for a current list of LTP archives.

1.2.12 Conclusions

Regarding persistent identifier policy all organizations have a similar policy on similar issues, though some details differ. Such details are with regard to what organizations can register identifiers and what type of objects these can be assigned to. Though the workflows are more different, it still has general equalities. The differences are because the workflows are custom to the organizations they are implemented in. Still they are all similar in terms of their local maintenance and their central registration via OAI-PMH.

The implementation is naturally different in each country. The important differences are the different formats that are used: The format of the identifiers, the metadata schemas used when registering them and the interface for resolving them. Luckily, none of these exclude the creation of a common model.

Most important for persistent identifiers is the trust that they can be used safely. Such trust is primarily achieved by creating clear and reliable policy agreements. None of the partners' policies is yet defined clearly and most of them are currently under consideration. The common policy refers to these policies to be. Moreover, these should be communicated to all stakeholders.

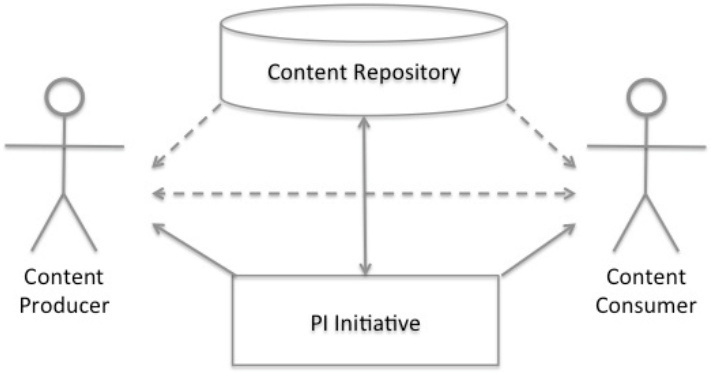
1.3 State-of-the-art in PID systems

The previous chapter gave a comparative overview of how the participating countries currently deal with persistent identifiers. This chapter will expand the scope: it will look at how alternative initiatives such as Handle and DOI are implementing persistent identifiers and it will look at developments around persistent identifiers by describing the different stakeholders, their responsibilities, their relations and the current/future developments in their area.

1.3.1 Stakeholders

The Data Seal of Approval (DSA)[6] identifies three stakeholders in the context of repositories: data producers, data repositories and data consumers. For this report we'll rename these to content producer, content repository and content consumer and add a fourth stakeholder: the PI providers.

[2]Trustworthy Repository Audit & Certification (TRAC), http://catalog.crl.edu/record=b2212602~S1
[3] Digital Repository Audit Method Based On Risk Assessment (DRAMBORA), http://www.repositoryaudit.eu/
[4] Data Seal of Approval (DSA), http://www.datasealofapproval.org
[5] A new initiative for audit and certification of Digital Repositories with support from the European Committee is under evaluation, http://trusteddigitalrepository.eu/Site/Trusted%20Digital%20Repository.html.
[6]http://www.datasealofapproval.org



The above diagram shows the relation between content producer, content repository, content consumer and PI initiative. This chapter will focus on the relation between PI initiative and the other actors (Content producer, consumer and repository). The other relations are of indirect (but not less) importance.

1.3.2 Content Producers

Content producers are responsible for depositing their content in a qualitative repository and for the quality of the deposited (and identified) content. This quality refers to: intrinsic scientific, scholarly or cultural quality of the content, quality of the deposited format and quality of the provided metadata and documentation.

Content producers deposit their contents for the following motives (either their own, or their funders'):

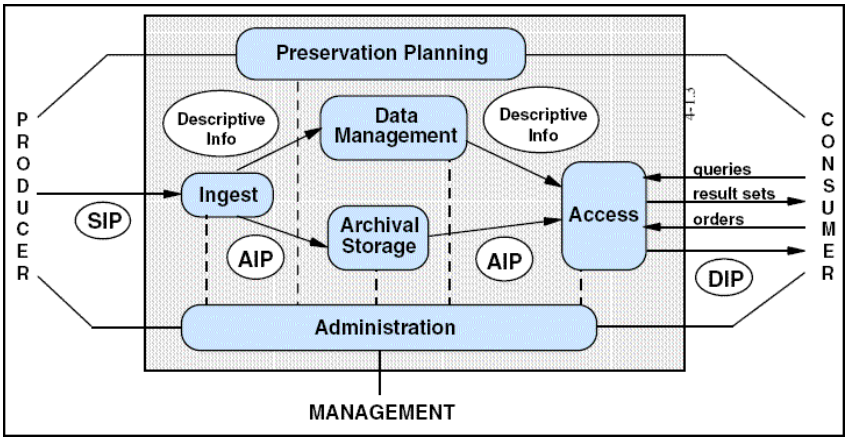
- Store the contents
- Publish the contents and receive recognition for the contents/work
- Make the contents available for online usage via e.g. European infrastructures (Europeana, CLARIN, DARIAH, …)

With regard to persistent identifiers, the biggest stake for content producers is to support citation and integration within European infrastructures.

Citation using persistent identifiers is already possible and practiced. CrossRef demonstrates how DOIs support transparency in the citation network. To achieve this CrossRef maintains a database of citations between DOI-identified publications. Identifiers that get registered need to provide the identifiers of resources that are cited by the registered resource.

1.3.3 Content Repositories

Content repositories are responsible for access and preservation of the digital contents. This requires them to have a qualitative organization and technical infrastructure: a qualitative organization has an explicit mission, complies with legal regulations, has a plan for long-term preservation, etc. A qualitative technical infrastructure complies to internationally accepted archival standards like OAIS.



Content repositories preserve their digital contents for reasons of integrity (science, national archive, national library) or collective memory (cultural heritage, national library). They are created and funded with this mission. They acquire their contents via (sometimes legal) agreements with content producers or by incidental deposits. With regard to persistent identifiers the three most important activities within the archiving process are ingest, preservation and access.

Ingest is more often based on self-archiving methods: the content producer uploads its content and provides the essential metadata. Experts then verify the quality of the contents and the metadata before accepting and/or publishing it.
In such cases it is important to define when an identifier is assigned and/or when it becomes valid. On the one hand the producers should know the identifiers beforehand so they can register it within the content itself, but on the other hand these identifiers only become valid once the repository accepts the contents.

Preservation is needed to ensure the usability of the contents. This usually implies monitoring the usability of the available formats in the repository. When one of these formats risks becoming un-interpretable (due to e.g. lack of support by modern software) they will be migrated to the most durable format at that time. In this case it is important to decide what will happen to the old formats and their identifiers. Ideally the old format is preserved with its existing identifier and the new format is assigned a new identifier. The solution of DNB allows registering a relation between these, allowing users of the old format to be notified about a more usable format, and allowing users of the new format to retrieve the original one.

Modern repositories like Fedora support access to virtual formats: contents can be requested in specific formats, which are then created on the fly. This makes identification of different files more difficult, especially as these dissemination formats come and go. Another issue regarding access is authentication. Many contents can only be made available to a selected audience. In such cases authentication is required before accessing the content. A standardized authentication mechanism is required when contents need to be accessed by automated systems, to ensure that these will be able to access the identified object, instead of an alternative splash-page informing them that they need to authenticate.

1.3.4 Content Consumers

The content consumer has the responsibility to use the contents in a qualitative way by complying with access regulations, respecting applicable licences and conforming to generally accepted codes of conduct.

The use of content ranges from viewing to analysing to modifying. Viewing requires unobstructed access in generally accepted dissemination formats or even via user-friendly online visualization services such as image-browsers, GIS-browsers, streaming AV 'YouTube' viewers. Analysis requires

availability in formats that can be imported by the tools of the researchers and the ability to browse, filter, sort, annotate and integrate it with other contents. Adaptation requires a workable and consistent versioning policy.

Content consumers desire trusted access to resources. They need control over the resources, either by having access and functionality on them or by copying it. The latter creates redundancy and complexity making it hard to manage. Ideally this is avoided by reuse (by reference) of existing resources, but this requires consumers to trust the availability and an advanced level of access.

Advanced usage of the identified contents requires more generic and more detailed identification. The desired access is on a far more detailed level: paragraphs of texts, data records, video fragments, etc. and the type of access needs to be more specific: direct access to either standardized metadata, specific formats of the content or a user friendly splash page. Authentication is essential for materials that cannot be publicly available due to copyright or privacy issues. If automated access to such contents is desired then standardized ways to authenticate are needed.

European infrastructures facilitate discovery and reuse of data. To this end they provide (or prepare) portal functions where contents can be searched and online workbenches where online tools can process online data. Persistent identifiers are needed to ensure the durability of these infrastructures by providing maintainable links. For proper functioning, these infrastructures need durable and standardized ways to retrieve specific representations of the identified resources, such as their metadata, a specific format or a specific part of the contents.

1.3.5 PI initiatives

Different PI initiatives have been created to support the activities of these content producers, consumers and repositories by providing services and policy that allow durable identification and/or location of content. The most important initiatives currently are PersID, CrossRef, Epic, DataCite and ANDS.

CrossRef
"CrossRef is an independent membership association, founded and directed by publishers. CrossRef's mandate is to connect users to primary research content, by enabling publishers to work collectively."[7] The services[8] that CrossRef exploits are nevertheless based on journal articles. The most interesting featue of CrossRef is their metadata registry, which registers the Cited-by linking system, which requires participants to "deposit references from their current material and must retrieve data (i.e. retrieve 'cited-by' links to link to other participants)"[9].

CrossRef is based on DOI (that in turn is based on the Handle System), causing every identifier to require an annual fee to finance the system.

DataCite
"The objectives of this initiative are to establish easier access to scientific research data on the Internet, to increase acceptance of research data as legitimate, citable contributions to the scientific record, and to support data archiving that will permit results to be verified and re-purposed for future study. DataCite will promote data sharing, increased access, and better protection of research investment."[10]

[7] http://www.crossref.org
[8] See for example this reverse-lookup: http://www.crossref.org/questquery/
[9] http://www.crossref.org/citedby.html
[10] http://www.datacite.org

DataCite is a relatively new initiative supported by a number of international organizations[11]. Their unique selling point is their focus on research data and their integration within the (publisher) DOI system.

EPIC
EPIC (European Persistent Identifier Consortium) is an initiative from Max Planck and CLARIN that provides allocation and resolution of persistent identifiers for the European research community. They acknowledge that:

> *...one needs a commonly agreed process and due to the importance of the resolution of the references to actual URLs for a lot of transactions, the needed resolution service has to have a high degree of robustness and reliability in the long-term.*
> http://www.pidconsortium.eu

Scientific institutions are eligible to use the system using a web-form or a web-service. Further policy is not (yet?) available.

Epic is based on the Handle System, with additional services (REST) to register persistent identifiers. They announced that the next Handle version will support identification of fragments by extending the registered location with a translation between fragment and more specific locations.

PersID
PersID[12] is a joint project of the current implementers of the URN:NBN namespace, facilitated and funded by Knowledge Exchange[13]. Their main goal to harmonize the use of the systems of the different URN:NBN implementers on the level of policy, technology and communication. Such transparency is required to support the different stakeholders of the URN:NBN system, but it also gives insight in how to deal with the use of 'different' persistent identifiers across an infrastructure.

The unique selling point of PersID is that the implementers are all national organizations with an explicit long-term vision.

All initiatives use a specific namespace and provide a resolution service. They differ in their focus on a specific community, their policy and their additional services.

ANDS
The Australian National Data Service (ANDS)[14] provides identifier services for staff at universities, government agencies, publicly funded research organizations, museums, galleries, archives, libraries and any custodian of data relevant to research. The service has an Australian focus but aims to assist the emergence of a global data commons. It expects its users to identify only objects that are relevant to the reseacrh data commons and that they have the ability and commitment to maintain the currency of the location information associated with your identifier over the long term.

1.4 PI applications

This section describes the contexts where persistent identifiers are used. These contexts imply the requirements on the services and policy for persistent identifiers.

[11] 12 organizations including the German National Library of Science and Technology (TIB), British Library (BL), Australie National Data Service (ANDS) and California Digital Library (CDL).
[12] http://www.PersID.org
[13] Knowledge Exchange
[14] http://ands.org.au/services/identify-my-data.html

Citation
The most well known application of persistent identifiers is citation and referral. Persistent identifiers can support the creation of durable and clickable links to referred resources and they can provide insight into impact of one work on the other. It requires either the submission of the relations between identified resources, or a standardized notation that can be interpreted by automated processes (like webcrawlers).

CrossRef succeeded in setting up an elaborate network of identified publications and citation by allowing depositors to supply information on citations.

Research Infrastructures
Examples of research infrastructures are DARIAH, CLARIN, CESSDA, DRIVER or SURFshare. They usually provide portals, (online) tools, collaborative workspaces and/or access to data and publications. The stability or durability of these workspaces depends on the availability of (access to) the tools and resources: These are often hosted independent of the research infrastructure.

Most of these infrastructures require more than just a reference to the identified resource; they need to access it for processing. Such access goes further than current practice, where it is often unclear what an identifier leads to: a splash page introducing the resource, the resource itself or a bibliographic record. Furthermore, sometimes it is needed to access only a part of the resource, such as a record from a dataset, a chapter from a publication, a photo from a collection or a specific question from a survey. Since many resources will be part of multiple infrastructures, such access should be unified.

A (theoretical) example of such processing is e.g. the recognition of shapes in a movie. Some algorithm needs to open a specific format of a video file and tries to recognize the shapes of existing bridges. Once it recognizes them, it outputs a reference (via persistent identifier) to the scene using a start and end-time and a description of the bridge. A researcher who investigates different bridges searches the system and finds 20 references to the Golden Gate Bridge in San Francisco. When he clicks on one of the references, the specific fragment is automatically played.

When resources are modified after processing it then a transparent versioning policy will be required.

National Libraries
National Libraries have been amongst the first to implement persistent identifiers. Their mission is to provide permanent access to the resource they archive. Their early presence on the Internet and their long-term perspective made them aware of the need for durable identification and access to their contents. One important feature is the ability to use the library copies as a fallback: if the original can no longer be retrieved, people should be able to fetch the copy from the library. This requires a relation between the original and the library-copy.

Cultural Heritage
Collections of cultural heritage have a very heterogeneous nature: their contents and their representations vary. Explicit identification of a song is difficult: is it the notes, the recording, an MP3? For some items there the only representation is a photo, for others there is nothing more than the bibliographic record.

Semantic Web
The semantic web allows the creation of semantic relations between web resources and concepts. They mandate these to be identified using URIs. Within the Linked Data movement, the most active implementers of semantic technology, the use of protocol based URIs like http-URI's is recommended. This implies that persistent identifiers are to be wrapped within the resolver URI If

the 'bare' persistent identifiers are to be included, these should be added using a 'similar to' relation.
The semantic web provides opportunities for discovery of contents, especially those within a controlled environment of libraries and archives: Once resources, their creators, research projects, funders, etc. are all properly identified these can easily be related and/or used to define virtual collections.

1.5 Conclusions and Recommendations

Every participant has a policy but it is either not formalized or out-of-date. The policy that is in place in general is very similar to the policies of others. In order to benefit on a global scale of the potential of persistent identifiers and to maximize transparency, PersID has formulated a shared policy – a Mode of Conduct - that refers to the local ones. Given this shared policy, each organization can derive its own workflow and technology, where the latter can be harmonized by defining a shared interface, that can replace or complement the existing interfaces.

2 Annex III.a1 - Resolvers, Metadata and LTP facilities

The following resolvers are used by PersID partners:

Country	Resolver
Finland	http://urn.fi
Germany	http://nbn-resolving.org
Italy	http://nbn.bncf.firenze.sbn.it/NBN/
Netherlands	http://persistent-identifier.nl http://resolver.kb.nl
Sweden	http://urn.kb.se

External resolvers are:

System	Resolver
Handle	http://hdl.handle.net
DOI	http://dx.doi.org

The metadata available at the resolvers:

Germany	Italy	Finland	Sweden	Netherlands
	Dublin Core			
Owner			Owner	Owner
Checksum	Checksum			
Mimetype				
primary(y/n)				
frontpage(y/n)				
archive(y/n)				
Created				
last modified		Last modified (URL)	Last modified (URL)	
	MPEG21			
	METS			

The following LTP facilities exist:

Country	Organization	Domain
Denmark	National Library	Publications, websites
Finland	National Library	Publications
Germany	National Library	Publications
Italy	CNR	Publications
Netherlands	National Library DANS 3TU Beeld & Geluid	Publications Research data Technical research data A/V resources
Sweden	National Library	Publications, websites

3 Annex III.a2 - Statistics and Expectations

These estimations are provided by the PersID participants.

	Italy	Finland	Germany	Sweden	Netherlands	Total
Current #id	1.500	63.000	2.1M	0.5M	0.5M	**3.1M**
Next year #id	100.000	90.000	3.9M	0.75M	1.5M	**5.5M**
Five year #id	1 – 5M	1M	5 – 10M	3-8M	5 – 10M	**15 – 35M**

All participants foresee an increase in the use of URN:NBNs and they distinguish various reasons:

- more publications and datasets will be deposited, for instance because research funders require this;
- increase in the use of persistent identifiers in government's document management;
- e-deposit laws;
- digitization projects lead to more digital objects;
- increase in the demand for identiying more fine-grained material (e.g., separately referable images instead of collections).

4 III.b – User Requirements

4.1 Revision History

Name	Date	Reason For Changes	Version
Laurents Sesink, Maarten Hoogerwerf	2 Feb 2010	Draft for PersID wp3 & wp4 meeting Frankfurt	0.1
Laurents Sesink	8 Feb 2010	Draft for comments wp3 & wp 4	0.2
Laurents Sesink	9 Feb 2010	Priorizations of requirements	0.3
Jürgen Kett	10 Feb 2010	First part of DNB feedback	0.4.jk (temporary version)
Nicole v. d. Hude Kirubel Legasion Karaca Kocer Christa Schoening	16 Feb 2010	Second part of feedback	
Christa Schoening, Laurents Sesink	10 mar 2010	Integration of comments	0.4
Laurents Sesink	19 mar 2010	Wording.	0.5
Laurents Sesink	1 April 2010	Layers	0.6
Laurents Sesink	14 April 2010	Added Use Cases	0.7
Laurents Sesink	15 April 2010	Added information on Appendix A	0.8
Laurents Sesink	18 April 2010	Added General functionality of PID resolvers	0.8
Laurents Sesink Maarten Hoogerwerf	20 April 2010	Consistent terminology	0.9
Maarten Hoogerwerf	5 Aug 2010	Processed comments Clarified delegate vs forward Updated persid to PersID	1.0
Marjan Grootveld	17 March 2011	limited redundancies in Functional requirements; limited sections on how to carry out the project	1.2
Marjan Grootveld	28 March 2011	Processed comments	1.3
Bas Cordewener	8 April 2011	Final edit / publication ready	1.4

4.2 Introduction

4.2.1 Purpose

Current persistent identifier (PID) systems do not offer the research and cultural heritage community a system that is open and international and is interoperable with existing PID solutions. This can cause problems. It is often not clear what kind of policy an authorizing organization has implemented when they assign a PID to a resource. It is often not clear where to resolve the PID. Most common users are not aware of the difference between a PID and an URL.

To solve this problem it is important to establish a PID infrastructure that is open, global and interoperable, that is controlled by the research and cultural heritage community and serves their particular needs with regard to guaranteed access and long term preservation.

In order to benefit on a global scale from the potential of PIDs and to maximize transparency, this should be transformed into a shared policy, with a minimal level of extensibility to allow minor local differences. The implementation of the shared policy will be based on trust rather than certification.

The purpose of the project is to develop atrusted persistent identifier infrastructure.

4.2.2 Project Scope

The PersID project deals with multiple types of organizations serving different communities: The project partners represent both national and academic libraries, cultural heritage organizations and data archives. The audience consists of web-users from these communities and beyond and will also be used within durable automated processes.

The different communities represent a broad range of requirements. *Citation* and *validation* are important in scholarly communication. For organizations in cultural heritage retrieval and the *authenticity* of resources is important. *Versioning* is a main issue for archives. For research and education purposes it is necessary to deal with issues like *reuse*, *analysis*, automation and *integration* of resources. *Availability*, *discovery*, *security* and *authenticity* are main topics of libraries and archives.

The proposed *trusted persistent identifier infrastructure* should support the common needs of these different communities and has to deliver a distributed resolution system that provides a unified interface to the end-user, but allows different systems to support their own communities.

The examples of the various community requirements illustrate the need for additional information and services, which extend the basic functions of the proposed *trusted persistent identifier infrastructure*.

The proposed *trusted persistent identifier infrastructure* will be basic and simple and integrate value added services which can handle functionality like granularity, authenticity and other functionalities.

The interaction of resolvers within the proposed *trusted persistent identifier infrastructure* and the maintenance/governance of the underlying infrastructure must be outlined in a transparent policy document, which in turn refers to national policies. Within the scope of the PersID initiative partners developed a Code of Conduct, as a prelude to such a policy document (see PersID Report V)

4.2.3 Intended Audience

This document describes the User Requirements identified in PersID work package 3 (WP3) and implemented by WP4 by the end of the PersID project. This is the starting point for further development within the current PersID community and for continued interaction with experts and stakeholders. The documents will be a means to explore opportunities for cooperation with projects like Europeana, ESFRI projects and e-InfraNET, with organisations that use, will use or consider the use of persistent identifiers as well other PID infrastructures.

The User Requirement Document reflects the necessary communication, coordination and alignment between the different work packages of the PersID project. For WP4 (Development and Implementation) the content of this document was essential input to implement the desired system. For WP5 (Sustainability) the overview of requirements provided ingredients to explore the cost model, and draft the Mode of Conduct and Roadmap recommendations. For WP2 (Communication) this document was a pillar for external communications about the scope and urgency of the project.

4.2.4 References

1. Knowledge Exchange and SURF PersID initiative: Global Resolution Service (GRS)
2. Work plan for URN based Persistent Identifier Infrastructure project, Final version (1 June 2009)
3. Report: Task 3.1 Current state on PID implementations, case studies and technologies

4.3 Overall Description

4.3.1 Background

Research and cultural heritage communities all have different attitudes and behaviour regarding the identification and localisation of resources. For researchers it is important that the content of the resources is authentic and that they can link to parts of a resource (fine-grained links). Usually it is not important if the information is in a Word document or in a PDF document. For national archives, however, the look and feel of resources is important because documents can have a legal status. For these purposes it is important that the document format (Word or PDF) is unchanged.

In scholarly communication there is a clear need for best practices related to the citation of publications, data sets and other web based resources. At the moment different communities covering nearly all of the sciences are building up research infrastructures. These e-research environments deal with publications, data, analyses tools and all of them express a need for PIDs. Issues like how to link a PID to a dynamic dataset and how to deal with granularity are currently investigated.

More and more educational resources are web-based. For educational purposes it is important that the resources can be re-used. Questions like how to deal with different manifestations and how to discover complex objects are being raised.

The cultural heritage community is heterogeneous and fragmented. There are large organizations that are well capable of managing a PID resolver for their own needs. On the other hand there are many small organizations that don't have the resources to maintain an infrastructure for PIDs. There is a strong need for a shared infrastructure where small organizations can join shared PID infrastructures.

Currently there are several PID infrastructures. They vary in technical solutions, policy, organization and maturity and range from well organized and operated services through technical implementations where sustainability is not underpinned by a solid organization and business model.

The Use Case scenarios (Annex III.1) illustrate that both for users in these communities and for the tools they (want to) apply it is important that there is a *trusted persistent identifier infrastructure* where they can resolve all of their PIDs and where they can find information about the reliability of the resource which the identifier locates.

4.3.2 Service Perspective

The *trusted persistent identifier infrastructure* does not only provide a technical solution, but uses open, interoperable URN:NBN-based technology in a framework of policy, organization and communication

- that can be *trusted* by the research and cultural heritage communities to *serve the community needs* and will stimulate usage of PIDs,
- that allows permanent access to resources underpinned by *global long term preservation and resolving* using PIDs,
- that is common enough to build *worldwide interoperable services* upon regarding usage statistics, global citation computing services, and preserve relations between objects within objects from a variety of resources
- that is embedded in a framework of *policy agreements* about maintenance, guarantees and quality.

The *trusted persistent identifier infrastructure* demonstrates how the fragmented systems of PIDs can be harmonised in order to increase the necessary trust and use by the end users. It starts by

integrating current national URN:NBN solutions on the level of technology, policy and communication while respecting the distributed nature of the current national systems, which is caused by differences in domain, national policy/laws, etc.

Nationally, URN:NBN PIDs issued by national libraries and other national agencies taking care of long term preservation have proven to be an effective instrument for permanent access. The *trusted persistent identifier infrastructure* aims to be as effective on an international scale by enabling the global academic and cultural heritage organizations to maintain the links to their resources via an international URN:NBN PID framework and by establishing a basic set of policy agreements between the participants.

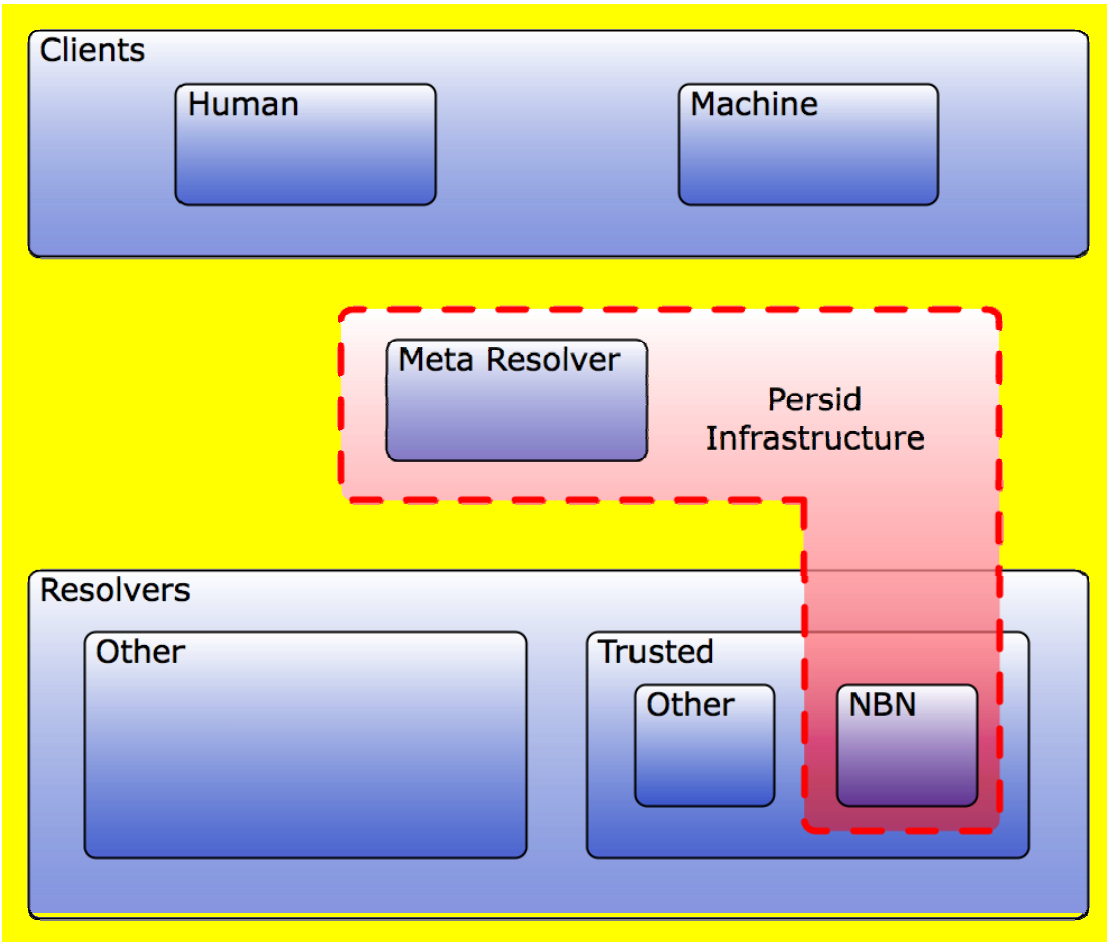
4.4 Requirements

A *user requirements document* is a document used in software engineering that specifies the requirements the user expects from software to be constructed in a software project. In the next sections three aspects are distinguished:

- The *Functional Requirements* describe the type of behaviour the user wants the system to perform. It is important to note that the requirement specifies what is wanted, and not how it will be delivered.
- There are also *Non-Functional Requirements*. Their purpose is to restrict the number of solutions that will meet a set of requirements.
- *Performance Constraints* describe how the system should perform when it is delivered.

4.4.1 Functional Requirements – general picture

Consider the following situation with various users, multiple PID systems (URN:NBN and others) and multiple resolvers per PID system:



The Meta Resolver contains a service that redirects all PIDs to the registered resolvers. The Meta Resolver also contains services to register resolvers and to provide information about all registered resolvers (in so far as a resolver provides information about itself).

There are two types of clients who can use the Meta Resolver: A human client and a client application. Both clients can use the same functionality offered by the Meta Resolver. A client application will communicate with the Meta Resolver in an automated way. Communication for human users is also supported by web pages.

PersID distinguishes three types of registered resolvers and two types of redirection:

Resolvers \ Redirection	Delegate	Forward
Trusted NBN Resolvers	X	
Other Trusted Resolvers	X	
Other Resolvers		X

Two types of redirection:

1. *Delegation* is a form of redirection that complies with the PersID policy. This policy describes what requests a resolver can expect and what type of response will be returned. This ensures that a user can trust that he receives the expected response that is well-formatted and contains the valid, identified contents according to the policy.
2. *Forwarding* is a form of redirection where it is not certain what will happen with the request: whether a response will be given, how this response will be formatted and/or what the validity of the contents is. In this case a client machine can receive a splash page when it expects the identified document, or it can receive version 1.1 of a document with a docx-extension while it

previously received version 1.0 in with a .doc extension. In other words, the response is outside the PersID remit.

Three types of registered resolvers:
1. *Trusted URN:NBN resolvers* are resolvers that service URN:NBN identifiers and comply with the policy of the PersID infrastructure to guarantee permanent access to the identified resources. The Meta Resolver will delegate resolution of PIDs to the appropriate Trusted URN:NBN resolvers.
2. Resolvers that service other identifiers like Handle, DOI or ARK but do comply with the policy of the PersID infrastructure are regarded as *Other Trusted resolvers*. At the time of writing no Other Trusted resolvers exist, but the technology is ready for it. An Other Trusted resolver will connect to the PersID Meta Resolver via or along the lines of the API that Trusted URN:NBN resolvers use. The Meta Resolver will delegate resolution of PIDs to the appropriate Other Trusted resolvers.
3. Resolvers that do not comply with the policy of the PersID infrastructure are regarded as *other resolvers*. The Meta Resolver will forward resolution of PIDs to the appropriate other resolvers.

There is a clear separation of responsibilities between the Meta Resolver on the one hand and the Trusted URN:NBN resolvers, Other Trusted resolvers and other resolvers on the other hand. This becomes clear in the MoSCoW listing of specific functional requirements hereafter. The resolvers communicate with the Meta Resolver by means of well-defined transparent communication protocols.

The *PersID infrastructure* is the service layer which provides an interoperability framework for Trusted URN:NBN resolvers based on harmonised policies and APIs. Organization, policy and business models are required to operate the PersID infrastructure, but are outside the scope of the User Requirements.

To summarize: The *trusted persistent identifier infrastructure* will exist of different service layers.
- A service layer – the PersID infrastructure – that will delegate PIDs to Trusted URN:NBN resolvers.
- A service layer that will delegate PIDs to Other Trusted resolvers.
- A service layer that will forward PIDs to other resolvers.
- All service layers are accessible for clients (human or machine) by a single access point.

4.4.2 Functional requirements – specifics

In this section the functional User Requirements are presented in two ways: first, in relation to three particular components: the Meta Resolver, the three types of resolvers introduced before, and the interface between the Meta Resolver and the resolvers. This is only a summary, in which the numbers following the requirements refer to the numbers in the so-called MoSCoW representation. The MoSCoW overview represents priorities: this method distinguishes what the system *M*ust have, *S*hould have, *C*ould have or *W*ishes to have. Because the MoSCoW representation provides a more explicit list for monitoring what functionality has already been developed and implemented, the requirements are listed there in more detail.

Component presentation
The numbers following the requirements refer to the numbers in the MoSCoW representation: Must have functionality (1-9), Should have functionality (10-13), and Could have functionality (14-22).

- Meta Resolver
    - About the Meta Resolver
        - Information about the Meta Resolver: 2
        - End User Feedback: 11
    - Meta Resolver register

        - Information about resolvers that are registered with the Meta Resolver: 3
        - Information about how to register with the Meta Resolver: 10
        - Add, Edit, Delete resolvers (note that the requirements on Other resolvers are a subset of the requirements on Trusted URN:NBN and Other Trusted resolvers): 4, 10, 17
    - Resolve a Persistent Identifier
        - There is one base URL to resolve PIDs for all registered resolvers: 5
        - A client (human or machine) can choose one of the following options (…): 5, 15, 18
    - PID Metadata in the PersID infrastructure layer
        - Authority metadata establishes identifier provenance and trustworthiness. A client (human and machine) can retrieve metadata about the PID: 6, 19
    - Resource Metadata in the PersID infrastructure layer (optional)
        - A client (human or machine) can retrieve metadata about the resource when this metadata is available: part of 8, 21
    - Information about how to reference a PID in the PersID infrastructure layer: 1
    - Error handling: 7, 16, 20
    - System Administration: 13
    - Interface between the three types of resolvers and the Meta Resolver. A standard common interface must be defined for the communication between the Meta Resolver and the registered resolvers.
    - The interface should be able to communicate the following information (…): 8, 21
    - Metadata
        - Provide PID metadata in a standard metadata format: 8, 21
        - Provide resource metadata in a standard metadata format (optional): 8, 21
- Types of resolvers
    - All resolvers must provide the following functionality (…).(Note that the requirements on Other resolvers are a subset of the requirements on Trusted URN:NBN and Other Trusted resolvers): 9, 15, 22
    - Trusted URN:NBN resolvers and Other Trusted resolvers are able to communicate with the Meta Resolver through the defined Interface: 9, 22

MoSCoW presentation: priorities

Must have

1 Reference a PID	**Meta Resolver**
1.1 Information how to reference a PID in the PersID infrastructure layer	

2 Information about Meta Resolver	**Meta Resolver**
2.1 Information page about Trusted URN:NBN resolvers	
2.2 Information page about Other Trusted resolvers	
2.3 Information page about other resolvers	

3 Information about all registered resolvers	**Meta Resolver**
3.1 Browse through all registered resolvers	
3.2 Search for a specific resolver	
3.3 Browse through all registered Trusted URN:NBN resolvers	
3.4 Search for a specific Trusted URN:NBN resolver	

4 Register Trusted URN:NBN resolvers	Meta Resolver
4.1 Add, Edit, Delete resolvers Resolver name Base URL Policy information Governing organization Business model Link to the resolver website Contact information (administrator).	
4.2 Browse through Trusted URN:NBN registered resolvers	
4.3 Search for a specific Trusted URN:NBN resolver	
4.4 View provenance metadata about Trusted URN:NBN registered resolvers	

5 Resolve a Trusted URN:NBN identifier	Meta Resolver
5.1 There is one base URL to resolve PIDs for all registered resolvers.	
5.2 A client (human or machine) can: Retrieve the resource. Retrieve the location(s) of the identifier. Get information about the resolver. Get information about the organization that authorized the PID.	

6 PI metadata Trusted URN:NBN identifier	Meta Resolver
6.1 A client (human or machine) can retrieve PID Metadata Date of last update of identifier Identifier creator Identifier owner(s) A link to the metadata of the object	

7 Error handling	Meta Resolver
7.1 Give feedback to users when: A resolver does not respond. An object is not found.	
7.2 Notify the maintainer of the registered Trusted URN:NBN resolver when: The resolver does not respond. An object is not found.	

8 Interface	Interface (pil)
8.1 The Interface must allow communication between the Meta Resolver and the Trusted URN:NBN resolvers about the following topics. Retrieve the resource (object itself). Retrieve the location (URL) of the identifier. Retrieve the locations (list of URLs) of the identifier. Provide information about the resolver. Provide PID Metadata Provide Resource Metadata (optional)	

9 Requirements for Trusted URN:NBN resolver	Trusted URN:NBN resolvers
9.1 Trusted URN:NBN resolvers must provide functionality to: Retrieve the resource (object itself). Retrieve the primary location (URL) of the identifier. Retrieve the locations (list of URLs) of the identifier. Provide information about the resolver. Provide PID Metadata Provide Resource Metadata (optional) Communicate by means of the Interface with the Meta Resolver	

Should have

10 Information about how to register a resolver	Meta Resolver
10.1 Information page how to register a Trusted URN:NBN resolver	
10.2 Information page how to register a Other Trusted resolver	
10.3 Information page how to register an other resolver	

11 End User Feedback	Meta Resolver
11.1 A human client can give information about a PID which does not retrieve a resource	
11.2 A human client can give information about slow response time	
11.3 Collect and forward this information to the responsible resolver.	

12 PID Syntax	Trusted URN:NBN resolvers
12.1 There should be a clear policy on the syntax of identifiers. What characters and what semantics are allowed?	

13 System Administration	Meta Resolver
13.1 System administrators must be able to easily check the status of various parts of the Meta Resolver, such as load.	
13.2 System administrators must be able to easily get some basic statistical data about the usage of the Meta Resolver, e.g. about users, requests, or partners.	

Could have

14 Register Other resolvers	Meta Resolver
14.1 Add, Edit, Delete resolvers Resolver name Base URL Link to the resolver website Contact information (administrator)	
14.2 Browse through registered other resolvers	
14.3 Search for a specific other resolver	
14.4 Provenance metadata of registered other resolvers	

15 Resolve an other identifier	Meta Resolver
15.1 A client (human or machine) can: Retrieve the resource Get information about the resolver	

16 Error handling	Meta Resolver
16.1 The Meta Resolver gives feedback to clients (human or machine) when: A resolver does not respond. An object is not found.	
16.2 The Meta Resolver gives feedback to Other resolvers when: A resolver does not respond. An object is not found.	

Wish to have

17 Register Other Trusted resolvers	Meta Resolver
17.1 Add, Edit, Delete resolvers: Resolver name Base URL Policy information Governing organisation Business model Link to the resolver website Contact information (administrator)	
17.2 Browse through Other Trusted resolvers	
17.3 Search for a specific Other Trusted resolver	
17.4 Provenance metadata of registered Other Trusted resolvers	

18 Resolve a Other Trusted identifier	Meta Resolver
18.1 A client (human or machine) can: Retrieve the resource. Retrieve the location(s) of the identifier. Get information about the resolver. Get information about the organization that authorized the PID.	

19 PI metadata Other Trusted identifier	Meta Resolver
19.1 A client (human or machine) can retrieve PID Metadata Date of last update of identifier Identifier creator Identifier owner(s) A link to the metadata of the object	

20 Error handling	Meta Resolver
20.1 The Meta Resolver gives feedback to clients (human or machine) when: A resolver does not respond. An object is not found.	
20.2 Give feedback to registered Other Trusted resolvers when: A resolver does not respond. An object is not found.	

21 Interface	Interface (other Trusted layer)
21.1 The Interface must allow communication between the Meta Resolver and the Other Trusted resolvers about the following topics. Retrieve the resource (object itself). Retrieve the location (URL) of the identifier. Retrieve the locations (list of URLs) of the identifier. Provide information about the resolver. Provide PID Metadata Provide Resource Metadata (optional)	

22 Requirements for Other Trusted resolver	Other Trusted resolvers
22.1 Trusted URN:NBN resolvers must provide functionality to: Retrieve the resource (object itself). Retrieve the location (URL) of the identifier. Retrieve the locations (list of URLs) of the identifier. Provide information about the resolver. Provide PID Metadata Provide Resource Metadata (optional) Communicate by means of the Interface with the Meta Resolver	

4.4.3 Non-Functional Requirements

Trusted
Hosting and governance of the system should be shared amongst different partners.

Secure
The proper functioning of the resolver should not be endangered because of security flaws. This issue ranges from system security to social security.
The meta resolver should have an infrastructure and workflows

- against information loss
- to avoid total failure of the system (need of backup strategies)

Monitored
The functioning of the Trusted URN:NBN resolvers and Other Trusted resolvers should be monitored (only technical: does the server respond?).

Responsibility
The Meta Resolver delegates responsibility for the correctness of the links and the authenticity and validity of the objects to the Trusted URN:NBN resolvers, the Other Trusted resolvers or other resolvers. The Meta Resolver does not deal with access restrictions for legal, commercial or other issues that might be assigned to a resource.

Delegate authenticity of the identified resources
In a web-based environment it can be important to ensure the authenticity of the digital objects and the metadata. This pertains to the degree of reliability of the original and to the provenance of the resource. When a PID is an aggregation of resources it is important that existing relationships between the resources and explicit links are maintained. An example of these kind of relationships are enhanced publications, where publications are linked to data sets. This is not part of the PersID infrastructure. It is the responsibility of the content providers to manage the relationships between digital entities.

Language
The common language of the system is English. The system should support other languages.

Impact on Trusted URN:NBN resolvers and Other Trusted resolvers
In the first stage of building the system it must require only minor changes to adjust the Trusted URN:NBN resolvers to communicate with the Meta Resolver by means of the defined interface layer.

Standardisation
PersID partners will participate in the IETF URNBIS working group, which is currently revising the URN-related Internet standards. These standards specify for instance the URN syntax and register a namespace for NBN. In addition to the work in IETF, it is necessary to provide guidelines for e.g. URN assignment. National libraries should also consider the possibility of standardizing at least some aspects of the URN system in ISO, and investigate ways in which URN and other persistent identifier systems will co-operate in for instance specification of resolution services.

4.4.4 Performance Constraints

Reliable
The Meta Resolver is an essential part in different systems and processes. Users worldwide require the Meta Resolver to be available 24*7. The Meta Resolver will be used for different kinds of resources. Some of them could be popular for a while and attend a lot of traffic. The Meta Resolver, the Trusted URN:NBN resolvers and the Other Trusted resolvers should respond without a recognizable delay, and the response should be consistent. Furthermore, the Meta Resolver should be able to handle multiple requests simultaneously.
Maintenance of the Meta Resolver is essential and thus implies redundancy and strict procedures.

Extensible
There is a great potential for using the Meta Resolver. Administration of the processes should not be difficult or time consuming. It should be an easy task to integrate new Trusted URN:NBN resolvers, Other Trusted resolvers or other resolvers.

Support Value Added Services
PIDs and their resolvers are a crucial component in different systems, that each might require different features. To ensure longevity, the feature set of the Meta Resolver should be kept to an absolute minimum. Specific features are in principal additional services that should be kept external where possible. Only features that are shared by all Trusted URN:NBN resolvers, that will evidently be there forever and that logically belong to a resolver, might be integrated into the meta resolver.

Caching
To avoid the risk of delivering a non-functional link caching of PIDs should not be used in the Meta Resolver.

4.5 Obtaining the user requirements

Managing user requirements focuses specifically on identifying, gathering, communicating, and documenting user requirements for an IT system. From the start of the PersID project there has been close cooperation with different users of the system in order to obtain "the" user requirements.

National URN:NBN resolvers
Several representatives of national URN:NBN resolvers are involved in the specification of the user requirements: the German National Library, The Swedish National Library, and The Italian National Library.

PID infrastructures (other than URN:NBN)
Two of the most important persistent identifier infrastructures have been consulted. There is close cooperation with DataCite and EPIC regarding requirements of the re-directing service. (See Annex III.2)

Cultural Heritage Portals
There is close cooperation with the Europeana project. User Requirements have been discussed and adjusted to the needs of the cultural heritage community.

Research Infrastructures
There is a couple of European funded Research Infrastructures who are in the implementation or preparatory phase. All of them have a need for PIDs. PersID members cooperate with CESSDA, DARIAH, EHRI, eScholar, and CLARIN to discuss and fine-tune the user requirements in such a way that they can be included in the system. (See Annex III.2)

End users
A select group of international experts in the field of permanent access to web-related resources has reviewed the user requirements. The experts are Adam Farquhar (British Library), Clifford Lynch (Coalition for Networked Information) and Andrew Treloar (Australian National Data Service). They represent the end user of the system.

5 ANNEX III.b1 - Use Case Scenarios

5.1 A Researcher

Paul Weller is a researcher at the University of Amsterdam. His current research interest is the behaviour of politicians before elections and after they are elected. To investigate this behaviour he uses as input political programs (what did they promise before the elections) and minutes from the parliament (what did they actually do after being elected). Both resources are accessible on the Internet.
Paul wants to use these resources. For Paul it is important that the resources are authentic and that they will be available not only today, but also in 10 years time.

Based on these two resources he publishes his results.
Paul wants to refer in his publication to these resources for reasons of verification and validation for his peer reviewers.

For Paul it is important to know:
- Are the resources authentic?
- Are the resources permanent accessible?
- Where can a resource be retrieved?
- How to refer to the resources?

The *trusted persistent identifier infrastructure* can provide Paul with answers to these questions. The PID is the key to unlock the answers for Pauls basic questions.

Paul can go to the Meta Resolver (the one stop shop) and retrieve information about the organization that authorized the PID. When the authorizing organization is part of the *trusted persistent identifier infrastructure* then Paul knows it is an organization he can trust. Paul can also retrieve the information of the organization regarding their policy on permanent access. The *trusted persistent identifier infrastructure* supplies Paul with good practices on how to reference the resources.

5.2 A machine

The Council of European Social Science Data Archives (CESSDA) is building up an European Research Infrastructure to give researchers better access to social science data. One of the building blocks of this infrastructure is the Data Documentation Initiative (DDI3.0) metadata standard. This standard covers the whole lifecycle of scientific data. Discovery & Planning, Initial Data Collection, Final Data Preparation & Analyses, Publication & Sharing, Long-Term Management.

On this DDI3.0 fundament there will be additional services like a Question Data Bank and Harmonization Tools. Most of these services will extract information from DDI3.0 resources in an automated manner. PIDs form an important part of the internal functioning of this infrastructure.

For a client application (a software tool) it is important to know:
- What kind of information can be expected?
- Are the resources authentic?
- Are the resources permanent accessible?
- Where can a resource be retrieved?

The *trusted persistent identifier infrastructure* can provide the CESSDA infrastructure with answers to these questions. The PID is the key to unlock the above questions.

For tools as well as for people it is important to have a single or unique location where all PIDs can be resolved. When resources in the *trusted persistent identifier infrastructure* are retrieved the tools can rely on the fact that these resources are permanent accessible and unchanged.

6 ANNEX III.b2 - Research and Persistent Identifier Infrastructure

6.1 Research Infrastructures

The following projects are investigating or building Research Infrastructures. All of them deal with the need for Persistent Identifiers. The project proposals are used to investigate the proposed user requirements for PIDs.

DARIAH – Digital Research Infrastructure for the Arts and Humanities.
The mission of DARIAH (Digital Research Infrastructure for the Arts and Humanities) is to enhance and support digitally-enabled research across the humanities and arts. DARIAH aims to develop and maintain an infrastructure in support of ICT-based research practices.
http://www.dariah.eu/

CLARIN – Common Language Resources and Technology Infrastructure.
CLARIN is committed to establish an integrated and interoperable research infrastructure of language resources and its technology. It aims at lifting the current fragmentation, offering a stable, persistent, accessible and extendable infrastructure.
http://www.clarin.eu/

CESSDA – Council of European Social Science Data Archives. The objective of CESSDA is an innovative European Research infrastructure for the social sciences, where scientists can find data, tools, advice and exchange knowledge.
http://www.nsd.uib.no/Cessda/

6.2 Persistent Identifier Infrastructures

DataCite - International Initiative to Facilitate Access to Research Data
Uses DOI
http://www.datacite.org/

EPIC - European Persistent Identifier Consortium
Uses Handle
http://www.pidconsortium.eu/

Handle System
Uses Handle
http://www.handle.net/

CrossRef
Uses DOI
http://www.crossref.org/

6.3 National URN:NBN resolvers

Netherlands: Basic Resolution Service (BRI), hosted by Data Archiving & Networked Services. Service for publications and research data.

Netherlands:	http://persistent-identifier.nl
	http://resolver.kb.nl
Germany:	http://nbn-resolving.org
Italy:	http://nbn.bncf.firenze.sbn.it/NBN/
Sweden:	http://urn.kb.se
Finland:	http://urn.fi