PersID
I – Project Report

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 – I
Project Report

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

Author/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-oss-dl6

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

1 Project Report ........................................................................................................... 5
  1.1 Aim and approach ............................................................................................. 5
  1.2 URN-NBN and other systems ............................................................................ 5
  1.3 The process and trajectory ............................................................................. 5
  1.4 Achievements ................................................................................................... 6
  1.5 Acknowledgement ............................................................................................. 6

2 Executive Summary ................................................................................................. 9

3 Annex – PersID Work Plan .................................................................................... 11
  3.1 Executive Summary ........................................................................................ 11
  3.2 General overview ............................................................................................ 12
    3.2.1 Current PId: status and problem .......................................................... 12
    3.2.2 The PId Challenge .............................................................................. 12
    3.2.3 The URN-NBN PId Solution ............................................................... 12
  3.3 Urgency for a global PId solution ....................................................................... 13
    3.3.1 Trust, proprietary ownership, future services .......................................... 13
    3.3.2 The URN-NBN option .......................................................................... 13
  3.4 The URN-NBN PId project ................................................................................. 14
    3.4.1 Proposal ........................................................................................... 14
    3.4.2 Partners ........................................................................................... 14
    3.4.3 Scope............................................................................................... 14
    3.4.4 Project Goal, Objectives, Outcomes ...................................................... 15
    3.4.5 Approach .......................................................................................... 15
    3.4.6 Benefits and beneficiaries.................................................................... 16
    3.4.7 NBN issues ........................................................................................ 17
    3.4.8 Key issues ........................................................................................ 17
  3.5 Objectives in detail .......................................................................................... 18
    3.5.1 Technical objectives ........................................................................... 18
    3.5.2 Policy objectives ................................................................................ 18
    3.5.3 Communication objectives................................................................... 18
    3.5.4 Open and inclusive............................................................................. 18
  3.6 Work packages ............................................................................................... 19
    3.6.1 Overview of work packages ................................................................... 19
    3.6.2 Deliverables ...................................................................................... 20
  3.7 Annex 1b: Current organisations participating in the project.............................. 22
  3.8 Annex 1c: Stakeholders that have been informed and are interested to be
    updated........................................................................................................... 22
  3.9 Annex 2: the wide world of Persistent Identifiers ................................................. 22
1 Project Report

1.1 Aim and approach

In 2009 the PersID initiative started as a follow up to several meetings by organizations and experts interested in using the URN-NBN persistent identifier system on a global level and as a response to the interest Knowledge Exchange organizations expressed in the current state and future scenarios for usage of persistent identifiers for research publications and data. There was a need for an overview of current practice and it was felt that the options, conditions and requirements for an international persistent identifier infrastructure needed to be studied and explored.

This led to a work plan (see Annex I.a. PersID Workplan) that combined studies on the Current State and State of the Art of persistent identifier usage, User Requirements and a Roadmap for a global persistent identifier infrastructure - funded by Knowledge Exchange, and a development project to build a working prototype of a Global Resolution Service infrastructure, later simply called Meta Resolver Solution.

The combination of studies and development provided the opportunity for an iterative process: studies could inform the development, development could inform the studies and reports. Both trajectories would help to shape a roadmap: conclusions and recommendations guiding future activities towards a global persistent identifier infrastructure.

1.2 URN-NBN and other systems

The organisations that took part in the PersID initiative, the PersID partners, all use and/or have a clear preference for the URN-NBN system because of its independence of vendors (unlike Handle/DOI) and technology (based on IETF standards). URN-NBN is also seen as the system that (more than others) can agree and between trusted partners guarantee accessibility and reliable long term availability of once persistently identified objects.

This pre-PersID preference for URN-NBN has no doubt influenced the outcomes of the work presented here. On the other hand it has been clear from the very start that PersID partners wanted the Meta Resolver Solution to be open to other systems, aimed to describe conditions and requirements without the prerequisite of URN-NBN usage, and engaged in open discussions with other systems. The PersID partners accept the notion that there will be more than one PID system; as a consequence the PersID partners have been seeking open exchanges, cooperation and convergence.

At the start of the project, and especially halfway the project, comments of non-URN-NBN biased experts have been sought and evaluated, to uncover blind spots. As a result the outcomes of the PersID initiative will be of interest and valuable to a wider audience than the URN-NBN friendly community.

1.3 The process and trajectory

To plan, coordinate, discuss and evaluate the work in PersID, eight workshops were organised. These took place in at the partners’ offices in the Netherlands (Utrecht), Denmark (Copenhagen, twice), Sweden (Stockholm), Germany (Frankfurt), Finland (Helsinki) and Italy (Rome and Florence). In between several face-to-face meetings and telephone conferences took place with mixed attendance. There have also been meetings by PersID partners with interested parties, such as the Europeana project and the IETF, and presentations at seminars/conferences.
In general the progress of PersID has been satisfactory and according to schedule. Promises were kept, and usually deadlines made after the first friendly reminder. This is a remarkable achievement, given the fact that the grant provided by SURF and Knowledge Exchange only covered the ‘joint’ or ‘meta resolver’ level activities – as you can see in the project plan, most of the hours are ‘internal hours’ for work that the partners should do using their own resources – either to provide detailed information, to organise their internal adaption or to inform and engage national stakeholders.

Less frequent than planned (once in writing and once in a meeting) a status report was given to the Knowledge Exchange Interoperable Digital Repositories working group that acted as a steering group for PersID. Due to many parallel activities those who should know were informed and kept in the loop about the progress.

Building a good relation, synchronising, and aligning PersID development choices with the Europeana project caused a slight delay; absence due to illness of the project manager added to the delay. The final delivery is about four months late.

1.4 Achievements

First of all PersID uncovered and published a lot of information on current usage of persistent identifier systems. A clear picture is presented of the current practice within the PersID countries, embedded in a rich overview of worldwide persistent identifier systems.

This knowledge is used to identify user requirements (for all relevant stakeholders/actors in the persistent identifier processes). PersID offers a clear view of the gaps between the everyday national use in each country and the potential of full development of a global infrastructure. What’s even more, it offers a working prototype that proves that the potential of a reliable, independent and open persistent identifier solution is within reach, and that technology is at hand to support a system, that if run by trusted partners, can provide the services (e.g. Long Term Preservation) that research communities and cultural heritage communities need. PersID acknowledges that it inspired the IETF to start work on updating existing, overdue internet standards related to URNs.

Most important achievement of PersID is that the partners managed to get more grip on the ‘trust’ factor. For practical use, to continue the ‘trusted global resolving’ after the project ends, the Mode of Conduct was drafted and agreed on. In this document the policies and actions are described that all partners will commit to, and will consider their organisations accountable for. This may be the basis for a future policy, and as such a basis for a governance model to organise the trust hat is needed to be able to give guarantees about permanence, persistence and un-ambiguity.

Finally, as a roadmap for further development PersID has listed essential actions to be taken by the PersID partners as well as by national organisations (either PersID partner or not). The roadmap is a plea for cooperation and interaction between all ‘schools’ of thinking. When we have the user in mind our expertise should be used to offer high quality services, users should be offered freedom of choice between those services, and whatever choice they make should not present them with difficulties to interact with the others. PersID partners all agree that the increased frequency of conversations, exchanging views and exploring cooperation with Handle (EPIC) and DOI (DataCite) may be the most valuable achievement of all. Raising the profile of URN-NBN users and their views made clarified the situation that we’re all in this together.

1.5 Acknowledgement

I wish to express my gratitude to all who contributed to the PersID project. The names of meeting participants, authors, scientific manager, work-package leaders, and experts can be found in the
documents, but given the work done I am certain that in the participating organisations there are many more to be thanked. It was a pleasure working with you.

Bas Cordewener, project manager PersID
2 Executive Summary

- The need for persistent identifiers is commonly accepted, although linked data purists and/or Cool URI adepts think differently.

- There are and will be more than one persistent identifier systems operating side by side, each with its advantages, disadvantages and/or specific purposes.

- The PersID project made an inventory of current technologies used and current practice regarding the usage of persistent identifiers, and concluded
  o There's a large variety in how countries, organisations, individuals make or not make use of persistent identifiers
  o The way persistent identifier systems are run is varied as well, some are more commercially led/owned than others
  o Some systems are easily available and offer customer ready packages and services, some systems are for free and others for profit; some are supported and/or promoted by publishers
  o Guarantees for the long term regarding the (lasting) availability of the identifiers and the object the identifier is assigned to, do not depend on technology but on the trustworthiness of the governing body

- PersID partners are long-time or start-up URN-NBN users. Their preference is based on the consideration that the system is and will be based on internet standards (technology independent), openly available and not owned by an organisation or vendor, and is best equipped for long term preservation services. Extra pro: URNs are issued by national libraries - reliable organisation serving research and cultural heritage communities.

- PersID participants using URN-NBNs have a policy on the conditions and expectations regarding the persistence of the object and the assigned identifier, but it is either not formalised or out-of-date.

- 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.

- User Requirements are the basis for setting up an infrastructure. PersID identified Functional Requirements (in Must Have, Should Have and Could Have order); Non-functional Requirements (such as ‘monitored’, ‘secure’, ‘responsibility’) and Performance Constraints (such as ‘reliable’, ‘extensible’). An Architectural Model was developed.

- A subset of the User Requirement has been realised in the Working Prototype (meta resolver) and the Mode of Conduct (policy description).

- Costs of current core persistent identifier service approaches in the partner countries have been mapped for two situations: the start-up phase and the regular service situation. The resulting cost model indicates that manual interference – especially error-handling - is the biggest cost factor and should be reduced by increased automation. Costs for the persistent identifier process as such is only a small part of the total costs of preservation and guaranteed accessibility.
• Partners in PersID agreed to a mode of conduct that formulates the actions they promise to take. They agree to be transparent about the measures they take nationally, confirm they are accountable to keep the service in place, and commit to work on improved uptake and quality of the URN—NBN meta resolver solution infrastructure, including its openness to other systems.

• PersID partners drafted a Roadmap for national organisations aiming for adequate, global, and user friendly usage of persistent identifier services:
  o 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
  o Build a community and do external outreach promoting usage of existing persistent identifier solutions
  o 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
  o 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)
  o Invest in cooperation with Handle and DOI, especially on the policy level, to eventually be able to provide user friendly services
  o 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.
3 Annex – PersID Work Plan

Integrated Work plan for URN based Persistent Identifier Infrastructure project

To be provided to KE:
Awareness & Dissemination, Technology Watch Report and Sustainability Report

To be developed on behalf of SURF:
First working instance of a URB:NBN based international Resolver Service

Project: Global Resolution Service (GRS)

3.1 Executive Summary

Worldwide the need has increased to store publications and other information objects with the certainty that they will remain intact and retrievable over a long time: persistent identification (PID). The challenge is to establish a persistent identifier infrastructure that is open, global and interoperable, that is trusted and controlled by the research and cultural heritage community and serves their particular needs with regard to long term preservation.

In a number of European countries national persistent identifier solutions are in place and working satisfactory, based on the URN-NBN system, usually managed by national libraries or other national organisations that work for the research and cultural heritage community.

To successfully establish a reliable and internationally interoperable persistent identification infrastructure the main conditions are a trusting community, non-proprietary ownership of the PID system and the capability to anticipate irrelevant future services. Given these conditions for satisfactory international performance, the good experience with current national solutions, and its inclusiveness (other PID systems can be resolved as well) the URN-NBN option is the best alternative.

The partners in this Knowledge Exchange project aim to overcome the confusing variety of existing persistent identifier systems, by using the current national URN-NBN initiatives as a start for an international harmonized persistent identifier framework that serves the long term preservation needs of the research and cultural heritage communities. The project will deliver the following outcomes:

1. a working global resolution service of URN-NBN identifiers at European level. This will elevate national persistent identifier functionality for all project partners to an international level.
2. a first instance of a policy framework and an organisation to operate, maintain and further develop the global resolver service, serving the scholarly community with more uniformity. Trusted and transparent administration, usage policies and governance will be established while maintaining participants’ national autonomy.
3. documentation, information and guidelines to use the global URN-NBN service will be online available. Information and increased knowledge will be provided to recommend and promote the use of NBN based persistent identifiers with the perspective to establish a full-fledged operational global resolution service.
4. a roadmap for further development. Recommendations for increased usage, priority future services to be built upon the results of this project to expand and enhance the established global resolver infrastructure.
5. enriched yield of NBN functionality by revising the URN:NBN-related Internet standards, especially RFC 2141 (URN Syntax) and RFC 3188 (Using National Bibliography Numbers as...
Uniform Resource Names). Other URN-related standards and/or specifications will be explored and marked for further action.

The project will last one year. Substantial efforts will be made to communicate with other initiatives and stakeholders in the PId area, e.g. European projects, existing PId services, national libraries.

3.2 General overview

3.2.1 Current PId: status and problem

Worldwide the need has increased to store publications and other information objects with the certainty that they will remain intact and retrievable over (a long) time. This so called 'long term preservation' requires a method to persistently identify information objects. All over the world initiatives have been deployed to achieve this goal by setting up systems for persistent identifiers. More on persistent identifiers see Annex 2.

A variety of solutions has been developed. The existence and use of persistent identifiers is a step forward but presents new problems. Technically, organisational and policy-wise the persistent identifier solutions differ considerably. Many are not really open or interoperable, they have different business models and policy approaches. Due to this ‘link rot’ (URL’s pointing to no longer existing information) still is a common plague.

For the research and cultural heritage community this is far from ideal: long term preservation and accessibility requirements are not met if there is no interoperability, if the persistence depends on proprietary systems or on third party policies and or commercial drives that can not be trusted to offer reliable long term references.

3.2.2 The PId Challenge

Current persistent identifier systems do not offer the research and cultural heritage community a system that is open and international and is interoperable with existing persistent identifier solutions. The community is not in command of policy and management of these systems and its specific needs may or may not be served.

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. These particular needs can be illustrated by the services the community would like to build on top of the persistent identifier system, opening up possibilities for citation counts, usage statistics and enhanced (multi object) publications.

3.2.3 The URN-NBN PIdSolution

In a number of European countries national persistent identifier solutions are in place, based on the URN-NBN system. These systems have been developed to serve national long term preservation needs. The infrastructure (such as the distribution of unique URN-NBN numbers and long term preservation policies) is managed by national libraries or other national organisations that work for the research and cultural heritage community.

The national systems, that all are based on the URN-NBN approach, work satisfactory on a national level. Jointly a demonstrator has been build to explore how the national remit of these systems can be elevated to an internationally working system. The URN-NBN system is based on open, existing internet technology standards and the success of this so called Global Resolver indicates that technically there are no real problems to develop and deploy a global persistent identifier system based on URN-NBN.
What needs attention are issues like the organisation and management of an international URN-NBN persistent identifier infrastructure that develops and guards policies and a business model that is sustainable. To ensure true interoperability an effort will be made to update internet standards that define NBNs.

### 3.3 Urgency for a global PID solution

#### 3.3.1 Trust, proprietary ownership, future services

Persistent identification of information objects plays an important role within the research life cycle approach to cultural and scientific digital library applications and the scholarly community. Though on the surface the existence of parallel PID systems may not seem to be a problem, in reality it presents several issues that need to be resolved before long term preservation and persistently identifying research and cultural heritage community information objects will be a trusted, reliable process.

**Lack of Trust**

Multiple PID mechanisms and many implementers have led to little or no harmonisation on policy about what to identify, how to identify, how to resolve, how long to identify, etc. This blocks researchers and others from trusting and using PIDs, there is no coherent information and guidance and without this the development of a durable network of academic output and cultural heritage and additional services will never be achieved.

**Proprietary ownership**

Current implementers and/or providers of persistent identifiers have no clear policies about their long term commitments, it is unclear who will guarantee the long term availability of both the persistent identifier and the information object it refers to; it is not transparent how international interoperability will be sought; proprietary drivers will define future decisions and this may not serve the research and cultural heritage community needs.

**Future Services**

The current situation is too complex and fragmented to build future services on that would really benefit the research and cultural heritage communities, such as worldwide comparison of usage statistics, global citation analysis, portals, enhanced publications, metadata interoperability facilities, global semantic thesauri, detection of duplicates, etc.

In this context it is currently difficult for researchers and applications to work with NBNs because they depend on the availability of a nationally available resolver, that does (not yet) have the level of trust required, and only offers a particular, nationally defined persistence. These national NBN services are relatively unknown.

#### 3.3.2 The URN-NBN option

It is urgent to build a trusted and durable infrastructure for scholarly communication and cultural heritage that provides permanent and transparent access to both research and cultural heritage resources on the Internet. Such a durable infrastructure requires an open technology approach, a trusted framework of management and policies directed at the research and cultural heritage community, and clear information to all involved about opportunities and procedures.

This durable infrastructure can be realised building on the URN-NBN concept, in particular on the basis of:

- previous satisfactory experience with URN-NBN in national contexts,
- the successful build of a global URN-NBN resolver demonstrator,
the willingness of national representatives of research and cultural heritage communities to engage in a shared URN-NBN infrastructure,
URN-NBN being fully based on open, interoperability standards and is able to resolve and harmonise other existing persistent identifiers.

This situation opens the opportunity for a project that will
• bring immediate benefits to URN-NBN using partners,
• establish a mature international URN-NBN infrastructure,
• explore URN-NBN potential to harmonise current PIid diversity.

### 3.4 The URN-NBN PIid project

#### 3.4.1 Proposal

This proposal brings parties together that use similar persistent identifier (PIid) solutions and that agree on the need for a more uniform global PIid infrastructure.

#### 3.4.2 Partners

The project aims to demonstrate and validate a service at European level on PIids for digital resources in the cultural and scientific area, in particular a URN:NBN-based Persistent Identifier Infrastructure, has been initiated by the Knowledge Exchange (KE), a collaboration of DEFF in Denmark, DFG in Germany, JISC in the UK and SURFfoundation in the Netherlands. The KE partner organisations, and/or other national organisations in their countries take part in the project. Other participants are national libraries in Sweden and Norway and Finland as well as national cultural heritage organisations in Italy.

#### 3.4.3 Scope

The partners’ perspective is defined by the needs of the Research and Cultural heritage community with regard to long term preservation. Partners share the thought that the main challenge is not to provide a technical solution, but to use open, interoperable URN-NBN based technology in a framework of policy, organisation and communication

• that can be trusted by the research and cultural heritage communities to serve the community needs and will stimulate usage of PIids,
• that allows global long term preservation and resolving using persistent identifiers,
• 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.

The project will explore and demonstrate how the fragmented systems of PIids can be harmonised in order to increase the necessary trust and use by the end users. It will start 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 PIids issued by national libraries and other national agencies taking care of long term preservation have proven to be an effective instrument for long term preservation. This project proposal aims to be as effective on an international scale by enabling the global academic and cultural heritage institutions to maintain the links to their resources via an international URN-NBN PIid framework.
3.4.4 Project Goal, Objectives, Outcomes

Goal
The project aims to overcome the confusing variety of existing persistent identifier systems, by using the current national URN-NBN initiatives as a start for an international harmonized persistent identifier framework that serves the long term preservation needs of the research and cultural heritage communities.

Objectives
In the first year participating URN-NBN partners want to start benefiting from a global solution that resolves all current nationally issued URN-NBN PIDs; they want to explore and establish a policy and business model for a sustainable, international URN-NBN PID infrastructure; they seek to work on a solution for the generic PID problem by examining current PID systems (function, performance, reliability, imperfections); examining the policy and business models of current PID systems; explore how the URN-NBN may help harmonizing current PID systems into a global PID framework.
Objectives therefore are:
1. Provide an interoperable global resolution service of NBN identifiers at European level. The GRS project will implement and test a framework that supports reliable access and retrieval of information objects via a central resolver, which will delegate resolution to the correct national NBN (or potentially other) resolution system. Here we will lift national persistent identifier functionality for partners to an international level (this is the technical objective).
2. Enrich the yield of NBN functionality to meet the requirements of the scholarly community by providing a more uniform and transparent policy. Here we will explore trusted and transparent administrative, governance and usage policies while maintaining participants’ autonomy (this is the policy objective).
3. Provide information on, increase knowledge and understanding, gain experience and promote the use of NBN based PIs. Analyse and recommend on technology, policy and service development issues (e.g. availability, scalability, security, maintainability, responsibility, financial) with the perspective to establish a full-fledged operational global resolution service (this is the communication objective).
4. Revise URN:NBN-related Internet standards, especially RFC 2141 (URN Syntax) and RFC 3188 (Using National Bibliography Numbers as Uniform Resource Names). Investigate the need for revising other URN-related standards such as namespace registration for ISBN (RFC 3187). Consider enhancement of the category of these RFCs (Standards track / Informational /). This work should be carried out in an IETF Working Group, open also for experts from outside the proposed project.

Outcomes
In the first year the project outcomes are the following:
1. a working global URN-NBN resolver
2. policy and organisation to operate, maintain and further develop the global resolver service
3. documentation, information and guidelines to use the global URN-NBN service (website)
4. revised RFCs
5. roadmap for further development, increased usage, priority future services.

3.4.5 Approach

Build on positive experience
A good start for international harmonising of PID solutions can be made by examining the urn:nbn (NBN) namespace for challenges and opportunities. The participating national NBN implementers, national libraries and/or national agencies responsible for long term preservation of cultural heritage and research information, have much experience and sound expertise with regard to URN-NBN PID that will be used for a global solution. This applies to technical issues, to policy issues and governance models, and to information and user support.
Emphasis on policy and organisation
The national implementations are tailored to different national situations with different policies and laws. Joint development of policies and design of a management organisation is the core element in the project. This will ensure a stable and reliable infrastructure that meets commonly agreed user requirements, allowing partners to have their own local policy in place. Existing practices will inform partners to define a business plan that supports a sustainable worldwide URN-NBN persistent identifier infrastructure.

Non proprietary solution
In order to avoid dependency on proprietary solutions, the global resolver policies and management organisation will be owned and controlled by the cultural heritage and research community members that participate in the project. The global resolver solution will develop common guidelines and criteria needed for an international transparent PID infrastructure.

Long term perspective
The one year project has four deliverables defined (see 4.2 under Outcomes) but global long term preservation based on URN-NBN persistent identifiers will not be finished in one year. Therefore the project will also issue a roadmap towards further development, increased trust and participation of stakeholders and future services that will support the communities needs, such as comparison of usage statistics, global citation analysis, portals, enhanced publications (relations between objects), metadata interoperability facilities, global semantic thesauri, detection of duplicates, etc. Recommendations will also address the (further) harmonisation with other PID systems.

3.4.6 Benefits and beneficiaries

International functionality
The NBN project will have a direct beneficial impact on the national URN-NBN communities. The national technical, policy and communication approaches will be enhanced and functionality of the URN-NBN system will be available at an international level.

Direct benefits for user community
The project will provide increased functionality to the listed target audiences:

- **Those who assign PIDs (e.g. content producers, authors, or organisations operating on their behalf)**
  - This community will be able to make use of a globally recognized and harmonised URN-NBN system for reliable long term preservation.
- **Those who use PIDs for citation (e.g. content users, readers or organisations operating on their behalf)**
  - This community will be able to refer to content/publications (e.g. for citations) in an easy and unambiguous way by using the URN-NBN PIDs, knowing the citations will be globally recognised, unambiguous and available.
- **Those who surf to information via a URN-NBN (e.g. web users, scholars)**
  - This community will find URN-NBNs reliable actionable web-links that directs them to the expected information objects and will discover more services based on URN-NBN PIDs.

Direct benefits for the URN-NBN management organisations
The project will provide essential guidance for future development of a PID infrastructure:

- **National IT governance organisations**
  - will be able to discuss at a much higher level of expertise about necessary global solutions concerning permanent accessibility and traceability of information objects on the basis of international experience with a concrete NBN infrastructure. Some national responsibilities and/or procedures may benefit from economy of scale
Other benefits
Outside the URN-NBN domain this project will result in knowledge and experience to help harmonising URN-NBN with other PID systems, which in turn will lead to a worldwide solution for long term preservation, a benefit to all user communities.

Overall gain
This project will help move issues that have been a problem for several years. Upgrading and integration to an international level has been talked about a lot but got smothered in discussions about definitions and lack of targeted resources. The impulse of Knowledge Exchange having an interest may help achieve a global resolver solution.

Overall, practical experience will build trust and confidence. The collaborative effort will establish a more trustable PID infrastructure that can provide unambiguous and permanent accessibility in a transparent way. It will be governed and managed collaboratively on an international level, based on agreed policies aimed at maintenance, service provision and wide usage. National organisations will successfully cooperate and develop a shared infrastructure while maintaining their local implementations.

3.4.7 NBN issues
As mentioned it is essential to have a clear policy on the use and the services for PIDs. This is where the NBN implementers themselves have a backlog: every implementation has its own policy and services. Therefore users cannot expect the same service from a NBN from Norway and one from the Netherlands. A transparent usage policy, interoperable services and a shared communication approach are essential in order to have researchers trust the PIDs from NBN.

It is important to acknowledge the fact that all existing NBN solutions have been designed to fulfill and conform to local policy and laws and have been successfully integrated into local systems. Top-down policy to update these will raise resistance and should be avoided at all cost.

Global uniqueness and persistence of NBNs is only achieved via their usage as URNs. Therefore it is essential that the URN system itself is solid. Unfortunately the Internet standards (RFCs) which specify the URN syntax, URN-based services and the way in which URNs are resolved in the Internet are out of date and require revision. This revision must target both the content and the status of these standards.

3.4.8 Key issues
In this federated framework of PID management and resolution services, key issues are to achieve trustworthiness, manage mutual dependability, and guarantee longevity of services, as well as establishing institutions maintaining them. The listed issues clearly indicate that we are mostly dealing with challenges at trust, managerial and policy level. Therefore the technical efforts will be modest, and the work on organisation, management and governance as well as promoting effective use will be significant.

The main focus will be on maintaining the autonomy of all parties to use their own metadata schemata and follow their own interests regarding the strategies to facilitate the circulation of electronic documents. At the same time the compatibility with existing techniques and protocols (policies) is considered to be crucial to fully satisfy the requirements of the international community. Currently used technology and protocols should seamlessly fit in the NBN based PID infrastructure.
3.5 Objectives in detail

3.5.1 Technical objectives

A global resolver is required to create a unified central entrance point where both researchers and applications can resolve NBNs. This project will implement a prototype that will:

- Cover primarily URN:NBN but will also try to encompass other URN namespaces such as URN:ISBN and URN:ISSN.
- Interoperability with other PI solutions such as Handle / DOI or ARK will be investigated, given that a partner for such testing is available and that integrating these services will not require significant efforts.
- Show how an operational global resolver can be created.
- Prove that a global resolver for different NBN solutions is feasible.
- Deal with issues like availability, scalability, security, maintainability by either integrating these within the prototype or by providing recommendations on them (if these are unfeasible within the scope of this project).
- Identify other technical issues that need to be considered for a (next) operational system.
- Issue a report that discusses the experiences, identifies issues to consider and gives recommendations for a production-level global resolution system.

3.5.2 Policy objectives

- Scope identification issues (content, persons, projects; scope of the project is on publications and research data).
- Issue guidelines on persistent identification.
- Validate RFC 2141 and RFC 3188 and update these URN standards/specifications if necessary.
- Implement trusted repositories persistence of the whole identifier system (guidelines for services provided by PID systems, updating and storing, connecting international long term preserving initiatives).
- Define the roles and responsibilities in the organisational model, regarding the national resolver and registration agencies, the global resolver, Long Term Sustainability of critical components and present a business model.
- Strengthen the status and awareness of the URN standards and services.

3.5.3 Communication objectives

- Organise and establish an effective communication and promotion regarding the project, the Global Resolver Service, NBNs, the urgency of persistent identification, the potential services that could be built on such an infrastructure.
- Achieve increased awareness, knowledge, involvement and usage regarding NBN within all relevant stakeholder groups.

3.5.4 Open and inclusive

The project will:

- make use of currently available technology based on the principles of URN and NBN;
- respect currently existing national NBN systems (a distributed approach);
- engage with all stakeholder groups (content providers, content managers, content users);
- engage in an open dialogue with existing projects that address PI issues;
• develop solutions that are capable to incorporate or work together with other existing PI solutions (e.g. DOI, ARK, Handle);
• provide lessons and unveil yet unknown issues regarding a full fledged operational global NBN based PI infrastructure.

### 3.6 Work packages

#### 3.6.1 Overview of work packages

The project will work according to the three focus areas 'technical', 'policy' and 'communication'. The work packages are defined as follows:

<table>
<thead>
<tr>
<th>Work package No</th>
<th>Work package title</th>
<th>Type of activity</th>
<th>Lead</th>
<th>Start month</th>
<th>End month</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP1</td>
<td>Management</td>
<td>Coordination</td>
<td>BC</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>WP2</td>
<td>Communication</td>
<td>Communication/Dissemination</td>
<td>BC</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>WP3</td>
<td>user Requirements</td>
<td>Definition</td>
<td>LS</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>WP4</td>
<td>Implementation</td>
<td>Development</td>
<td>RP</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>WP5</td>
<td>Sustainability</td>
<td>Documentation</td>
<td>CS</td>
<td>7</td>
<td>12</td>
</tr>
</tbody>
</table>
3.6.2 Deliverables

The aspired end result of this project is an organisation capable of delivering a sustainable registration agency for URN:NBN and a Global Resolution service. Four major deliverables will be produced.

- D1 Resolver service – including Technology Watch report for KE
- D2 General Awareness on PId and Information on how to use the Resolver Service
- D3 Organisation & policy (incl. RFCs) – incl. organisational part of Sustainability report for KE
- D4 Future recommendations – including business/cost model part of Sustainability report and Roadmap for KE

(See Annex 1. for person names corresponding with initials below)

<table>
<thead>
<tr>
<th>Deliverable nr.</th>
<th>D1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Resolver Service including Technology Watch Report</td>
</tr>
<tr>
<td>Delivery date</td>
<td>Month 12</td>
</tr>
<tr>
<td>Coordinator</td>
<td>WP3 and WP4: RP (lead), LS</td>
</tr>
<tr>
<td>Short description</td>
<td>A working Global Harmonisation Resolution Service. This service is able to resolve Identifiers from existing URN:NBN resolvers which are national solutions from the partners, plus! Other popular Identifier schemes, like DOI, ARK, HANDLE... etc. that is used in science and cultural heritage.</td>
</tr>
<tr>
<td>Deliverable and contributing Milestones</td>
<td>M3.1, M3.2, M3.3 (the above are the components of the Technology Watch Report), M4.1 (is the organisational part of the Sustainability Report), M4.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deliverable nr.</th>
<th>D2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Information package and facility</td>
</tr>
<tr>
<td>Delivery date</td>
<td>Month 12</td>
</tr>
<tr>
<td>Coordinator</td>
<td>WP2 = BC (lead), AP, SD, KE office</td>
</tr>
<tr>
<td>Short description</td>
<td>D2 = Communication and dissemination about the particular KE-PId global resolver project, about URN-NBN developments and related PId activities and development. A full information service for the community of existing or potential users (research and cultural heritage) or others holding a stake in URN;NBN based persistent identifier solutions (institutions, libraries, experts, developers, managers)</td>
</tr>
<tr>
<td>Deliverable and contributing Milestones</td>
<td>M2.1, M2.2</td>
</tr>
</tbody>
</table>
### Deliverable nr. D3

**Title**  
Organisation & policy, including RFCs and org part of Sustainability report

**Delivery date**  
Month 12

**Coordinator**  
WP3 and WP4 : LS (lead), RP

**Short description**  
D3 = “Registration agency in place” Established organisational infrastructure and policies for basic (minimal) global resolver service, capable to be expanded into a managerial and economically sustainable, international infrastructure and organisation to manage and maintain an extended global resolver service in support of Long Term Preservation of Research and Cultural Heritage content. Rewritten RFC 2141 and 3188.

**Deliverable and contributing Milestones**

- M3.1
- M3.2
- M3.3
- M4.1 (part of Sustainability report)
- M4.2
- M4.3

### Deliverable nr. D4

**Title**  
Future roadmap

**Delivery date**  
Month 12

**Coordinator**  
Mainly the outcome of WP5 – CS and WP2 – BC

**Short description**  
D4 = report of collated information, guidelines and recommendations on further realisation of required infrastructure services, roadmap document, future services, and sustainable growth

**Deliverable and contributing Milestones**

- M4.1
- M4.2
- M5.1
- M5.2

### Milestones contribution to Deliverables

<table>
<thead>
<tr>
<th>Milestone n°</th>
<th>Title</th>
<th>Part of WP/ contrib. to Deliverable</th>
<th>Relates to KE output</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1.1</td>
<td>Brief description of quality approach</td>
<td>WP1 Proj Coord</td>
<td></td>
</tr>
<tr>
<td>M1.2</td>
<td>IPR statement</td>
<td>WP1 Proj Coord</td>
<td></td>
</tr>
<tr>
<td>M2.1</td>
<td>Promotional material and web architecture, snapshot</td>
<td>D2 Awareness &amp; Comm</td>
<td></td>
</tr>
<tr>
<td>M2.2</td>
<td>Agenda of events, community building strategy and results</td>
<td>D2 Awareness &amp; Comm</td>
<td></td>
</tr>
<tr>
<td>M3.1</td>
<td>Current state on PId implementations, case studies and technologies</td>
<td>D1/D3 State of the Art /Tech Watch rep</td>
<td></td>
</tr>
<tr>
<td>M3.2</td>
<td>User requirements definition for a PId NBN service</td>
<td>D1/D3 State of the Art /Tech Watch report</td>
<td></td>
</tr>
<tr>
<td>M3.3</td>
<td>Minimum specification of a European resolver service</td>
<td>D1/D3 State of the Art /Tech Watch report</td>
<td></td>
</tr>
<tr>
<td>M4.1</td>
<td>Create preliminary blue print for roles and responsibilities</td>
<td>D1/D3/D4 Org part of Sust report</td>
<td></td>
</tr>
<tr>
<td>M4.2</td>
<td>Document the software and infrastructure</td>
<td>D1/D3/D4</td>
<td></td>
</tr>
</tbody>
</table>
3.7 Annex 1b: Current organisations participating in the project

- Knowledge Exchange  www.knowledge-exchange.info
- DEFF    www.deff.dk
- SURF Netherlands  www.surffoundation.nl
- DANS Netherlands  www.dans.knaw.nl
- CNR Italy  www.cnr.it
- FDR Italy    www.rinascimento-digitale.it
- National Library Finland  www.nationallibrary.fi
- National Library Germany  www.d-nb.de
- National Library Sweden  www.kb.se

3.8 Annex 1c: Stakeholders that have been informed and are interested to be updated

- Europeana project / European Digital Library
- British Library
- Dutch Royal Library
- National Library Norway
- Ministry of Education, Flanders, Belgium
- Coalition for Networked Information USA
- and others

3.9 Annex 2: the wide world of Persistent Identifiers

- A link to information on persistent identifiers by the Digital Curation Centre (UK): http://www.dcc.ac.uk/resource/briefing-papers/persistent-identifiers/
- Below follows an overview of current PID technologies and approaches, documented by Maurizio Lunghi

Persistent URL (PURL)

PURLs were developed by the Online Computer Library Center, Inc.(OCLC). Functionally, a PURL is a URL that, instead of pointing directly to the location of an Internet resource, points to an intermediate resolution service. The latter redirects the request to the actual location of the resource by using the standard web servers capabilities. http://purl.org
Uniform Resource Name (URN)

The purpose of a Uniform Resource Name is to provide a globally unique, persistent, location-independent resource identifier which can be used for identification, for access to resource characteristics or for access to the resource itself. The URN specification is part of the IETF family of specifications encompassed by the Uniform Resource Identifier (URI) framework. This framework also includes URLs, which specify both a protocol and a location in order to give access to resources on the Web. IANA (Internet Assigned Numbers Authority) is the registration authority for URN namespaces. URNs are designed to enable heterogeneous namespaces mapping (which share URN the properties) onto a URN-space, and therefore enable the reuse of well-known identifiers.

Unlike URLs, URNs are not directly actionable (browsers generally do not know what to do with a URN) because they have no associated global infrastructure that enables resolution (such as the DNS supporting URL). Although several implementations have been made, each proposing its own means for resolution through the use of plug-ins or proxy servers, an infrastructure that enables large scale resolution has not been implemented.

As an example in the library domain, the NBN PI schema follows the URN standard. The NBN namespace, as a Namespace Identifier (NID), has been registered and adopted by the Nordic Metadata Projects but is being separately implemented by individual systems with no reference implementation which enable the coordination of information sources. In fact, several national libraries have developed their own NBN systems for national and international research projects; several implementations are currently in use, each with different metadata descriptions or granularity levels.

Handle system

The Handle System was developed by the Corporation for National Research Initiatives (CNRI). The Handle System is a general-purpose global name service that allows secured name resolution and administration over networks such as the Internet. The Handle System manages “handles”, which are unique names for digital objects and other Internet resources. A naming authority is authorised to create and maintain handles; the identifiers must be unique to that authority but have no prescribed syntax.

Digital Object Identifier (DOI)

DOIs are names assigned to objects with intellectual properties such as electronic journal articles, images, learning objects, eboSKO, and any kind of content. They are used to provide current information, including where resources (or information about them) can be found on the Internet. The DOI is therefore an actionable identifier and its use will immediately enable resolution over the Internet. The DOI identifier itself is part of the DOI system that provides a framework for managing intellectual content, for linking customers to content providers, facilitating electronic commerce, and enabling automated copyright management for all types of media. The DOI system is an application of the Handle System; it also adds to the latter an approach based on additional metadata formats and interoperability standards, such as the Online Information Exchange (ONIX) metadata format, policies, procedures, business models, and application tools.
Archival Resource Key (ARK)

The ARK schema is another proposal intended to facilitate the persistent naming and retrieval of information objects. A founding principle of the ARK is that persistence is purely a matter of service and is neither inherent in an object nor conferred on it by a particular naming syntax. ARK aims at gaining acceptance by providing peculiar functionalities, such as the capability of separating the univocal identifier assigned to a resource from the (potentially multiple) addresses that may act as a proxy to the final resource. An ARK, however, provides extra services above and beyond that of an ordinary URL. Instead of connecting to one thing, an ARK should connect to three things: the object itself, metadata, and a maintenance commitment from the current server. The resolution service returns a brief machine- and human-readable metadata record in an Electronic Resource Citations (ERCs) format.

http://www.cdlib.org/inside/diglib/ark

Info URI

The Info URI schema was developed within the library and publishing communities to expedite the referencing of information assets that have identifiers in public namespaces but have no representation within the URI allocation. The “info” URI scheme exists primarily for identification purposes. Implementations must not assume that an "info" URI can be dereferenced for obtaining a digital representation of the resource identified by the URI. It is simply an identifier with no associated resolution or metadata, although some of the registered schemes may provide this functionality.

http://info-uri.info

OASIS Extensible Resource Identifier (XRI)

The purpose of XRI is to define a URI scheme and a corresponding URN namespace for distributed directory services that enable the identification of resources (including people and organizations) and the sharing of data across domains, enterprises and applications.

http://www.oasis-open.org/committees/xri

Library of Congress Control Number (LCCN) Permalink

The Library of Congress has developed the “LCCN Permalink” a new persistent URL service for creating links to bibliographic records in the Library of Congress Online Catalogue using the LCCN. It can be used to cite items from the Library’s collection in bibliographies, reference guides, emails, blogs, databases, Web pages, etc. This category of permalinks is referenced by bibliographic records in the LC Online Catalogue in MArchine-Readable Cataloguing (MARCXML), Metadata Object Description Schema (MODS), and Dublin Core formats.

http://lccn.loc.gov

Enhancement of Persistent Identifier Services - Comprehensive Method for Unequivocal Resource Identification (EPICUR)

A notable experience that tries to address this challenge is the EPICUR Project at the German National Library. It is linked to other national registries as well as to the commercial DOI platform. The resolution service is able to resolve the NBNs generated by National Libraries linked to the DOI namespace.

http://www.persistent-identifier.de
<table>
<thead>
<tr>
<th><strong>NBN Italy</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>In Italy, a consortium led by Fondazione Rinascimento Digitale is developing a novel persistent identifier architecture, based on the NBN namespace, with additional features (e.g., for resolution and robustness) and solutions that recall the DNS architecture. They have designed a distributed system, according to which registered institutions can generate and resolve the NBN identifiers associated with their sub-namespace. Institutions can also resolve NBNs generated by others linked to the infrastructure, as well as the DOI namespace, through communication with the central node. The open source code is going to be reused by other countries. An Internet Draft proposal has come out from this project and the research will continue in GRS.</td>
</tr>
<tr>
<td><a href="http://www.rinascimento-digitale.it/indexEN.php?SEZ=531">http://www.rinascimento-digitale.it/indexEN.php?SEZ=531</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Name to Thing (N2T)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>N2T proposed by John A. Kunze at the California Digital Library (CDL), tries to handle the integration of different naming techniques and tackle the challenge of PI longevity and interoperability through the definition of a single consortium. This consortium manages a global access point redirecting requests to individual PI resolution services.</td>
</tr>
<tr>
<td><a href="http://www.n2t.info">http://www.n2t.info</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Life Science Identifiers (LSIDs)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>A LSID is a persistent, location-independent resource identifier for uniquely naming biologically significant resources on the Web. The LSID concept proposes an approach to naming and identifying data resources stored in multiple, distributed data stores that tries to overcome the limitations of naming schemata used today. By defining a simple, common way to identify and access biologically significant data, independently of whether data is stored in files, relational databases, in applications, or public data sources, LSID provides a naming standard underpinning for wide-area science and interoperability.</td>
</tr>
<tr>
<td><a href="http://lsids.sourceforge.net/">http://lsids.sourceforge.net/</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Network of Expertise in Long-Term Storage of Digital Resources (NESTOR)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>This is a cooperative project (funded by the BMBF, German Ministry of Education and Research) involving a consortium of libraries, archives and museums as well as leading experts, who form a network of expertise in long-term preservation and long-term availability of digital resources. In particular, the standards working group of the German NESTOR initiative has undertaken a project for the DIN (German Institute for Standardization) in order to explore the real long term validity of identifiers, what technical and organizational conditions are necessary for persistent identifiers to become interoperable with each other, and how to extend the notion of trustworthiness in the area of repositories and in the persistent identifiers domain.</td>
</tr>
<tr>
<td><a href="http://www.langzeitarchivierung.de/index.php?newlang=eng">www.langzeitarchivierung.de/index.php?newlang=eng</a></td>
</tr>
</tbody>
</table>
**Persistent Identifier Linking Infrastructure (PILIN)**

Emphasis in the PILIN Project is put on building an identifier management infrastructure based on a technology (the Handle System) that is now under development through the auspices of CNRI in order to underpin sustainable global identifier infrastructure. The project aims to take advantage of existing governance and consultative mechanisms within the ARROW (The Australian Research Repositories Online to the World) environment for ensuring relevant and sustainable outcomes and an optimal return on investment. The project will be run in partnership between ARROW and the University of Southern Queensland to develop a technology-independent model for persistent identifiers and services. This model will extend the lifetime of the infrastructure by allowing services to be mapped to new technologies over time. It will also contribute to the international understanding of the requirements of PI infrastructures.  

http://www.arrow.edu.au/PILIN.php