Date submitted: 14/08/2009

Persistent Identifier Distributed System for Digital Libraries

Emanuele Bellini[*], Chiara Cirinnà[*], Maurizio Lancia[**], Maurizio Lunghi[*], Roberto Puccinelli[**], Massimiliano Saccone[**], Brunella Sebastiani[**], Marco Spasiano[**]

[*]Fondazione Rinascimento Digitale
[**]Consiglio Nazionale delle Ricerche

Meeting: **193. Information Technology**

WORLD LIBRARY AND INFORMATION CONGRESS: 75TH IFLA GENERAL CONFERENCE AND COUNCIL
23-27 August 2009, Milan, Italy
http://www.ifla.org/annual-conference/ifla75/index.htm

Abstract

In this paper we present an Italian initiative, involving relevant research institutions and national libraries, aimed at implementing an NBN Persistent Identifiers (PI) infrastructure based on a novel hardware/software architecture. We describe a distributed and hierarchical approach for the management of an NBN namespace and illustrate assignment policies and identifier resolution strategies based on request forwarding mechanisms. Starting from the core motivations for the assignment of "persistent identifiers" to digital objects, we draw a state of art in PI technologies, standards and initiatives, and illustrate other NBN implementations. We then describe structure and goals of our initiative and illustrate the features already implemented in our system and the results of our testing activities. Finally, we draw some conclusions and point out the future directions of our work.

1. Introduction

Stable and certified reference of Internet resources is crucial for digital library applications, not only to identify a resource in a trustable and certified way, but also to guarantee continuous access to it over time. Current initiatives like the European Digital Library (EDL) [1] and Europeana [2], clearly show the need for a certified and stable digital resource reference mechanism in the cultural and scientific domains. The lack of confidence in digital resource reliability hinders the use of the Digital Library as a platform for preservation, research, citation and dissemination of digital contents [15]. A trustworthy solution is to associate to any digital resource of interest a Persistent Identifier (PI) that certifies its authenticity and ensures its accessibility. Actually some technological proposals are available [24], but the current scenario shows that we can't expect/impose a unique PI technology or only one central registry for the entire world. Moreover, different user communities do not commonly agree about the granularity of what an identifier should point to.

In the Library domain the National Bibliography Number (NBN – RFC3188) has been defined and is currently promoted by the CENL. This standard identifier format assumes that the national libraries are responsible for the national name registers. The first implementations of NBN registers in Europe are available at the German and Swedish national libraries.

In Italy we are currently developing a novel NBN architecture with a strong participation from the scientific community, leaded by the National Research Council (CNR) through its Central Library and ITC Service. We have designed a hierarchical distributed system, similar to the DNS[1], in order to overcome the criticalities of

[1] DNS (Domain Name System) is the infrastructure that resolves server names, such as www.ifla09.it, into IP addresses, such as 150.146.1.5, on the Internet.

a centralised system and to reduce the high management costs implied by a unique resolution service. Our approach implies a central node responsible for the NBN:IT top-level Italian domain, and lower-level nodes each responsible for managing one of the Italian sub-domains (NBN:IT:UR, NBN:IT:UR:CNR, NBN:IT:FRD, etc.). The number of levels within this hierarchy is virtually unlimited. Only the nodes at the lowest level harvest metadata from the actual repositories and create NBN identifiers. The upper level nodes just harvest new NBN records from their child nodes and store them within their databases. In this way each node keeps all the NBN records belonging to its sub-domain. It is easy to see that within this architecture the responsibility for name creation/resolution is distributed and information about persistent identifiers is replicated in multiple sites, thus providing the necessary redundancy and resilience for implementing a reliable service.

Before describing our system in detail, we will provide in the following sections an overview of available persistent identifiers technologies.

2. Persistent identifier standards

The association of a Persistent Identifier (PI) to a digital resource can be used to certify its content authenticity, provenance, managing rights, and to provide an actual locator. The only guarantee of the actual persistence of identifier systems is the commitment shown by the organizations that assign, manage, and resolve the identifiers [25], [26].

At present some technological solutions are available but no general agreement has been reached among the different user communities. We provide in the following a brief description for the most widely diffused ones. Only the NBN [3] standard will be described in details in the next section.

The Document Object Identifier system (DOI [11]) is a business-oriented solution widely adopted by the publishing industry, which provides administrative tools and a Digital Right Management System (DRM).

Archival Resource Key (ARK [10]) is an URL-based persistent identification standard, which provides peculiar functionalities that are not featured by the other PI schemata, e.g., 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.

The Handle System ([12], [26], [27]) is a technology specification for assigning, managing, and resolving persistent identifiers for digital objects and other resources on the Internet. The protocols specified enable a distributed computer system to store identifiers (names, or handles) of digital resources and resolve those handles into the information necessary to locate, access, and otherwise make use of the resources. That information can be changed as needed to reflect the current state and/or location of the identified resource without changing the handle.

Finally, the Persistent URL (PURL [13]) is simply a redirect-table of URLs and it's up to the system-manager to implement policies for authenticity, rights, trustability, while the Library of Congress Control Number (LCCN [14]) is the a persistent identifier system with an associated permanent URL service (the LCCN permanent service), which is similar to PURL but with a reliable policy regarding identifier trustability and stability.

This overview shows that it is not viable to impose a unique PI technology and that the success of the solution is related to the credibility of the institution that promotes it. Moreover the granularity of the objects that the persistent identifiers need to be assigned to is widely different in each user application sector.

3. NBN overview

The National Bibliographic Number (NBN) [3] is a URN namespace under the responsibility of National Libraries. The NBN namespace, as a Namespace Identifier (NID), has been registered and adopted by the Nordic Metadata Projects upon request of the CDNL and CENL. 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. Moreover, each URN name-domain is isolated from other systems and, in particular, the resolution service is specific (and different) for each domain.

Each National Library uses its own NBN string independently and separately implemented by individual systems, with no coordination with other national libraries and no commonly agreed formats. 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.

Examples are the DIVA project [16], EPICUR [18], and ARK at National Library of France [17].

There are some important initiatives at European level like the TEL project that it is in the process of implementing a unique system based on NBN namespace within the European Digital Library (EDL). The

adoption of NBN identifiers is needed for implementing the 'National Libraries Resolver Discovery Service' as described in the CENL Task Force on Persistent Identifiers report [19].

In our opinion NBN is a credible candidate technology for an international and open persistent identifier infrastructure, mainly because it is based on an open standard and supports the distribution of the responsibility for the different sub-namespaces, thus allowing the single institutions to keep control over the persistent identifiers assigned to their resources.

4. The NBN initiative in Italy

The project for the development of an Italian NBN register/resolver started in 2007 as a collaboration between "Fondazione Rinascimento Digitale" (FRD), the National Library in Florence (BNCF), the University of Milan (UNIMI) and "Consorzio Interuniversitario Lombardo per l'elaborazione automatica" (CILEA). After one year of work a first prototype was released demonstrating the viability of the hierarchical approach. The prototype leveraged some features of DSpace and Ark and provided a basic PHP web interface for library operators and final users. The hierarchy was limited to a maximum of two levels.

The second and current phase of the Italian NBN initiative is based on a different partnership involving Agenzia Spaziale Italiana (ASI), Consiglio Nazionale delle Ricerche (CNR), Biblioteca Nazionale Centrale di Firenze (BNCF), Biblioteca Nazionale Centrale di Roma (BNCR), Istituto Centrale per il Catalogo Unico (ICCU), Fondazione Rinascimento Digitale (FRD) and Università di Milano (UniMi). At the beginning of 2009 the Italian National Research Council (CNR) developed a second prototype based on Java Enterprise technologies, which eliminated the need for DSpace and Ark and the two-level limit and introduced new features.

Future objectives have been defined in order to extend functionality and integrate the system within an international infrastructure. To this end the Italian group is currently establishing international collaborations.

In the following we provide a description of objectives, governing structure and licensing policy defined for the Italian initiative.

Objectives

The project aims at:

- creating a national stable, trustable and certified register of digital objects to be adopted by cultural and educational institutions;
- allowing an easier and wider access to the digital resources produced by Italian cultural institutions, including material digitised or not yet published;
- encouraging the adoption of long term preservation policies by making service costs and responsibilities more sustainable, while preserving the institutional workflow of digital publishing procedures;
- implementing a new service based on URN, similar to other national systems but with a more advanced architecture in order to achieve distribution of responsibility for name management;
- extending as much as possible the adoption of the NBN technology and the user network in Italy;
- developing an inter-domain resolution service (e.g., NBN Italy and NBN Germany, or NBN Italy and DOI) with a common meta-data format and a user-friendly interface (pre-condition for global resolver);
- creating some redundant mechanisms both for duplication of name-registers and in some cases also for the digital resources themselves;
- overcoming the limitation imposed by a centralised system and distributing the high management costs implied by a unique resolution service, while preserving the authoritative control.

Governing board

In order to define organization and policies for the Italian register, a governing board has been established, where BNCF, BNCR, CNR, FRD, ICCU are represented. The governing board defines the top-level structure of the Italian NBN domain hierarchy and the policies for overall infrastructure management, sub-domain creation/removal and PI assignment.

Software licensing policy

The project partners are currently working to define a software license that is appropriate to pursue their objectives. In particular, in order to encourage the adoption of our software by other national registers and to foster the creation of a supporting community, an open source license would be a facilitating factor.

5. The distributed architecture approach

The proposed architecture (see Figure 1), starting from [22], [23] and taking into account the URN standard requirements [20], [21], introduces some elements of flexibility and additional features. At the highest level there is a root node, which is responsible for the top-level domain (IT in our case). The root node delegates the responsibility for the different second-level domains (e.g.: IT:UR, IT:FRD, etc.) to second-level naming authorities. Sub-domain responsibility can be further delegated using a virtually unlimited number of sub-levels (eg.: IT:UR:CNR, IT:UR:UNIMI, etc.). At the bottom of this hierarchy there are the leaf nodes, which are the only ones that harvest publication metadata from the actual repositories and assign unique identifiers to digital objects.

Each agency adheres to the policy defined by the parent node and consistently defines the policies its child nodes must adhere to.

It is easy to see that this hierarchical multi-level distributed approach implies that the responsibility of PI generation and resolution can be recursively delegated to lower level sub-naming authorities, each managing a portion of the domain name space. Given the similarity of the addressed problems, some ideas have been borrowed from the DNS service.

Within our architecture each node harvests PI information from its child nodes and is able to directly resolve all identifiers belonging to its domain and sub-domains. Besides, it can query other nodes to resolve NBN identifiers not belonging to its domain. This implies that every node can resolve every NBN item generated within the NBN:IT sub-namespace, either by looking up its own tables or by querying other nodes. In the latter case the query result is cached locally in order to speed up subsequent interrogations regarding the same identifier.

This redundancy of service access points and information storage locations increases the reliability of the whole infrastructure by eliminating single points of failure. Besides, reliability increases as the number of joining institutions grows up.

In our opinion a distributed architecture also increases scalability and performance, while maintaining unaltered the publishing workflows defined for the different repositories.

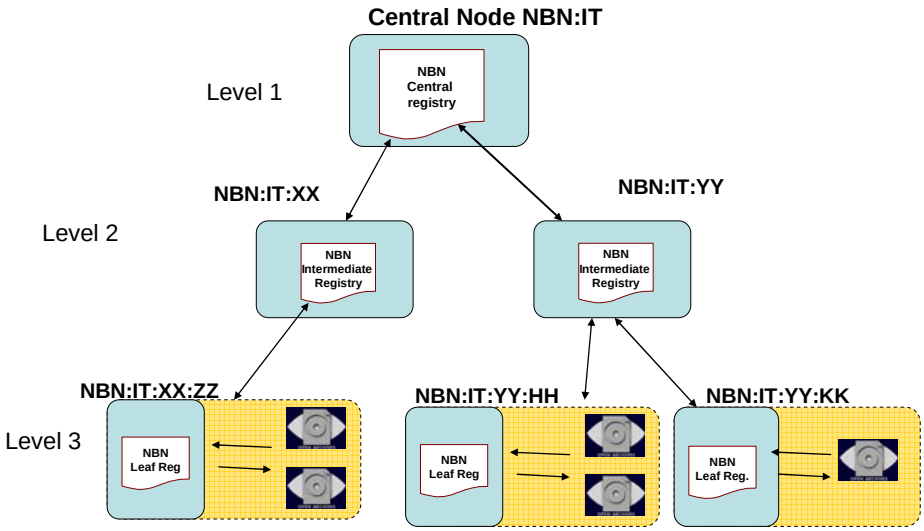

Figure 1: The multi-level distributed architecture

5.1 Policy

The trustability and reliability of an NBN distributed infrastructure can be guaranteed only by defining and enforcing effective policies. To this end the Italian NBN governing board is going to release a general policy that will have to be signed by all the participating agencies.

We have performed an initial analysis to detect problems and issues that the policy should address. In our opinion each agency should satisfy some requirements, which are both technical and organisational, and should commit in respecting some guidelines.

Organisational requirements

Each participating agency should indicate an administrative reference person, who is responsible for policy compliance as regards the registration and resolving procedures as well as for the relationships with the upper and lower level agencies, and a technical reference person, who is responsible for the hardware, software and network infrastructure.

Technical requirements

The hardware hosting an NBN register/resolver should be housed in a managed hosting infrastructure, with uninterruptable power supply and high-speed network connection. An agency that does not have an internal server farm may outsource hosting services to an external provider, which fulfils the technical requirements.

The hardware architecture should be redundant in order to guarantee no single point of failure.

In our opinion it would be also useful to identify and monitor some simple service level indicators, such as service response time and up time, and define thresholds that each agency should respect. Each domain maintainer could monitor its child sub-domains and notify them service level violations. The policy should also define how violations should be dealt with.

Guidelines

The policy should define rules for:

- generating well-formed PIs;
- identifying the digital resources which "deserve" a PI;
- identifying resource granularity for PI assignment (paper, paper section, book, book chapter, etc.)
- auditing repositories in order to assess their weaknesses and their strengths (the Drambora toolkit may help in this area).

5.2 Architectural elements

All the architectural elements of our system feature a web interface, which allow operators to perform administrative tasks and final users to resolve names and access metadata and contents. Below is a description of the different components.

1. **Central Node**
   At the top of the hierarchy there is the central node, which maintains the central register, where all NBN names generated by any NBN leaf nodes are stored. The central node can assign sub-domains to institutions that accomplish a registration procedure; it can resolve a user-query directly or redirect it to the appropriate lower level agency; it checks the NBN records harvested from sub-domain registers for policy compliance and uniqueness. The Central Node supports the management of a Central Agency, which is responsible for the quality and reliability of the answers provided by the whole infrastructure to the user requests.
2. **Inner Node**
   The architecture defines the inner node in order to manage specific sub-domains (as NBN:IT:UR in Figure 1), which control other lower level domains (as NBN:IT:UR:CNR). The Inner Agencies can define their own policies for generating NBN names or registering institutions in relation to the specific domains (e.g.. Cultural Heritage, Scientific, Broadcasting, etc.). The inner node harvests the NBN records from the leaf and/or inner nodes under its responsibility and performs checks that are similar to those described for the central node but for a smaller set of resources. The inner node cannot generate NBN names but can resolve them directly or by redirecting requests to the appropriate nodes.
3. **Leaf Node**
   The Leaf nodes are responsible for the bottom-level sub-domains, which are assigned to the agencies that manage the actual digital libraries. They harvest resources from the repositories under their responsibility and generate resource names on operator's demand. Also leaf nodes can resolve all the NBN names.

6. Implemented functions

In the present section we will provide short descriptions of the main functions implemented in our software.

6.1 Registering a new node

The new node registration process is an important step for the trustability and credibility of the entire network. It is important to define strict off-line (i.e. based on human interaction) and on-line (i.e. based on machine interaction) authentication procedures for the enrolling institutions. These procedures are under development by the NBN Italian consortium. At the moment the software implements the following steps:

1) At the first execution, the node must be configured as inner or leaf node and the operator have to insert the URL of the parent node and some additional information, such as an administrator's e-mail and a description of the managing institution.
2) The software sends this information to the parent node via web service.
3) The new node waits for an answer from the central node that assigns a sub-namespace to the new node. During this waiting time the leaf nodes are enabled to register OAI repositories and harvest the OAI records but cannot generate NBNs.
4) When the node receives the acknowledgement from parent node, it can perform all the actions associated to its role.

6.2 Managing an NBN leaf node registry

Each leaf node is responsible for harvesting publication metadata from one or more repositories and for generating NBN PIs. Harvesting is performed via OAI-PMH [5] protocol and allows gathering metadata in DC [6], METS [7] and MPEG21 [8] formats as presented in [9].

The leaf node administrator can manage the repository list and manually start the harvesting procedure, which gathers metadata regarding all publications deposited after the last "harvest". She/he can then select new publications for PI generation. Once generated, PIs are in "inconsistent" state, which means that the digital resource may have another PI in a different sub-domain. Upper level nodes periodically harvest metadata, which are thus propagated up through the hierarchy until they reach the Central node, which checks PI uniqueness (see following section). If the digital resource already has an NBN PI, the leaf node is requested to delete the record; otherwise it must change the status to "consistent". Uniqueness check is performed via resource hash code[2] comparison.

6.3 Synchronization between NBN registries

As stated before, the central node and the inner nodes periodically harvest new NBN records from their child nodes and update their registries. Node administrators can configure this automated synchronization procedure from the administrative console by defining one or more schedulers. For each scheduler the administrator must indicate a name, the synchronization interval and the sub-domain registers to be harvested. Each node can only harvest its child nodes.

The central node also verifies whether the harvested NBNs are policy-compliant, performs uniqueness checks and notifies the leaf nodes whether the new records must be kept or deleted and whether their state must be changed to "consistent".

6.4 Resolving NBN names

Any node belonging to the NBN infrastructure can be queried to resolve an NBN identifier. If the queried node is able to directly resolve and identifier, it answers immediately to the user, otherwise it parses the URN in order to identify the leaf node, which is responsible for the associated sub-domain. If the queried node knows the leaf node's address, it tries to contact it via a particular web service. If the leaf node is up, the name is resolved and the answer is passed back to the user, otherwise the queried node climbs up the hierarchy until it finds a server that is able to resolve the name.

The described procedure is feasible because each node keeps information about the other nodes (at least about some of them) and is thus able to forward resolution requests. Currently it is an administrator's task to manually insert this information in the node configuration but we are studying automated mechanisms to broadcast infrastructure topology information at fixed intervals.

The user can submit a resolution request either by using a form available within the node's web interface or by specifying a particular http URL within the browser's address box. We have also developed a Mozilla plugin, which enables the browser to accept NBN URNs.

[2] During the harvesting procedure the leaf nodes fetch the files containing the digital resources and use their binary stream as input for a hash function. For each file the hash function returns a so-called hash code, which can be considered as a file fingerprint. If two files have the same hash code there is a very high probability that their content is identical. The probability depends on the adopted hash function.

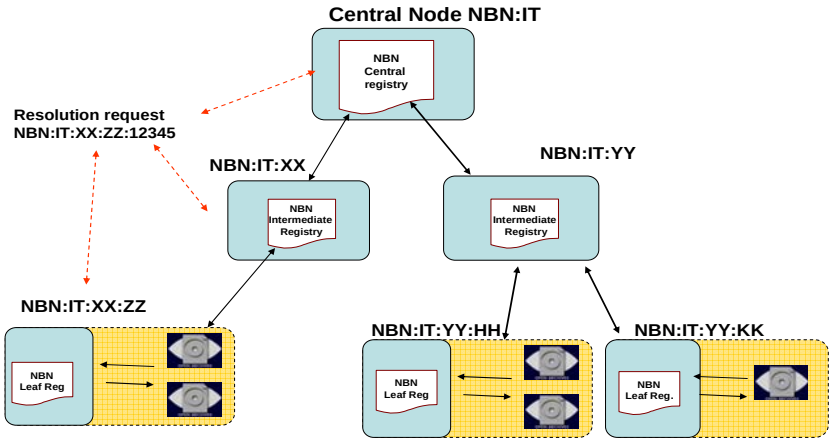

Figure 2: name resolution.

6.5 Digital object relocation

When the content of a digital library (or part of it) is transferred to another library, a problem arises for an NBN resolving infrastructure, because identifier persistence is a mandatory requirement. The comparison with the DNS is of no help in this case. There are two main scenarios which must be considered: a) the digital library dies and no other library can (at least for the moment) legitimately host its digital objects; b) one or more digital objects are transferred from one library to another belonging to a different sub-domain. In the first case our software keeps the NBN record, changing its status to “withdrawn”. The name can still be resolved but it is not possible to access the resource. If the same document is subsequently ingested in a different library and its record is harvested by a leaf node, the infrastructure detects via hash comparison that the object already has an NBN PI an does not create a new one. When this occurs, not only the old name can be resolved but also the resource can be accessed again. In case b) the NBN PI does not change. The register that originally generated the identifier simply redirects resolution requests to the new maintainer. In both cases our infrastructure keeps and resolves the originally assigned NBN PI, thus implementing actual name persistence.

7. Testing activities

After developing a first working prototype, collaborations have been established with several research institutions in order to create a community where final users and software developers are both represented. Several institutions are already involved in user requirement definition or have declared their availability to join the NBN network. These institutions are: the GRIDS group (ISS, INAF, INFN, INGV, ASI, ENEA, INOA, APAT, Università degli Studi di Pisa, Università degli Studi di Roma ‘La Sapienza’), the University of Florence, the Florence University Press.

A first testbed has been deployed where users can execute test cases and provide feedback to the developers in terms of bug/defect notifications, change or enhancement requests and new requirements. On the other hand the developers perform technical tests to evaluate performance, scalability and reliability of the infrastructure and implement what needed to satisfy user indications.

The testbed is configured as follows:

a) central node at BNCF, responsible for the Italian sub-domain (NBN:IT),
b) a second level inner node at CNR, responsible for the “University and Research” sub-domain (NBN:IT:UR),
c) a second level leaf node at FRD, responsible for the local NBN:IT:FRD sub-domain,
d) a third level leaf node at UNIMI, responsible for the local NBN:IT:UR:UNIMI sub-domain,
e) a third level leaf node at CNR, responsible for the local NBN:IT:UR:CNR sub-domain.

The second level CNR inner node (NBN:IT:UR) aims to implement the University and Research National Registry. It currently aggregates the records generated by the UNIMI and CNR leaf nodes for the resources stored in their local repositories. The FRD node generates NBNs for resources stored in a local Dspace repository. A first set of tests has been performed to verify functionalities and behaviour in a distributed

environment using different metadata sets. Performance was not the main focus in this phase and this is the reason why the servers used to set up the infrastructure are neither particularly powerful nor up to date.

First feedbacks from users are positive as regards registering and resolving functionalities. The system harvests resources, assigns NBNs and provides access to metadata and documents as expected. As regards duplicate discovery via hash comparisons, it has been pointed out that this mechanism works only if the compared files are identical, but fails even if they differ for a single bit. It has also been remarked that currently it is not possible to represent within the identifier the "part of" relation between two digital objects. This means that if we want to assign identifiers both to an entire document and to parts of it (e.g. a picture) there is currently no commonly agreed way to represent this inclusion relation in the final part of the persistent identifier. Finally, the need for higher-level services has been expressed by several parties, first of all the possibility of producing reports about the number of publications deposited in a sub-domain within a certain period. This problem is tightly related to the duplicate detection one. If the latter is not solved, resource accounting statistics may be affected by errors whose impact cannot be estimated at the moment.

Some measurements have been carried out to test behaviour when performing bulk harvesting of different metadata sets. In particular, we harvested 1000 digital resources in each trial, using Dublin Core and METS metadata. For a leaf node harvesting from a digital library the average time was about 25 minutes using Dublin Core and 11 minutes using METS. The absolute values are not relevant within this context because the servers involved were neither recent nor powerful and no performance tuning has been performed on any software or hardware component. The relevant aspect is that the time required when using METS is less than a half of the time required when using Dublin Core. This is because the METS set includes by default the MD5 hash code used to perform uniqueness checks, thus avoiding the system to calculate it. It is important to point out that MD5 could be specified also in a Dublin Core record using a custom field.

For an inner node harvesting 1000 records from an child node the average measured time was about 40 minutes. This value is strongly influenced by the uniqueness check, which depends on the underlying DBMS performance. This can be improved by defining indexes on the involved fields and obviously by using a high-performance hardware/software configuration.

8. Conclusions and future directions

In this paper we have described a new software application for a distributed and hierarchical NBN register/resolver infrastructure. The first tests have shown the feasibility of this approach and the group of supporting institutions includes important Italian research organisations and national libraries.

The main technical problems pointed out so far regard the identifier uniqueness guarantee. The proposed solution of using MD5 hash codes partly resolves this issue but poses performance problems and does not cover cases where the same content is represented in different formats. A more comprehensive solution will probably involve the comparison of a strictly defined set of metadata. This means that strict rules and clear responsibilities must be defined as regards data entry in the digital libraries.

From a political point of view the short-term objective is to enlarge the group of supporting institutions in order to create a first nucleus of a credible NBN national infrastructure. On a larger scale, contacts have been established with the Knowledge Exchange consortium in order to cooperate in the development of a top level European NBN resolver. The adoption of our software as top-level node manager will be taken into consideration in the following months.

In our opinion it is also important to identify high-level value-added services (such as digital resource accounting) that could be built on top of the infrastructure. This would probably favour the diffusion of NBN persistent identifiers.

From the technical point of view the next steps will include performance testing and tuning, in order to define the hardware requirements for a production infrastructure that would guarantee the necessary service levels.

The testbed will be enlarged in order to include a leaf node installed at the University of Bologna, which will harvest records from the "Magazzini digitali" project repository. The goal of this project is to enable the BNCF digital library to harvest doctoral thesis from the University of Bologna Eprints repository, in order to accomplish their legal deposit. In this case the resources already have an NBN name. A new NBN record will be created in our registry using the existing identifier, which will be associated to the new URL assigned by legal deposit at BNCF.

A research group has also been established to thoroughly examine the duplication problem and its possible solutions. In this field hash codes different from MD5 could provide better performance with respect to comparison operations. The same group will also address the problem of the "part of" relation representation.

Finally, we are going to investigate ways to establish permanent and reliable connections between NBNs and other persistent identifiers such as DOI, which would favour the implementation of a multi-standard global resolver.

Acknowledgements
We would like to thank Luciana Trufelli from Agenzia Spaziale Italiana (ASI) for her contribution to the activity described in the present paper.

References
[1] European Digital Library
http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edlproject/
[2] Europeana www.europeana.eu/
[3] IETF RFC 3188 Using National Bibliography Numbers as Uniform Resource Names
http://tools.ietf.org/html/rfc3188
[4] IETF RFC 2141 URN Syntax http://tools.ietf.org/html/rfc2141
[5] C. Lagoze and H. V. de Sompel. The Open Archives Initiative Protocol for Metadata Harvesting, version 2.0. Technical report, Open Archives Initiative, 2002.
http://www.openarchives.org/OAI/openarchivesprotocol.html.
[6] Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, Version 1.1.
http://dublincore.org/documents/dces/.
[7] MPEG-21, Information Technology, Multimedia Framework, "Part 2: Digital Item Declaration," ISO/IEC 21000-2:2003, March 2003.
[8] METS, http://www.loc.gov/standards/mets/
[9] Herbert Van de Sompel et al. Resource Harvesting within the OAI-PMH Framework D-Lib Magazine December 2004 Volume 10 Number 12 ISSN 1082-9873
[10] J. Kunze. The ARK Persistent Identifier Scheme. Internet Draft, 2007.
http://tools.ietf.org/html/draft-kunze-ark-14.
[11] Norman Paskin. Digital Object Identifiers. Inf. Serv. Use, 22(2-3):97–112, 2002 http://www.doi.org
[12] Sam X. Sun. Internationalization of the Handle System - A persistent Global Name Service. 1998.
http://citeseer.ist.psu.edu/sun98internationalization.html. www.handle.net
[13] Persistent URLhttp://purl.oclc.org
[14] Library of Congress Control Number http://www.loc.gov/marc/lccn.html
[15] David Giaretta, Issue 1, Volume 2 | 2007 The CASPAR Approach to Digital Preservation The International Journal of Digital Curation
[16] Andersson, Stefan; Hansson, Peter; Klosa, Uwe; Muller, Eva; Siira, Erik Using XML for Long-term Preservation : Experiences from the DiVA Project
[17] Bermes, Emmanuelle, International Preservation News, Vol 40 December 2006, pp 23-26 Persistent Identifiers for Digital Resources: The experience of the National Library of France
http://www.ifla.org/VI/4/news/ipnn40.pdf
[18] Kathrin Schroeder. Persistent Identification for the Permanent Referencing of Digital Resources - The Activities of the EPICUR Project Enhanced Uniform Resource Name URN Management at Die Deutsche Bibliothek. The Serials Librarian, 49:75–87(13), 5 January 2006.
[19] CENL Task Force on Persistent Identifiers, Report 2007
http://www.nlib.ee/cenl/docs/CENL_Taskforce_PI_Report_2006.pdf.
[20] Sollins, Karen Architectural Principles of Uniform Resource Name Resolution (RFC 2276) http://www.ietf.org/rfc/rfc2276.txt
[21] Masinter, Larry; Sollins, Karen Functional Requirements for Uniform Resource Names (RFC 1737) http://www.ietf.org/rfc/rfc1737.txt
[22] E. Bellini, M.Lunghi, E. Damiani, C. Fugazza, 2008, Semantics-aware Resolution of Multi-part Persistent Identifiers, WCKS 2008 conference.
[23] E. Bellini , C. Cirinna, M. Lunghi, E. Damiani, C. Fugazza, 2008 Persistent Identifiers distributed system for cultural heritage digital objects, IPRES2008 conference
[24] H.-W. Hilse, J. Kothe Implementing Persistent Identifiers: overview of concepts, guidelines and recommendations, 2006, ix+57 pp. 90-6984-508-3 http://www.knaw.nl/ecpa/publ/pdf/2732.pdf
[25] DCC Workshop on Persistent Identifiers, 30 June – 1 July 2005 Wolfson Medical Building, University of Glasgow, http://www.dcc.ac.uk/events/pi-2005/
[26] ERPANET workshop Persistent Identifers, Thursday 17th - Friday 18th June 2004-University College Cork, Cork, Ireland , http://www.erpanet.org/events/2004/cork/index.php
[27] Handle System website, http://www.handle.net/
[28] Wikipedia Handle page, http://en.wikipedia.org/wiki/Handle_System