



PersID V – Sustainability

Overview and studies on
persistent identifier
infrastructure
commissioned by
Knowledge Exchange

and

Prototype development of
Meta Resolver Solution
commissioned by
SURFfoundation



About this publication

PersID - V Sustainability

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

Nicole von der Hude	- German National Library
Maurizio Lunghi	- FRD
Marjan Grootveld	- DANS
Maurice Vanderfeesten	-SURFfoundation
Bas Cordewener	- SURFfoundation
Roberto Puccinelli	- CNR
Stina Degerstedt	- Royal Library Sweden
Esa-Pekka Keskitalo	- National Library Finland
Adrian Price	- Copenhagen University

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

© Knowledge Exchange partners
April 2011

URN:NBN:nl:ui:13-o4p-8py

This publication is published under the Creative Commons Attribution 3.0 Netherlands Licence.



Contents

1	Sustainability: Cost model to run a URN Resolving Service.....	5
1.1	Purpose	5
1.2	Approach	5
1.3	Regular service in the Netherlands, DANS.....	10
1.4	Regular service in Germany/German National Library (DNB).....	11
1.5	Regular service in Sweden / National Library.....	12
1.6	Regular services in Finland / National Library	13
1.7	Regular Services in Italy	14
1.8	Generic Regular Services Model (based on available information)	15
1.9	Conclusions and recommendations	17
2	Sustainability: Mode of Conduct	19
2.1	Introduction PersID Mode of Conduct	19
2.2	PersID Mode of Conduct.....	19
3	Sustainability: Roadmap.....	21
3.1	Roadmap for PersID partners.....	21
3.2	Roadmap for national organisations	22
4	Annex 'Work in Progress' - Policy Draft.....	23

1 Sustainability: Cost model to run a URN Resolving Service

1.1 Purpose

The main aim of the cost model is to get an estimation of input of work time and a general survey of the possible supply of services concerning the 'URN Resolving Service' product. The URN Resolving Service includes assigning, ingesting and retrieval of URNs and the persistent identifier service in general. Services like LTP and repository are not included.

1.2 Approach

Within the PersID project a matrix was developed containing all kind of services, the departments involved and the working hours. Working hours are net values without holidays, sick days etc. With the help of the matrix, project partners could choose what kind of services concerning the Persistent identifier system they provide and how many hours are spent on the service.

Project partners agreed on a multilevel model about costs for a URN Resolving Service:

- Level 1: Resources needed to establish an URN-service (start-up phase)
- Level 2¹: Resources needed to keep up an URN-service on intermediate level
- Level 3: Resources to keep up a complete URN-service including conceptualizing and implement new service, attend workshops, etc.

On this basis, project partners can compare among each other the different services and the associated input of manpower. These figures are collected mainly for internal use of the project partners, but interested colleagues can have a look at them in detail on request.

In a second step, these figures help new participant of the Persistent Identifier System URN to get a concrete idea of the Benefit-Cost Analysis. Moreover, this survey can serve as a checklist to support the decision making process regarding what scope the new service should have and if it will be affordable.

Finally we draw conclusions and give recommendations about an optimised URN Resolving Service.

¹Level 2 was eliminated because it proved impossible to get a sufficient amount of valid data that could be attributed to this level.

Level 1. Resources needed to establish an URN-service (start-up phase)

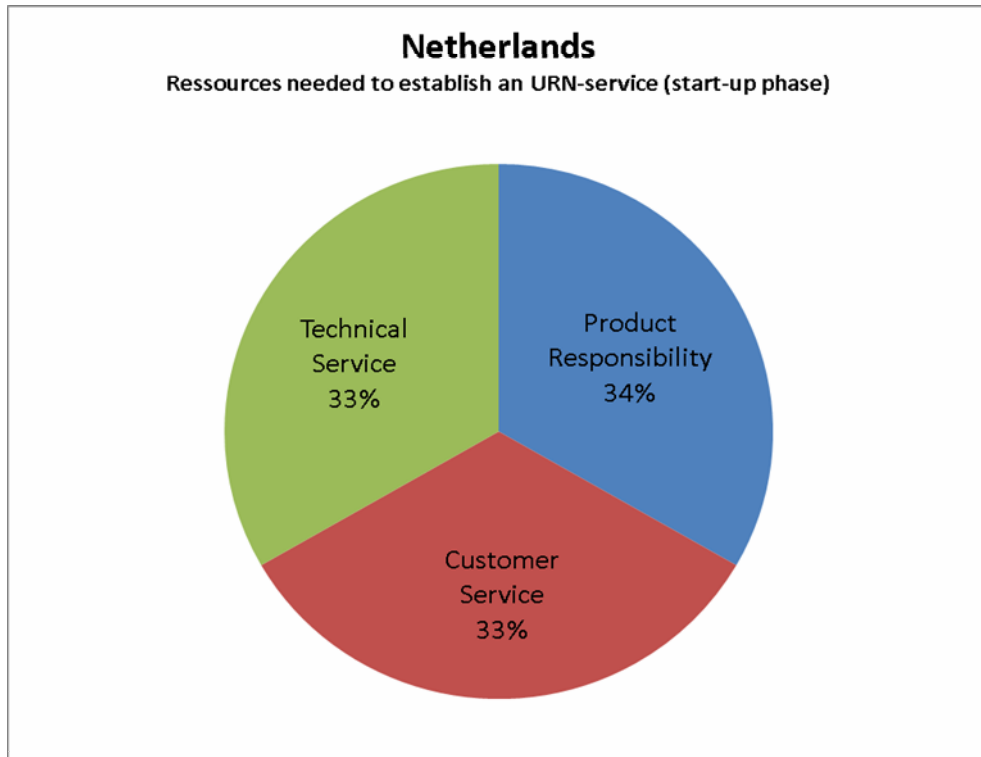
Table of costs – Netherlands, start-up phase.

Level 1: Netherlands / DANS Resources needed to establish an URN-service (start-up phase) (Netherlands: to keep the resolver up and running)				
start-up phase (15 months)	Product Responsibility Strategy and policy, overview of established URN services and first drafts Public Relations and Marketing (website, press relations, participation in meetings, conferences and expert groups) Participation in projects (local, national and european) Release Management Customer Relationship Management	Customer Service First Level Support Helpdesk for prospective customers intending to assign URN's Administration of namespaces Assign URNs for online resources archived by the National Library URN Database Maintenance Support for corrections, deletions etc. of URN's or URL's Deletion of errors caused by customers delivering urn's or / and metadata for online publications Helpdesk for error messages Complaint Management (Part of CRM) Maintenance of transfer interfaces (e.g. OAI, e-mail ingest)	Technical Service Second level support Third level support / Bug fixing.	Service Evolution Overview about existing URN services Participation in meetings, conferences and expert groups Contact and exchange of experiences with established URN services Work out user requirements Continuous communication with technical development Definition of work routines for the URN service like customer and technical service
	34%	33%		33%

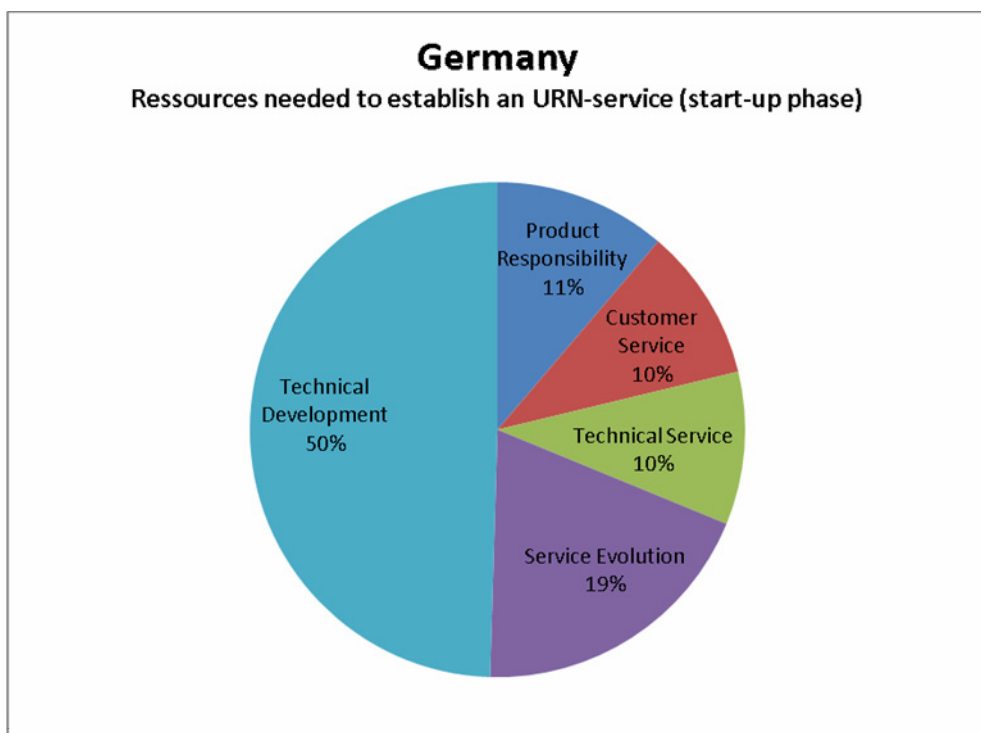
Table of costs – Germany, start-up phase.

Level 1: Germany / DNB Resources needed to establish an URN-service (start-up phase)					
start up phase (12 months)	<p>Product Responsibility Strategy and policy, overview of established URN services and first drafts</p> <p>Public Relations and Marketing (website, press relations, participation in meetings, conferences and expert groups)</p> <p>Participation in projects (local, national and european)</p> <p>Release Management Customer Relationship Management</p>	<p>Customer Service First Level Support</p> <p>Helpdesk for prospective customers intending to assign URN's</p> <p>Administration of namespaces</p> <p>Assign URNs for online resources archived by the National Library</p> <p>URN Database Maintenance</p> <p>Support for corrections, deletions etc. of URN's or URL's</p> <p>Deletion of errors caused by customers delivering urn's or / and metadata for online publications</p> <p>Helpdesk for error messages</p> <p>Complaint Management (Part of CRM)</p> <p>Maintenance of transfer interfaces (e.g. OAI, e-mail ingest)</p>	<p>Technical Service Second level support Third level support and Bug fixing.</p>	<p>Service Evolution Overview about existing URN services</p> <p>Participation in meetings, conferences and expert groups</p> <p>Contact and exchange of experiences with established URN services</p> <p>Work out user requirements</p> <p>Continuous communication with technical development</p> <p>Definition of work routines for the URN service like customer and technical service</p>	<p>Technical Development Overview about existing technical solutions</p> <p>Functional requirements considering the user requirements</p> <p>Architectural design</p> <p>Decision about software, hardware</p> <p>Connection with the entire IT-infrastructure</p> <p>Development of a prototype</p> <p>Test phase / approval system</p> <p>Move to production</p> <p>Bugfixing</p> <p>Continuous communication with service part</p>
	11%	10%	10%	19%	50%

Circle diagram of costs – Netherlands, start-up phase



Circle diagram of costs – Germany, start-up phase



For the start up phase there are different solutions in Germany and the Netherlands as shown in the graphics above. The total number of hours spent on the service range from 200 hours to 1.500 hours per year. The services offered are not the same and the main focus and/or emphasis is different, for details see the tables.

Level 3: Resources to keep up a complete URN service including conceptualizing and implement new service, attend workshops, etc.

Table of costs – regular service.

Product Responsibility	Customer Service	Technical Service	Service Evolution	Technical Development
<ul style="list-style-type: none"> • Strategy and policy • Public Relations and Marketing (website, press relations, participation in meetings, conferences and expert groups) • Participation in projects (local, national and European) • Release Management • Customer Relationship Management 	<ul style="list-style-type: none"> • First Level Support • Helpdesk for prospective customers intending to assign URN's • Administration of namespaces • Assign URNs for online resources archived by the National Library • URN Database Maintenance • Support for corrections, deletions etc. of URN's or URL's • Deletion of errors caused by customers delivering urn's or / and metadata for online publications • Helpdesk for error messages • Complaint Management (Part of CRM) • Maintenance of transfer interfaces (e.g. OAI, e-mail ingest) 	<ul style="list-style-type: none"> • Second level support • Third level support / Bug-fixing. 	<ul style="list-style-type: none"> • Service Design • Continuous improvement of the service 	<ul style="list-style-type: none"> • Technical Design • Continuous improvement

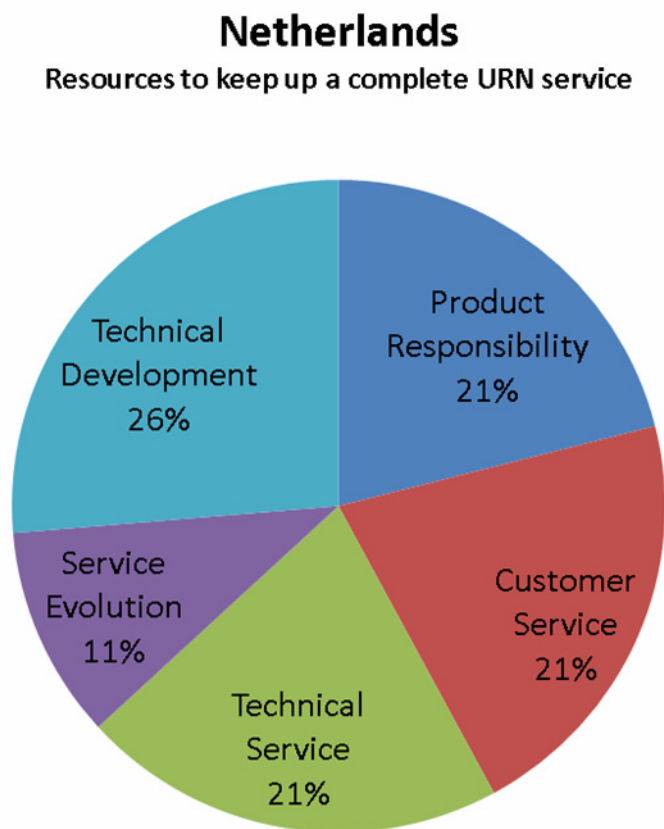
These services are necessary to keep up a complete URN service in the opinion of the PersID partners. All partners offer these services but priorities are set different among the partners. The differences are documented in the graphics on the following pages.

1.3 Regular service in the Netherlands, DANS

Number of URNs about 400.000. 30 Dutch organisations are owners of subname spaces and allowed to assign PIDs, but in practice just Utrecht University and DANS do so actively.

- Currently, DANS spends very little on outreach/marketing/PR.
- The cost of the two virtual production servers DANS uses amounts to €10.000 a year.

Circle diagram Regular service – Netherlands



1.4 Regular service in Germany/German National Library (DNB)

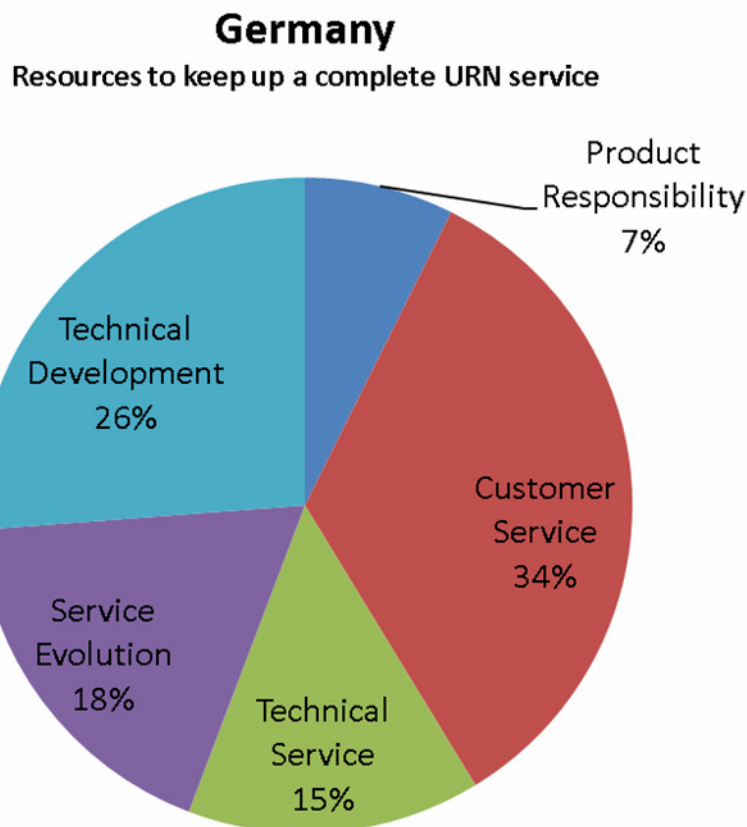
Number of URN's circa 4.400 000 in the resolver Database. At the moment DNB is also the host for the National Libraries of Switzerland and Austria. The number of URN's of these two partners is not significant (Switzerland about 10.000 and Austria less than 100).

DNB assigns URNs for digital objects being part of the [collection mandate](#). All digital object being part of the collection mandate are long term preserved in the DNB repository. DNB administrates at the moment 420 URN sub-namespaces for those institutions and publishers, who wants to assign URN:NBN on their own. Only those are allowed to have a 'sub name space' that guarantees the long term accessibility of the object.

Assigned URN identifiers and corresponding URLs must be registered at the National Library

German National Library spends significant time on outreach/marketing/PR, standardization, projects concerning PI and LTP etc. Moreover, DNB spends a lot of time in the development of the new resolver.

DNB counted 823.929 requests to resolve during a 12-month period. The monthly average is around 68,660 and daily average around 2,257 request.



Circle diagram
Regular service -
Germany

1.5 Regular service in Sweden / National Library

Number of URN's circa 1.000 000. There are for the time being approximately 50 registered 'sub-name spaces' in the Swedish URN:NBN register.

Assigned URN identifiers and corresponding URLs must be registered at the National Library.

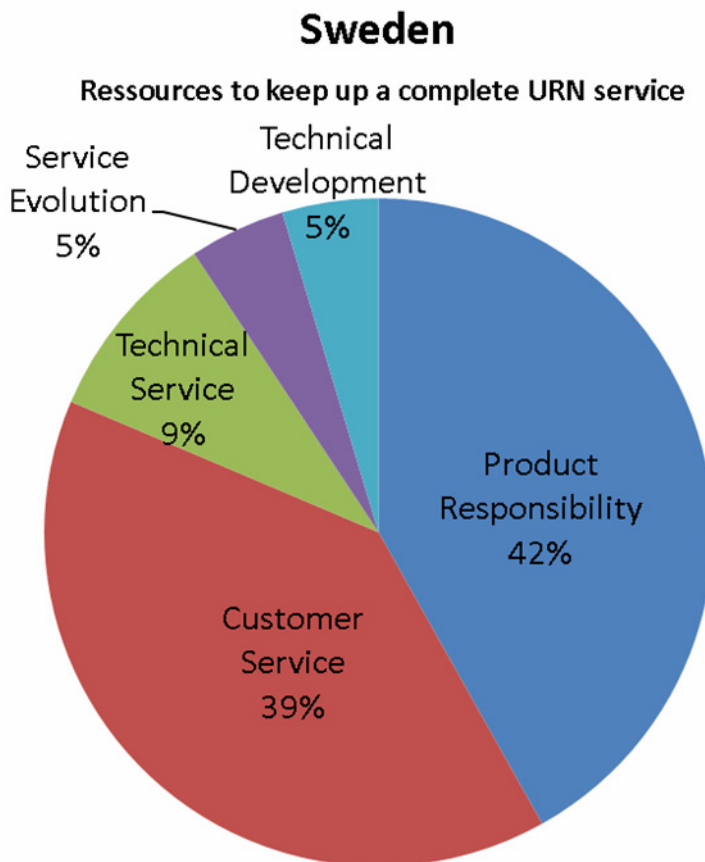
The URN:NBN register and resolver at the National Library of Sweden works as a centralized service for Swedish universities and other publishers (including the National Library's own digitization projects).

NBN numbers may be obtained from the URN generator maintained by the National Library. A publisher may apply for an NBN sub-namespaces and generate numbers independently.

The library is currently investigating a significant upgrading of both the resolver and register services.

The average amount of requests is around 25,000 per day.

Circle diagram Regular service - Sweden



1.6 Regular services in Finland / National Library

Number of URN's circa 51,000 in the resolver database.

The National Library of Finland's URN register works as a decentralized service for Finnish universities and other institutions.

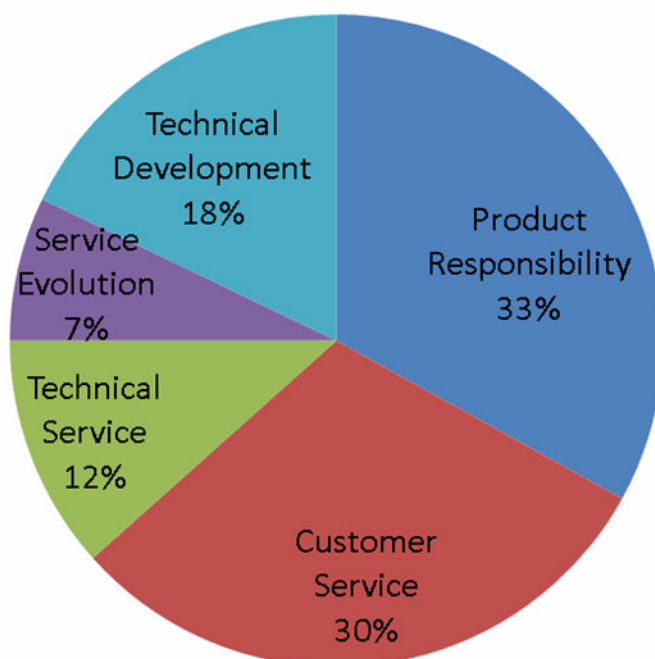
Currently private persons cannot register URN identifiers. NBN numbers may be obtained from the URN generator maintained by the National Library. A publisher may apply for a NBN sub-namespace and generate numbers independently

Assigned URN identifiers and corresponding URLs must be registered at the National Library. A persistent web address comprises the URN identifier prefixed with the address of the resolving service

There has been round 6,000,000 requests to resolve during a 12-month period. (5,936,754 to be exact.) Thus the monthly average is around 500,000, and daily average around 16,500.

Circle diagram Regular service –Finland

Finland
Resources to keep up a complete URN service



1.7 Regular Services in Italy

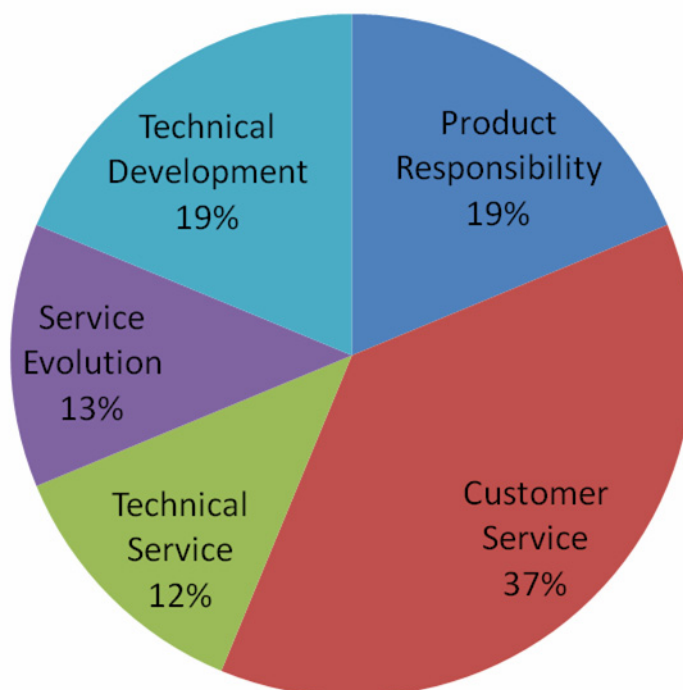
The National Libraries in Florence and Rome are responsible for the legal deposit service for a long term preservation; in that context they manage also the NBN domain for Italy. The policy is that the central node will certify a certain number of second level agencies, if they meet some criteria, and will delegate to them the authority and responsibility to select content, assign a name and preserve the digital material. The central node will harvest all the second level agencies and collect all the names in order to create a unique national NBN register. A controlling process will avoid duplication of names for the same resource.

The resolution service is different if addressed from Italy or from abroad. In the first case any second level agency can act also as resolver, in the second case a user must address the national resolver offered by the central node. This distributed architecture for roles and responsibilities can reduce costs for the central node and economy of scale. Unfortunately up to now the name-generator is still a stand-alone programme and this causes a rise of costs. In future the name - generator should be embedded in the management system.

Nobody can buy a NBN, the regular route is to ask the National Library to assign a name or a sub-domain to a specific request. The number of URN's no more than 10.000. The system is still under testing and assessment.

Circle diagram Regular service –

Italy
Ressources to keep up a complete URN service

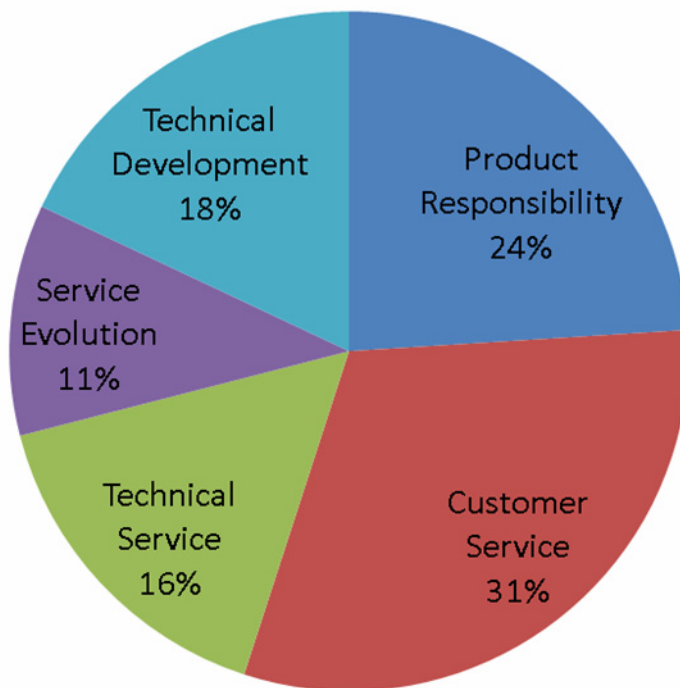


Italy

1.8 Generic Regular Services Model (based on available information)

The generic model shown below is the result of combining the figures and graphics per country and builds the average of the figures delivered by the partners. The total number of hours spent on the URN service shows a high bandwidth from 450 hours per year up to 2,900 hours per year. Explanations for this will be found in the following chapter 'conclusions and recommendations'.

Model
for resources to keep up a complete URN service



The tasks 'Technical Development', 'Product Responsibility', 'Service Evolution', 'Technical Service' and 'Customer Service' are listed in detail on the next page in the Model for the tasks belonging to a full service level URN service (Level 3).

Model for the tasks belonging to a full service level URN service (Level 3)

<p>Product Responsibility</p> <p>Tasks:</p> <ul style="list-style-type: none"> • Policy and Strategy • Public Relations and Marketing (Website, Press Relations, Attendance of meetings, conferences and expert groups) • Participation in projects (national and international) • Release Management • Customer Relation Management (CRM) for National Libraries and international organizations 	
<p>Customer Service, Support and Clearing</p> <p>Tasks:</p> <ul style="list-style-type: none"> • First level support • Helpdesk for prospective customers (intending to use URNs) • Administration of name spaces • Assign URNs • URN database management • Support for corrections, deletions of URNs, URLs • Deletion of errors • Helpdesk for error messages • Complaint management (part of CRM) • Maintenance of transfer (e.g. OAI, email ingest) 	<p>Service Evolution</p> <p>Tasks:</p> <ul style="list-style-type: none"> • Service design • Continuous improvement of the service
<p>Technical Service</p> <p>Tasks:</p> <ul style="list-style-type: none"> • Second level support • Third level support and bug fixing 	<p>Technical Development</p>

1.9 Conclusions and recommendations

There is a wide variety between the project partners' approach to the URN service and therefore the costs vary as well. Sharing and comparing experiences leads partners to conclude that the costs largely depend on different preconditions and requirements. A more in depth investigation is required but PersID partners consider the following conclusions and recommendations valuable for current decision making:

- In general the task "URN service / PI service" is comparatively new and therefore includes many costs for "trial and error"
- There is a correlation between the number of customers and the costs of customer service , e.g. in Germany there are 420 customers, in Sweden 50 and in the Netherlands 30. The resources spent vary accordingly. Conclusion: with more customers the costs will rise.
- There is a correlation between the number of URNs and the costs in general, e.g. Germany has nearly 4 millions of URN's and the highest costs but the lowest cost per URN. Conclusion: absolute costs per URN will go down significantly when the number of URNs per service increase.
- Costs related to customer services are independent from the PI system (e.g. Handle, DOI, URN:NBN, others)
- Costs can be reduced by more transparency about the responsibility for errors
- Currently human interaction represents a significant part of the cost. Costs can be reduced by
 - Distribution of responsibility and the adequate error management
 - Automation of harvesting, feedback and other services
- In relation to the total costs for Long Term Preservation, the costs for URN services are very low (compared to the actual costs of permanent storage)
- Hardware costs for URN services are moderate (at an average 4,500 EUR) and no special, costly technology (hardware, software etc) is needed.

2 Sustainability: Mode of Conduct

2.1 Introduction PersID Mode of Conduct

Continuous existence of digital resources and their accessibility on the internet are often taken for granted. Reality is that digital resources are bound to be lost if not maintained and kept accessible. Solving this is not just a technology issue but a matter of governance.

The PersID project aimed to establish a first instance of an agreed policy to achieve guaranteed accessibility (retrievable) and persistence of digital resources (remain intact). Agreed policies are the fundament to build trust between partners and provide a trustworthy service to users.

Over time such a first instance of agreed policy – modestly called Mode of Conduct here – will evolve into a governance infrastructure that facilitates all procedures and operations required and allows partners to depend on the infrastructure and to trust each other. Trust is based on a shared view on the preferred governance approach and expectations about partners' compliance and effort to make the governance infrastructure work.

Based on the inventories and studies on current practice and user needs and on the acknowledgement that technology alone cannot guarantee accessibility and persistence, current PersID partners have formulated a Mode of Conduct. The Mode of Conduct describes a set of principles, agreements and policies that represent their shared view on the governance infrastructure and the expectations about the common effort to make it work. Participants within the PersID community agreed to apply and work according to the Mode of Conduct in their persistent identifier related processes. This Mode of Conduct is basic and practical, and only goes as far as is possible within the constraints of current development of national persistent identifier infrastructures, PersID project resources and currently established cooperation.

Partners will remain autonomous in their national approaches, policies and operations; cooperation is based on the policies in the Mode of Conduct.

2.2 PersID Mode of Conduct

- Partners agree that resources linked to a Persistent Identifier should be resolvable and accessible now and in the future. An agreed governance infrastructure must establish persistency and accessibility, thereby fulfilling the expectations of the research and cultural heritage community.
- Partners agree that a global resolution service is essential to enable (1) Persistent identification of resources (2) Persistent localisation of resources (3) Persistent access to electronic resources on an international scale.
- Partners agree that establishment of persistency and accessibility through persistent identification at a global level requires cooperation between national responsible entities (Registration Agencies).
- Partners agree that their main interests are primarily best served by the URN:NBN namespace.
- Partners agree that open communication, interaction and where possible cooperation with other persistent identifier systems is in the interest of the research and cultural heritage community.
- Partners agree that the governance infrastructure related to the persistent identification of digital resources does not include issues regarding security and authorisation nor legal matters.
- Partners will work to establish/identify national entities (Registration Agencies) in the respective countries to support and further develop policies. Until Registration Agencies are established/identified partners will provisionally act as placeholder for that role.

- Partners (in their role as Registration Agency placeholder) will enforce that (other) parties in their country that are involved in the chain of persistent identifier operations act in accordance with the agreements stated in the Mode of Conduct (inheritance principle).
- As far as governance and policies are not (yet) commonly agreed between partners, the cooperation will be based on full transparency: current national governance practice, roles and responsibilities of entities in the chain of operations will be described in detail and this information will be openly shared.
- Partners will rely on and comply to current IETF standards and will support development of (draft) future IETF standards regarding the URN:NBN persistent identifying process.
- Partners will increase the awareness of the importance to build up trustable Persistent Identifier systems in general and URN:NBN in particular and will accentuate the inseparability of Persistent Identifier Systems and Long-Term Preservation.
- Partners will actively engage in exchanges with other Persistent Identifier systems to establish a better Persistent Identifier service to the research and cultural heritage community.
- Partners agree to the aim to fulfil user expectations regarding persistency and accessibility of digital resources and therefore will:
 - ensure that registered digital resources will be persistently preserved and will be accessible; additional conditions may apply (e.g. authorisation, copyright) for actually obtaining the digital resource.
 - invest in high quality maintenance of the linking between a Persistent Identifier (PID) and resolution of resource locations, PID and metadata, and PID and provenance information.
 - offer high availability and quick responding local services that operate in a multitude of world languages and support a global resolution service.
 - only register digital resources via URN:NBN assignment if these are archived at a national archive or another trusted digital archive, that uses strategies for enabling or delegating Long-Term Preservation activities.
 - explore appropriate certification for trusted digital archives.
 - provide clarity and transparency regarding the persistency strategy of the trusted digital archive. This could include archival certification.
 - provide clarity and transparency regarding the roles and responsibilities in the chain of persistent identification processes - human interaction and technical service levels alike.
 - specify the service level of technical maintenance of the hardware and software supporting the persistent identifier processes.
 - contribute and comply to (updated) IETF URN:NBN standards and
 - not use a URN:NBN for more than one entity, and
 - not reuse a once published URN:NBN for a different digital resource.
 - specify on which conceptual level a resource will be assigned a persistent identifier (e.g. using the conceptual levels of the FRBR model as a reference).
 - proactively communicate with other partners and/or other stakeholders about
 - (planned) activities, events, and developments;
 - knowledge and experience;
 - cost and business models;
 - qualified error messages and/or restricted access messages;
 - a vocabulary with agreed terms and definitions.

The Mode of Conduct needs further development and refinement towards a full-fledged policy. In '4 Annex Work in Progress – Draft Policy' current work and thoughts on such a policy are presented.

3 Sustainability: Roadmap

The challenge for PersID is to establish a persistent identifier 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 long term preservation (LTP). Long term preservation needs a trusted and high quality persistent identifier solution with guaranteed accessibility.

The ambition of PersID is to deliver a working meta resolver infrastructure which will be able to resolve requests to local URNs but also to other persistent identifiers. This requires technical solutions, but more important: an established organization and agreed policy.

During the project many steps in the right direction were made towards:

- *Visibility*
established usage of PersIDs' trusted URN:NBN Meta Resolution Service, stakeholders are aware of PersIDs' URN-NBN activities, PersID is a credible factor in expert and strategy discussions about the future persistent identifier infrastructure
- *Know-how*
national experience has been bundled, studies on the current state of affairs have been conducted, a prototype has been developed and tested, a cost model with recommendations how to establish and run a URN resolving service has been developed, guidelines for a policy have been prepared
- *Community*
eight organisations in six European countries have worked successfully together for 16 months and connections to other stakeholders have been established e.g. Europeana; in the partner countries URN-NBN communities have been established
- *Infrastructure*
the first trusted URN:NBN Meta Resolution Service is in place at European level, serving national URN-NBN resolvers; organisations agreed to technology and policies regarding quality of persistence)
- *Standardization*
participating in an IETF URNbis working group for URN:NBN, especially in the revision of the URN:NBN namespace registration
- *Cooperation*
initiating contacts between the different persistent identifier systems

The PersID project had only limited time and limited resources at its disposal. The outcomes are only the beginning and need further development and continuation. PersID project partners agree in a Letter of Intent, to use the results from the project for future cooperation and to be aware of the risk factors that endanger the sustainability of a persistent identifier system.

3.1 Roadmap for PersID partners

(to further develop, exploit and refine the work started in the PersID initiative):

- **Operate and improve an international URN-NBN Meta Resolution Service (MRS) infrastructure, that**
 - serves current partners' national resolution services
 - is open source and based on internet standards for URN:NBN and URN related technology
 - is governed by agreed policies to guarantee trust and quality
 - is open to all organisations that agree to the policy
 - is open to identify all types of objects resulting from research or education

- is open to and aims for integration with other persistent identifier systems
- **Operate and improve the MRS as follows:**
 - secure the current trusted URN:NBN MRS by continued use and maintenance of existing hardware and software
 - establish a regular PersID partner consultation on performance, policy matters, governance structure and business model
 - contribute to the work of the IETF URNIB group, developing and renewing internet standards for a URN infrastructure
 - raise awareness for the PersID trusted URN:NBN MRS
 - outreach to potential organisations that might support or participate in PersID,
 - take part in the discussions about persistent identifier solutions
- **Arrange coordinated action to:**
 - keep the current trusted URN:NBN MRS up and running at minimal cost and effort
 - bring forward automation of components of the MRS to reduce costs by
 - more transparency about the responsibilities. Detailed information about the participating resolvers is needed such as: resolver name, base URL, policy information, governing organization, business model, link to the local resolver website and contact information. If an error occurs, the MRS notifies the maintainer of the resolver about the problem
 - enabling the end user to give feedback to the MRS,
 - if a PI does not retrieve a resource
 - if response time is slow
 - and send these information to the responsible local resolver
 - enabling system administrators to check easily the status of various parts of the MRS such as load
 - enable system administrators to get easily some basic statistical data about the usage of the Meta Resolver, e.g. about users, requests, or partners to have a valid image of the increasing demand
 - draft a proposal for a sustainable model to migrate the current trusted URN:NBN MRS from semi-production to production level
 - formalise agreed policies on trust and quality into a membership charter
 - exchange information between national organisations on developments relevant to PersID and other persistent identifier systems
 - create a clear policy on the syntax of identifiers, what characters and what semantics should be allowed

3.2 Roadmap for national organisations

(to realise adequate, global, and user friendly usage of persistent identifier services):

- Provide a reference point for people and organizations who are interested in URN:NBN and/or other persistent identifier systems, such as Handle and DOI
- Build a community and do external outreach promoting usage of existing persistent identifier solutions
- Urge the national libraries to take responsibility and play an active role to promote usage of PID's (especially URN-NBN's) and establish a national governance structure
- Engage with other projects such as APARSEN, to suggest and think through further development of a global persistent identifier infrastructure (URN-NBN, DOI or Handle)
- Invest in cooperation with Handle and DOI, especially on the policy level, to eventually be able to provide user friendly services
- Stay in touch with COOL-URI and Linked Data people – it is in the interest of the community to benefit from both 'schools' of thinking.

4 Annex 'Work in Progress' - PolicyDraft

This supplement is a first and unfinished attempt to describe the areas and issues that an agreed policy between partners should cover in detail – currently the cooperation between partners is formulated in the PersID Mode of Conduct.

Draft! Persistent Policy and Strategy for the PersID initiative and the Meta Resolution Service (MRS)

Contents

- 1. The PersID initiative**
- 2. Goals of the PersID initiative**
- 3. PersID infrastructure**
 - 3.1 Meta Resolution Service
 - 3.2 Policy
 - 3.3 Public Relations
 - 3.4 Standardization
- 4. Meta Resolution Service MRS**
- 5. Governance / Requirements for participation in the MRS**
- 6. Recommendations for a basic persistent identifier policy for participants (model for a local policy)**
 - 6.1 What can be assigned a persistent identifier?
 - 6.2 How will persistent identifiers assigned and managed? (terms and conditions for registration)
 - 6.3 What technical services are available at the local nodes?
- 7. Criteria to be classified as a trusted repository**
- 8. Suggestion for the minimum requirements for an agreed policy**

1. The PersID initiative in general

Persistent access to digital publications and other electronic resources over time is an increasingly important aspect of the scientific and cultural heritage communities. "Persistent access" to information requires an unambiguous identification of information objects through a persistent identifier.

The challenge is to establish a persistent identifier 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 long term preservation (LTP). Long term preservation needs a trusted and high quality persistent identifier solution with guaranteed accessibility.

The PersID organisations combine a large amount of experience already acquired through existing national persistent identifier solutions. Ten national organisations in eight European countries see that the time has come to co-operate with other European efforts and stakeholders to ensure reliable access to information objects.

The ambition of PersID is to deliver a working meta resolver infrastructure which will be able to resolve requests to local URNs but also to other persistent identifiers. This requires technical solutions, but more important: an established organisation and agreed policy.

The partners have chosen to use "Uniform Resource Names for National Bibliography Numbers" URN:NBN regarding this to be the most open and versatile system. URN:NBN is and will be based on IETF internet standards, it is not commercial or proprietary. It is able to accommodate existing national URN:NBN identifier schemes as well as other systems (DOI, Handle, ARK, etc.) within the overall URN scheme, using proven methods and technologies in an open and transparent way.

2. Goals of the PersID initiative

- PersID aims to enable
 - persistent identification of resources
 - persistent localisation of resources
 - persistent access to electronic resources
- PersID wants to achieve a governance infrastructure for persistent identifiers, not a technology infrastructure.
- Within this infrastructure one only can get an URN:NBN for a digital object if one has a strategy in place that takes care of the preservation of that digital object in the long run.
- This infrastructure relies on contracts with clear service level agreements between different roles in this infrastructure for persistent identifiers.
- These roles form a network of trusted repositories, resolvers, aggregators, LTP archives and registration agencies.
- A distributed URN resolver network makes sure resolvers can work independently gaining more autonomy, control, reliability and trust.
- PersID aims to ensure permanent access to electronic resources for education, research and cultural heritage communities using an open persistent identifier service infrastructure that is affordable, platform independent and non-proprietary

3. PersID infrastructure

PersID provides to the participants an infrastructure consisting of different parts:

3.1 Meta Resolution Service (MRS):

Single point of entry for accessing most of the registered digital entities regardless of the persistent identifier systems used (i.e. entities that are identified by a persistent identifier via an organization that adheres to the agreed PersID policies).

3.2 Policy

Agreed policy for the participants of the MRS with the goal to make objects durable resolvable and accessible. Advices for the partners how to build up a local trusted PI system, models for cost estimates and suggestions for a policy

3.3 Public Relations

PersID will increase the awareness of the importance to build up trustable PI systems in general and URN:NBN in particular and PersID will accentuate the inseparability of persistent identifier systems and long term preservation. To reach this goal PersID will:

- take part in conferences, meetings etc. concerning PI systems and LTP
- schedule and realise cooperation meetings
- initiate own workshops
- represent the cooperation on European and international events

3.4 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 standardising 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. Meta Resolution Service (efforts of the Resolver / advantages for participants)

- Single point of entry for several kind of PI systems regardless of the provenance
- Redirection of a PI query to the location where the item is currently stored
- Offers different alternative services for querying the URL of the full text of the digital resource, metadata, URL list, primary URL (primary: FRBR classification, first URL to be returned) and information about the identifier
- Delivers qualified error messages
- Ensures transparency by referring to contact partners being responsible for the quality of the PI
- takes care for the maintenance of the resolver
- provides and actualises the information about the participating local nodes to confirm the trustworthiness, high quality and reliability of the PI system
- pays attention to the compliance with the minimum standard for the local nodes
- guarantees high availability and security of the resolver
- fast resolution/ redirection on sub-second level

5. Governance / Requirements for participation in the MRS

PersID: The non-profit-organisation provides the social infrastructure. Other organisations are welcome to become partners also without fee, but by taking over a task.

PersID trusts in the work of the national organisation and delegates responsibility to the local resolvers, but gives advice for a common policy and takes care that the participants comply with the agreed minimum standards.

Participants, not complying with the minimum standards will be requested to do so. If necessary participants could be marked as not reliable, could be blocked or suspended.

6. Recommendations for a basic PI policy for participants (model for a local policy)

The participants declare the intent to aim in the medium term for the basic conditions for a trustable meta resolution service carried out below.

The PersID infrastructure is open for all identifier schemes. A persistent identifier (PI) used may be in principle a URN:NBN, a HANDLE, a DOI, etc. If the organisation that issued the PI is part of the PersID infrastructure the objects will be durable resolvable and accessible.

A local resolver should be able to answer the following questions:

- a) What can be assigned a persistent identifier?
- b) How will persistent identifiers assigned and managed? (terms and conditions for registration)
- c) What technical services are available at the local nodes?

6.1 What can be assigned a Persistent Identifier?

PIs may identify all kinds of things like media (books, movies, paintings, etc.), persons, organisations, places, events, research data or other concepts. The only prerequisite is that an entity is formally described in a way that allows disambiguation (i.e. delimitation of entities). We distinguish between *digital* and *non-digital* entities. *Digital entities are informationresources living on the web (a definition used in the semantic web)* that could be accessed by a generally used internet access mechanism such as a URL.

[Comment: The above is good for a local resolver. This gets difficult for cross-local-resolver resolution. E.g. an article with a DOI and has a urn:nbn in the NL and DE. (Springer article)].

[Comment: The semantic web also recognises ,non-informational' resources. Such as identifiers for people, places, etc. even a ,work' as defined in FRBR. Should PersID recognise these as well?]

The next step in the road map could be that a service will be established to give 'same_as' relationships between identifiers. The next step after that is to provide more sophisticated relations, such as FRBR: *expression_of*, *manifestation_of*, *version_of*, *sequential_part_of*, *part of*, *Child_of*, *parent_of*, etc...

A PI provides a reference to at least one (of more) URLs linked to an object, by means of which an object is addressed. A PI can also refer to several copies of the same object, accessible through multiple URLs, and can incorporate a variety of object presentation formats.

Registered entities can but need not be freely accessible. Digital entities may underlie access restrictions (e.g. usage only for authorized groups/persons or at certain places). Therefore the accessibility of resources may depend of the user's individual rights and his location (reference to Info:eu-repo/semantics/AccessRights).

A PI identifies exactly one entity, but one entity may be assigned in different PI systems (like DOI, URN, Handle and so on). It is not allowed to use a PI for more than one entity, or to reuse a published PI.

[Comment : So, if an entity has a DOI it may be given a URN:NBN.]

[Comment: What about situations that another FRBR expression, an author version, is in a repository and the publisher version is in at the publisher? Does the ,‘work’ then have the DOI?]

[Comment: It must be allowed to have several PIDs for one entity. One may discuss however if it should be allowed for a repository to give a urn:nbn to something that already has been assigned a urn:nbn. (In Germany there are policies in place to avoid double urn:nbns) to one object) We (KB Sweden) are planning to use a certain urn:nbn sub name space for our archival information packages. This means that a resource (a copy) delivered to us may have a urn:nbn assigned by the publisher (on the expression level) and when it reaches our archive it will be given one more urn:nbn for the AIP. You could argue that those two urn:s will be assigned on different “levels” in the FRBR model. You could also say that our urn:nbn is assigned to the “package” and not to the resource as such (?). It depends on how you define the word “entity”.]

And then have other named relationships to URL: like ‘from this DOI, this URL A is another expression of this URL B or even an URN:NBN B’

Even if the location (URL) of an entity changes, its PI remains the same. Once published PIs will never be deleted or become non-responsive, nor used to identify another digital entity.

Examples for digital and non-digital entities:

- Online Dissertations
- Audio files
- Online Publications
- Digitisations of printed objects, pictures etc.
- E-Journals
- Number of a E-Journal
- Articles of E-Journals
- Websites
- All printed materials
- Paintings
- Persons
- Places
- Events
- Research data
- Concepts
- ...

[Comments: Until the RFC have been accepted the approach to the non-information objects is unclear]

6.2 How will persistent identifiers assigned and managed? (terms and conditions for registration)

By fulfilling the following requirements a PI assignment for digital objects is possible:

- PIs can only be assigned for digital objects within one's own area of responsibility. *[Comment: example/explanation needed]* This applies regardless of how the object is technically published, for example by an external institution.

- the digital objects are archived at a National archive or *[Comment: does this mean that the national archive or library has to host a copy of everything that has been assigned a urn:nbn, or can it be stored at another institution accepted as a trusted?]*
- the digital objects will be administered on certified document servers (e.g. [DINI certificate](#), [Data Seal of Approval](#)) with a perspective to making them permanently available via long term archiving.
- PIs must have a metadata component

[Comment: We can recommend a minimum set : name, title, year, access rights.]

[Comment: Institutions should inform the PersID governance structure what their definition is of the minimal metadata they require for assigning a URN-NBN, and publish this on the persID website]

- PI registration at the local nodes within 24 hours after publication on the document server.
- URLs must be / are updated consequently.
- If a digital object for which a PI has already been assigned is administered as a copy at some other institution, then the PI must be transferred too, to prevent ambiguity.
- If the contents of objects can be shown to have changed, for example by a change in their MD5-fingerprint, then a new PI must be assigned.

[Comment : This implies that PI's are for digital objects on the Expression level FRBR!]

[Comment: What does this mean exactly? It cannot be stored in another resolver? According to FRBR, you can have Copies, which are duplications]

- Registered digital entities are backed up and converted/merged to agreed formats in at least one *registered, trusted long term preservation archive*. [OAIS] (e.g. a national archive or the archive of a national library).
- A workflow system have to guarantee that URLs were updated. Workflow is here being used to describe a firm arrangement between the local node and an institution with responsibility for a namespace. URL checks are to be carried out centrally by the local node as part of the PI management process and must identify inactive URLs. Institutions are then required to register new, valid URLs with local node. A deadline of X working days has to be specified.
- Relocation or duplication or deletions must be reported promptly by the institution that holds the copy that will or has been moved or copied. All *registered, trusted archives* have a compulsory responsibility in that area. They must agree to report relocations or duplications and deletions promptly and allow the national (or any other system under PersID policy *) resolver a seamless conversion to the new location.

6.3 What technical services are available at the local nodes?

- Local nodes guarantee high availability and security of the resolver
- aim for maximum response time (describe local service level)
- provide a constant link check to avoid broken links
- have a qualified error messages management

[Comment: Minimum error messages should be: 'The URN has no valid URLs, please contact xx', 'The URN urn:nbn:de:hbz:5N-19361 is not registered', 'The requested URL or URN is not valid', 'The URN ... is inactive and has the following new version:']

[Comment: At least provenance information pointing to the responsible party. Or making the trusted archive responsible for making appropriate error messages that tell the user 1. what happened to the digital object at hand , and 2. How the end-user can continue his quest for retrieving the digital object.]

- The primary function of the resolver is to redirect a PI query to the location where the item is currently stored.
- If only one URL is registered together with a PI, then the resolver will point to this URL.
- If multiple URLs are managed by one URN, then a standard resolver procedure exists to return the URL with the highest priority. This for example is the URL for a full text document in a specific format such as PDF-A, or it could be a front door or front page URL, i.e. an internet document containing a meta-description of the object.

[Comment: need a description from WP3 about the minimum technical standards for the local nodes to be a part of the PersID Resolver]

7. Criteria to be classified as a trusted repository / trusted archive

<http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf> Trusted repositories

Trusted archives

- Compliance with the Reference Model for an Open Archival Information System (OAIS)
- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological and procedural suitability
- System security
- Procedural accountability
- accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users;
- have an organizational system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility;
- demonstrate fiscal responsibility and sustainability;
- design its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it;
- establish methodologies for system evaluation that meet community expectations of trustworthiness;
- be depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly;
- have policies, practices, and performance that can be audited and measured;
- have a qualified error-messages management, describe actions that will be deployed when messages such as 'URN has no valid URLs please contact xxx'

Examples for trusted repositories / archives

- National libraries / archives
- Computer Service by a large University contracted to host the digital repository