Consiglio Nazionale delle Ricerche

Fondazione RINASCIMENTO digitale

Persistent identifiers:

jNBN, a JEE application for the management of a national NBN infrastructure

Mario Incarnato, Roberto Puccinelli, Marco Spasiano
Consiglio Nazionale delle Ricerche

Emanuele Bellini
Fondazione Rinascimento Digitale

Persistent identifiers

Purpose: 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**

Reliability: the **actual persistence** of identifier systems can only be achieved through the **commitment of the organizations** that assign, manage, preserve and resolve the identifiers

User communities: the existing **user communities differ by technologies and implemented registries** → Digital Object Identifier (DOI), Archival Resource Key (ARK), Handle System, URN, PURL, etc

No general agreement has been reached among the different user communities.

Requirements

In our view the problem can be tackled with a hierarchical distributed approach BUT a credible solution should be:

Policy-based: the organisations managing the identifiers must sign and commit to a well-defined and binding policy.

Open: it should be based on open standards and technologies.

Interoperable: it should provide mechanisms to keep record of other identifiers assigned to the associated resource.

Easy to use: It should be easy an ID to the corresponding resource.

National Bibliography Number

National Bibliography Number (NBN) is a Universal Resource Name (URN) namespace under the responsibility of National Libraries.

Adoption: 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.

Current limitations: URNs are not directly actionable (i.e. 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).

The NBN initiative in Italy 1/2

Management

- A **Steering Committee** for experimental activity management has been established, where **BNCF, BNCR, CNR, FRD, ICCU, ASI** are represented.

- The board **defines** the **structure of the Italian NBN testbed** (domain hierarchy) and the **policies for infrastructure management, sub-domain creation/removal and PI assignment**.

The NBN initiative in Italy 2/2

Software

- The **Italian National Research Council (CNR)** and **Fondazione Rinascimento Digitale (FRD)** worked together to define the requirements and features of a software for the management of a **distributed hierarchical NBN infrastructure**.

- The **Italian National Research Council (CNR),** leveraging internal human resources and facilities, developed the software.

- **CNR** is currently responsible for the **software development and license**.

Objectives

1. Define and propose **methodologies, tools and policies** for the management of a **national stable, trustable and certified NBN register of digital objects** to be adopted by **cultural and scientific** communities.
2. Allow an easier and wider **access to the digital resources produced by Italian cultural institutions**, including digitised or not yet published material; encouraging the adoption of **long term preservation policies**.
3. Develop an **inter-domain resolution service** (e.g., NBN:IT ←→ NBN:DE) with a common meta-data format and a user-friendly interface (pre-condition for global resolver).
4. Implement some '**accounting' mechanisms to be leveraged in the production of statistics about scientific publications and works**

Key points

“Trustability” model

The “trustability” of the whole system relies on a binding policy that all participating agency MUST sign in order to join the infrastructure.

Sustainable approach

Distributed responsibility and functionalities can overcome the limitations imposed by a centralised system and split up the management costs among several partners, while preserving the authoritative control.

Robust infrastructure

Redundant mechanisms must be implemented to ensure high availability of the service (data duplication, multiple service entry points, …).

Policy

The trustability and reliability of an NBN distributed infrastructure can be guaranteed only by defining and enforcing effective policies.

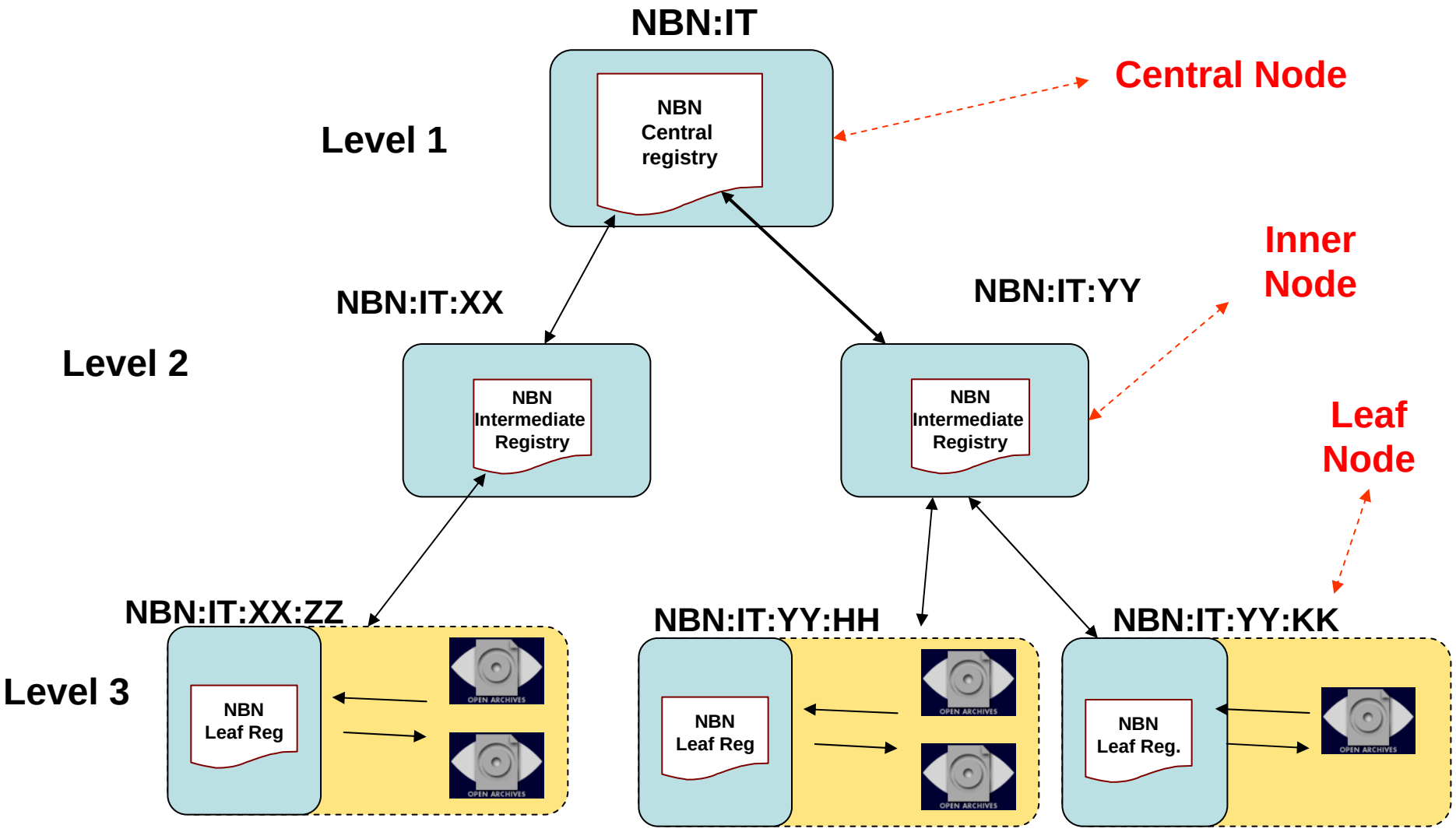
To this end the steering committee for NBN experimental activities in Italy is going to release a common policy, that will have to be signed by all the participating agencies, which will include:

a) Organisational requirements

b) Technical requirements

c) Guidelines for ID management

Distributed NBN architecture



Distributed approach

At the highest level there is a root or **central node (CN)**, 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 for University and Research) 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.). The intermediate sub-domains are managed by the so-called **inner nodes (IN)**.

At the bottom of this hierarchy there are the **leaf nodes (LN)**, 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.

Architectural elements
Central Node (CN)

1. manages the NBN:IT domain both for national and international issues.
2. registers sub-domains for institutions that accomplish a registration procedure.
3. harvests second level nodes and maintains the central register, where all NBN names generated by any leaf nodes are stored.
4. checks the NBN records harvested from sub-domain registers for policy compliance and uniqueness.
5. resolves user-queries directly or redirect them to the appropriate lower level agency.
6. manages cross-requests for names belonging to other NBN national domains or to DOI namespace, providing a minimum set of common metadata

Architectural elements
Inner Node (IN)

The architecture defines the IN in order to manage specific sub-domains which control other lower level domains (as NBN:IT:UR).

a) The Inner Agencies can define their own policies for NBN name generation or sub-agency registration, which apply to their specific domains (e.g.. Cultural Heritage, Scientific, Broadcasting, etc.).

b) The INs harvest the NBN records from the leaf and/or inner nodes under their responsibility and perform checks that are similar to those described for the central node but for a smaller set of resources.

c) The INs **cannot generate NBN names** but can resolve them directly or by redirecting requests to the appropriate nodes.

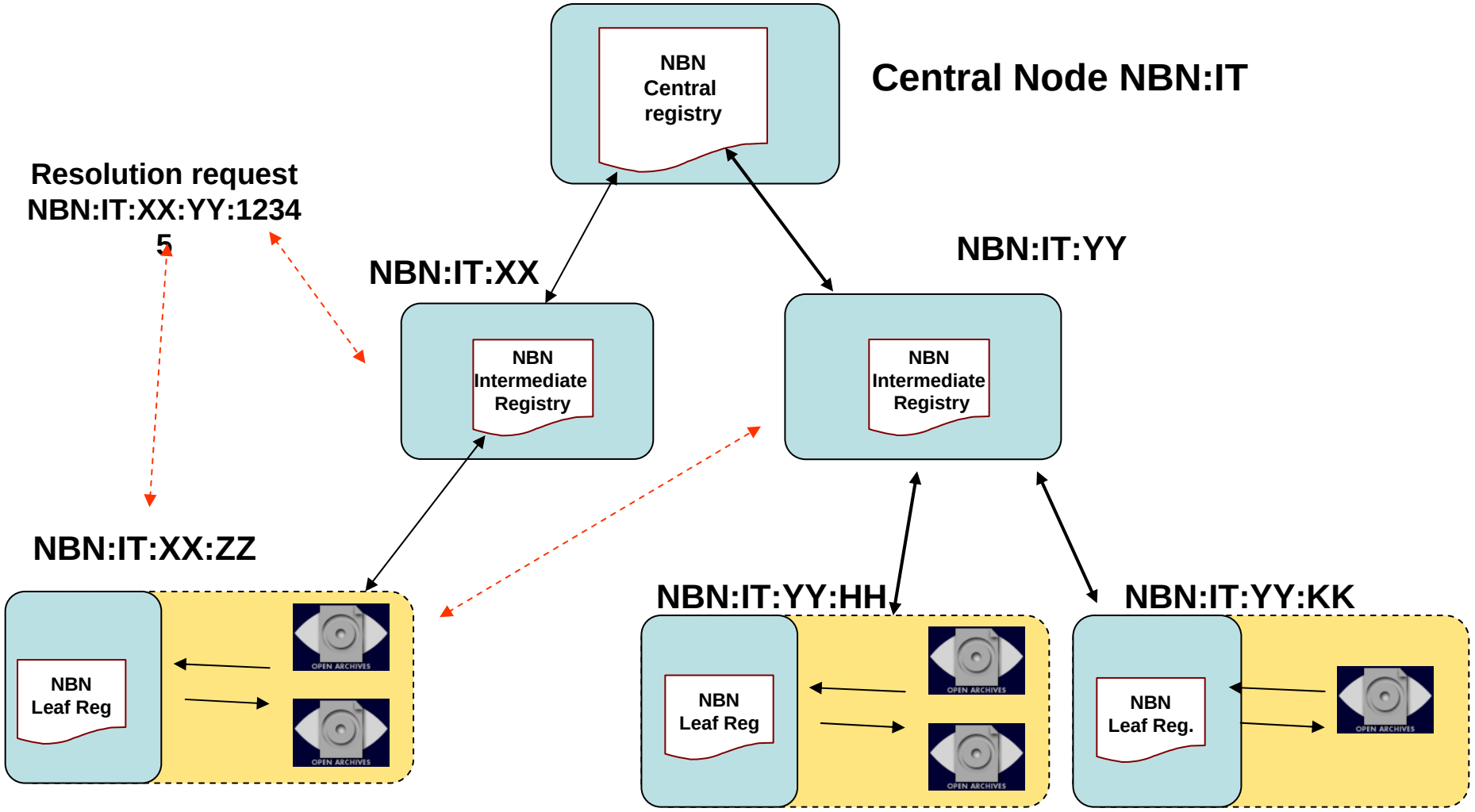
Architectural elements
Leaf Node (LN)

The LNs are responsible for the bottom-level sub-domains, which are assigned to the agencies that manage the actual digital libraries.

a) They harvest digital resources from the repositories under their responsibility and generate resource names on operator's demand.

b) Each LN can resolve NBN names directly or by redirecting requests to the appropriate nodes.

Resolving NBN names



Technologies

- Technology platform:Java Enterprise Edition
- AJAX Framework (user Interface): zKoss
- Development Framework: JADA®
- OAI-PMH library: OCLC

Tools and platforrms

- Application server: JBOSS
- DBMS: Postgres

Development Infrastructure

- Versioning: Subversion
- Change Request Management: Bugzilla
- Development Statistics: StatSVN

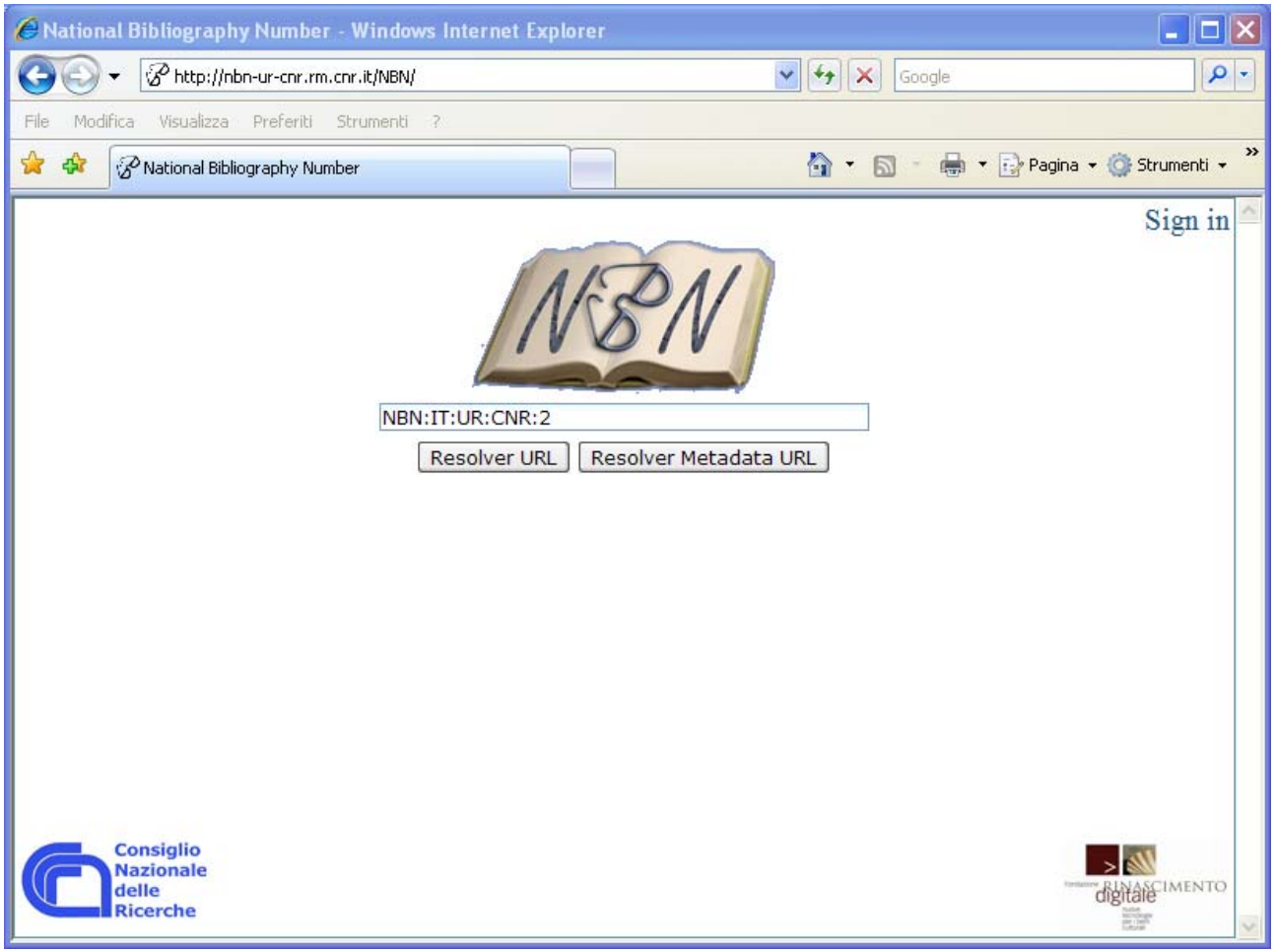
User profiles

- **Administrator**: configure node.
- **Operator**: triggers NBN assignment.
- **Registered User**: queries registry, access categorized view.
- **Public User**: queries registry.

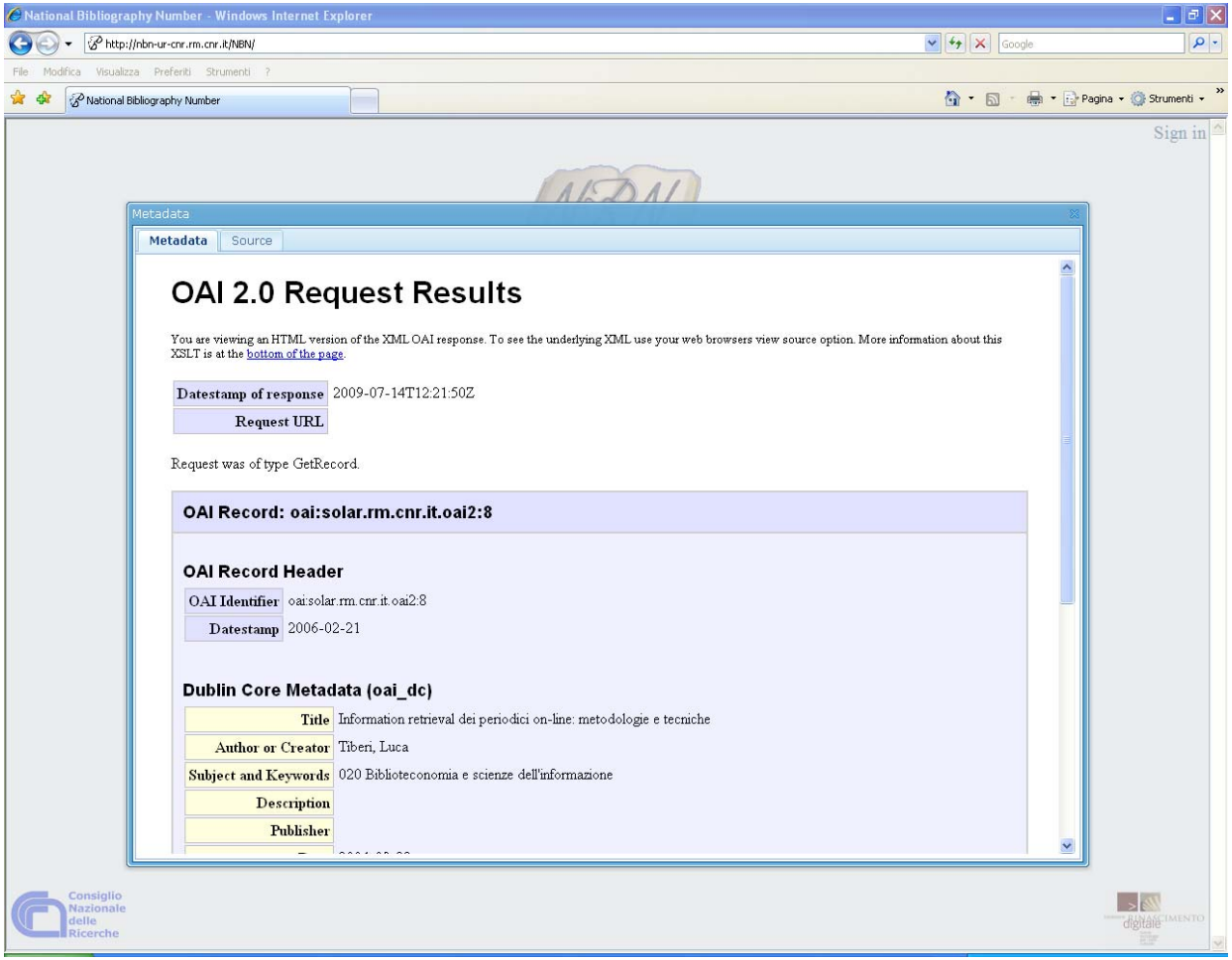
Features

- Resource and metadata retrieval
- Resolution request forwarding
- Scheduled metadata harvesting
- Triggered NBN assignment
- Ubiquitous name resolution
- Duplicate check
- Supported metadata schemata: Dublin Core, Mets.

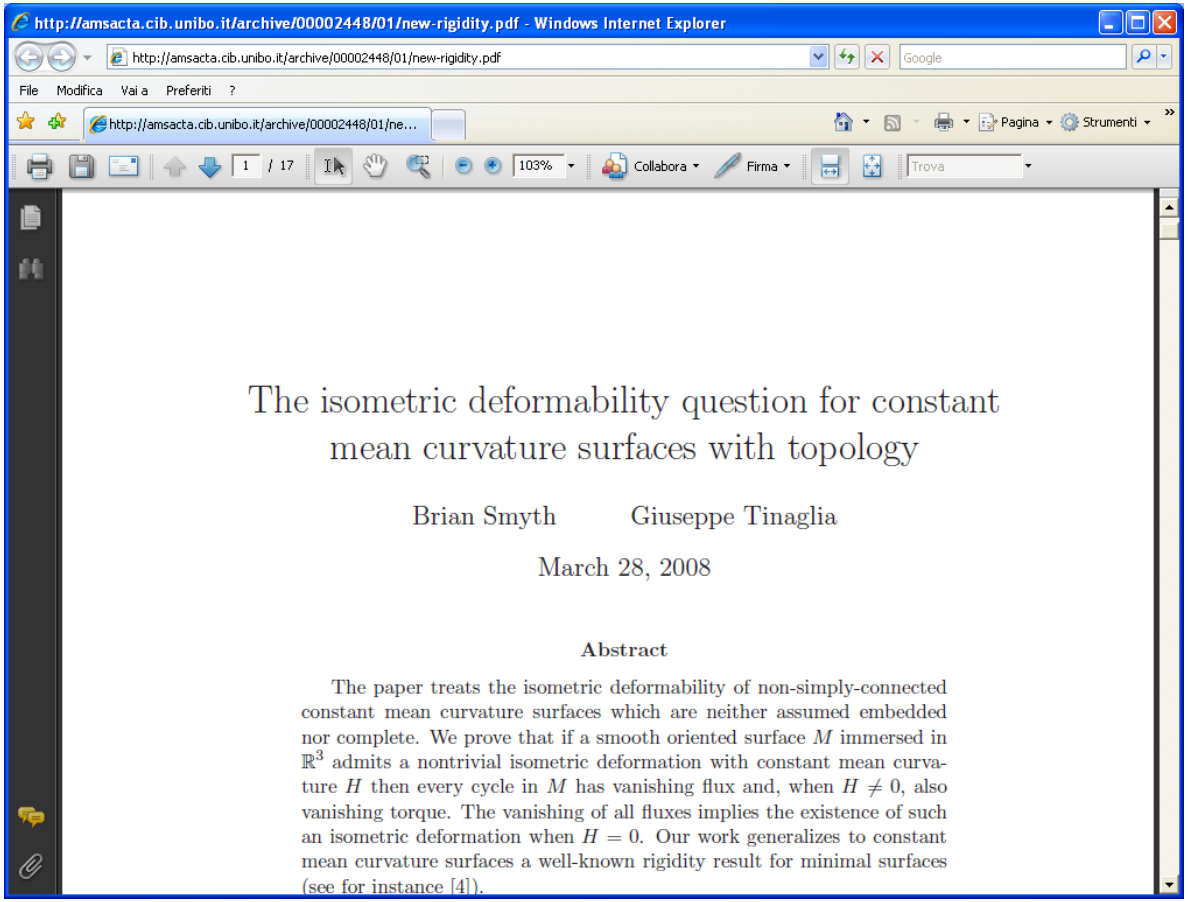
Simple search



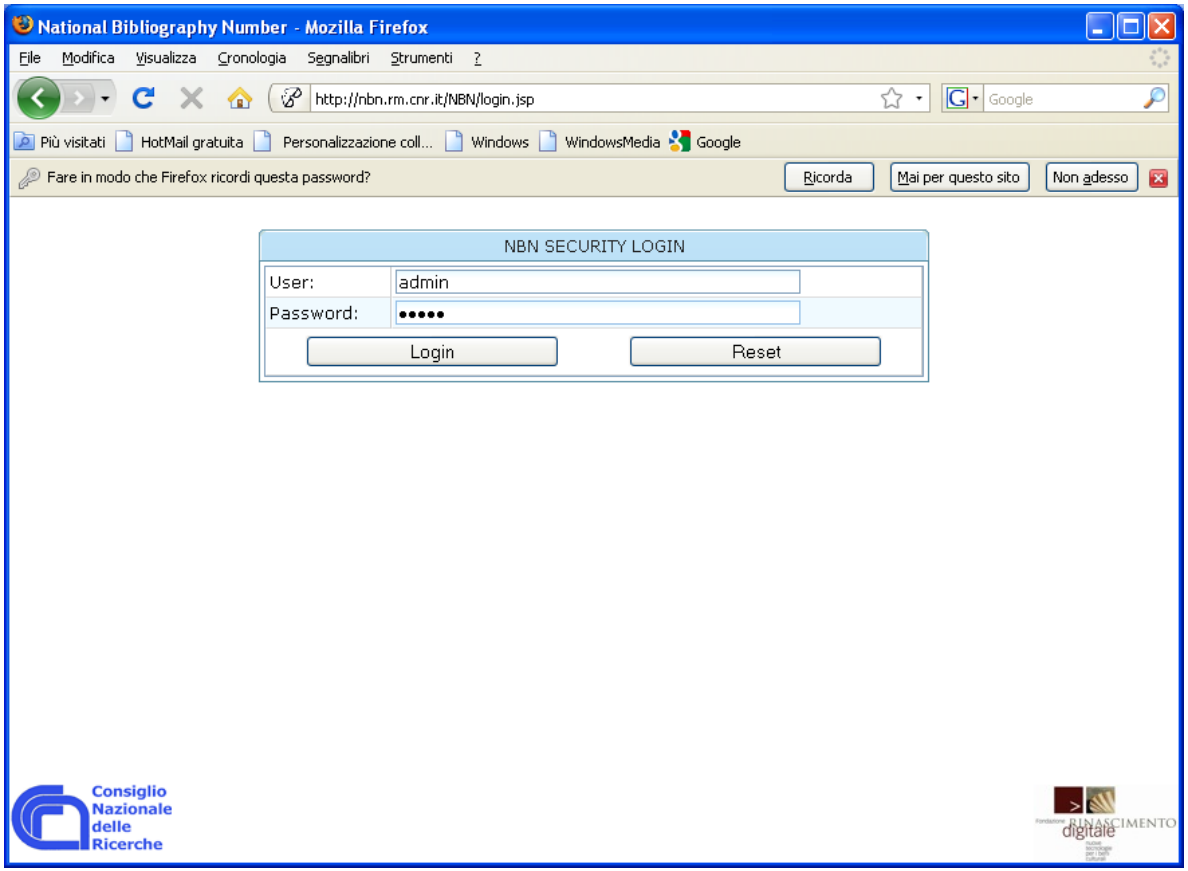
Metadata


National Bibliography Number - Windows Internet Explorer
http://nbn-ur-cnr.rm.cnr.it/NBN/
National Bibliography Number
Sign in
Metadata
Metadata
Source
OAI 2.0 Request Results
You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source option. More information about this XSLT is at the bottom of the page.
Datestamp of response 2009-07-14T12:21:50Z
Request URL
Request was of type GetRecord.
OAI Record: oai:solar.rm.cnr.it.oai2:8
OAI Record Header
OAI Identifier oai:solar.rm.cnr.it.oai2:8
Datestamp 2006-02-21
Dublin Core Metadata (oai_dc)
Title Information retrieval dei periodici on-line: metodologie e tecniche
Author or Creator Tiberi, Luca
Subject and Keywords 020 Biblioteconomia e scienze dell'informazione
Description
Publisher

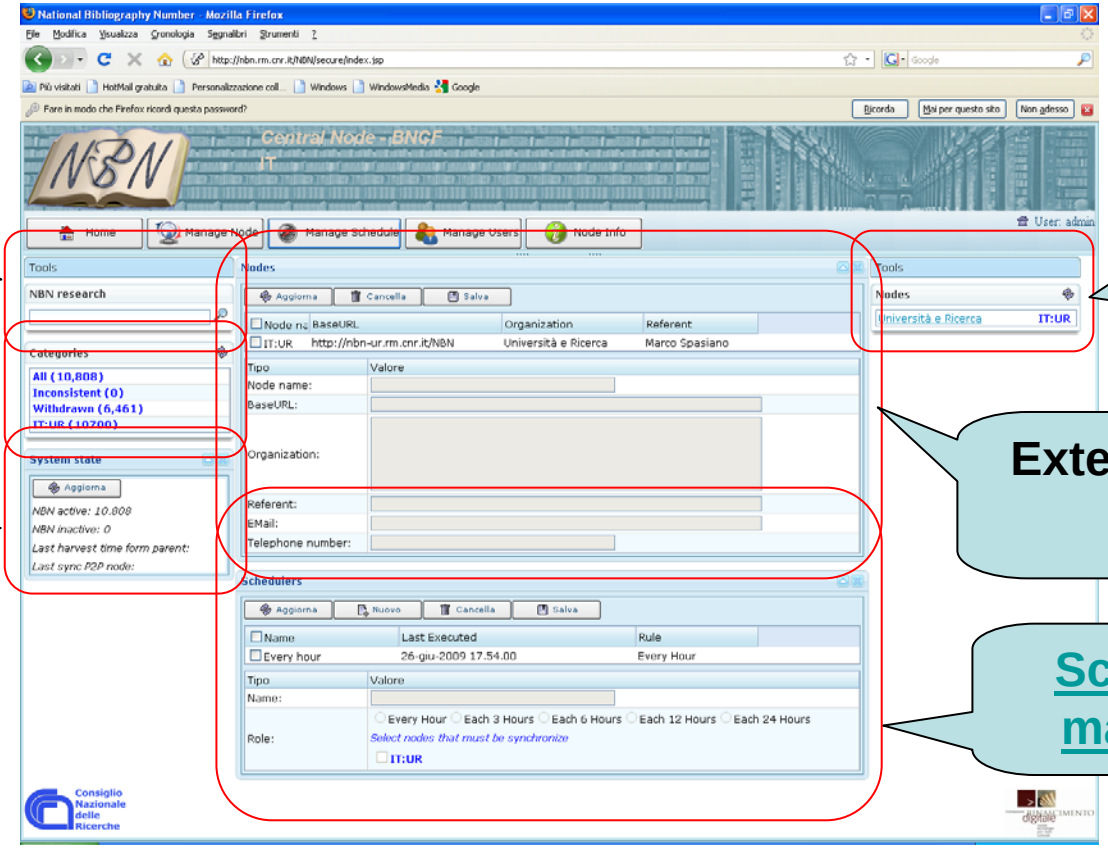
Actual document



Login


National Bibliography Number - Mozilla Firefox
File Modifica Visualizza Cronologia Segnalibri Strumenti ?
http://nbn.rm.cnr.it/NBN/login.jsp
Più visitati HotMail gratuita Personalizzazione coll... Windows WindowsMedia Google
Fare in modo che Firefox ricordi questa password?
Ricorda
Mai per questo sito
Non adesso
NBN SECURITY LOGIN
User:
admin
Password:
Login
Reset
Consiglio Nazionale delle Ricerche
Fondazione RINASCIMENTO digitale

Administrator home



Search mask

Categories

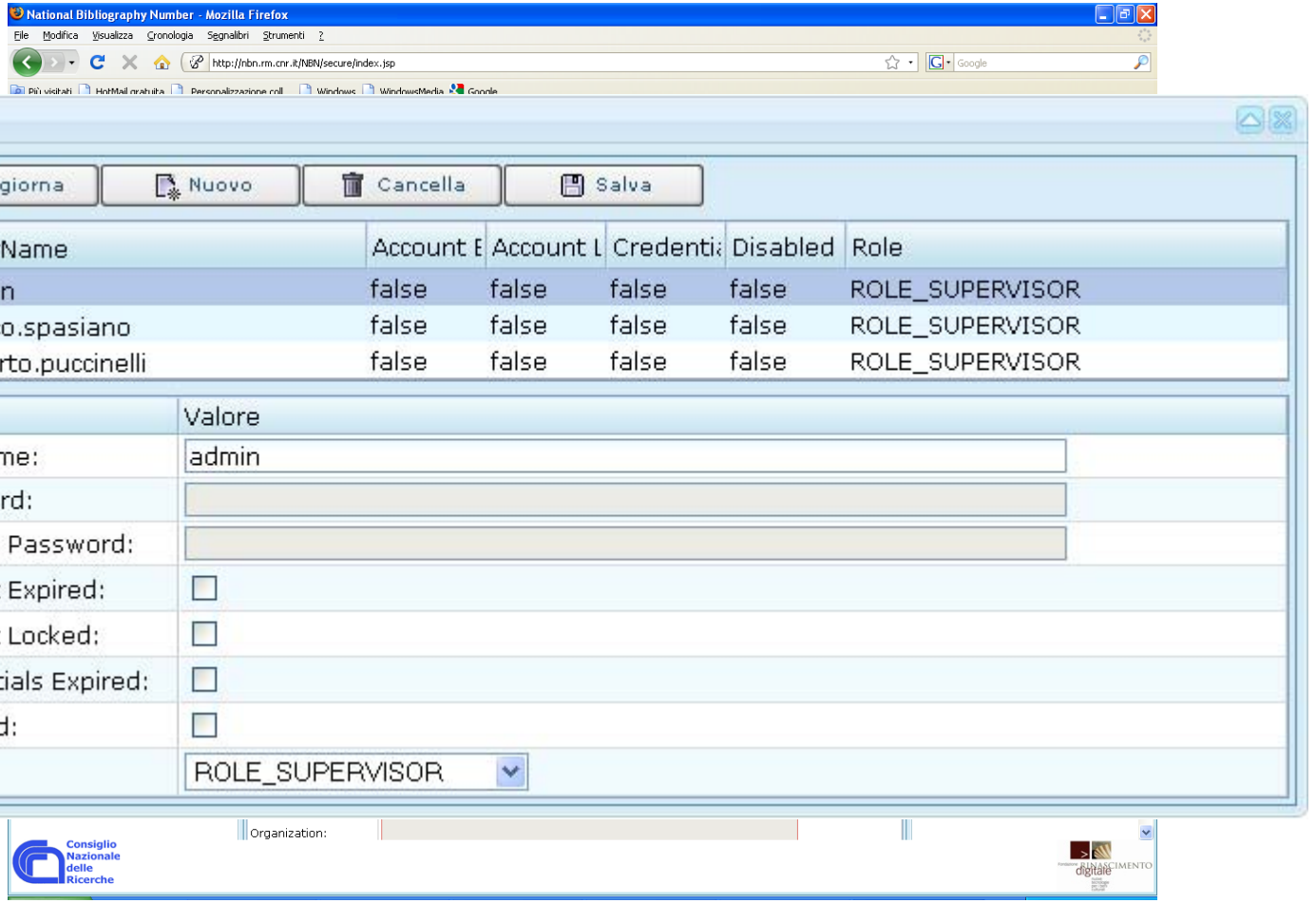
Node State

List of other known nodes

External node info

Schedule manager

User management



Testbed configuration

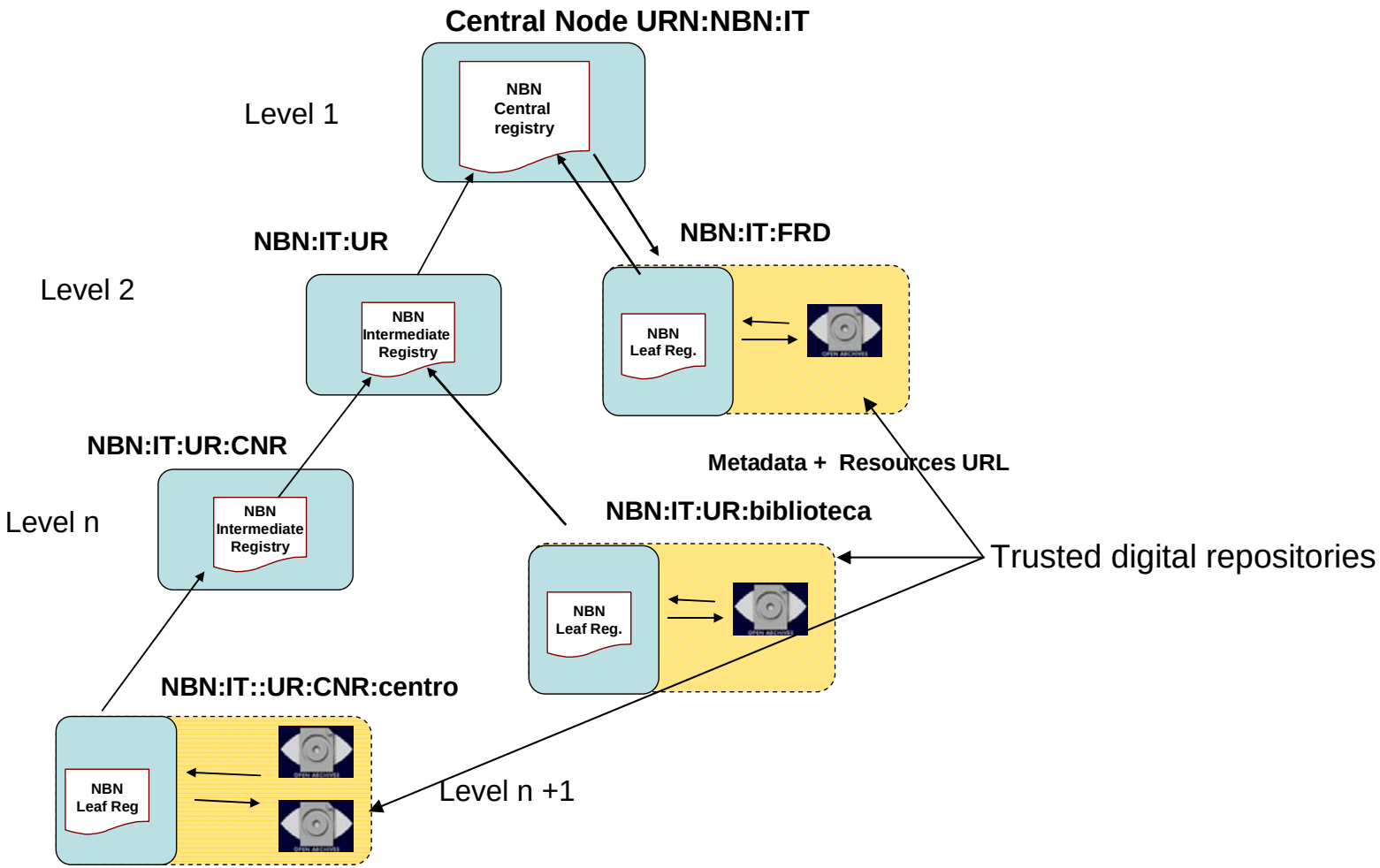
Architecture

- central node at BNCF and duplicated at BNCR, responsible for the Italian NBN:IT, generating and managing all the sub-domains, guiding the resolution service for Italy and the international coordination;
- a second level inner node at CNR, responsible for a sample thematic sub-domain (NBN:IT:UR); a third level leaf node responsible for the local NBN:IT:UR:CNR
- a third level leaf node at UNIMI, responsible for the NBN:IT:UR:UNIMI
- a second level leaf node at FRD, responsible for the local NBN:IT:FRD

Functions

- distributed name generation for digital resources
- uniqueness and authenticity certification for generated names
- armonisation of policies among different user communities
- distributed capacity to resolve the association name-metadata-resource

Italian testbed



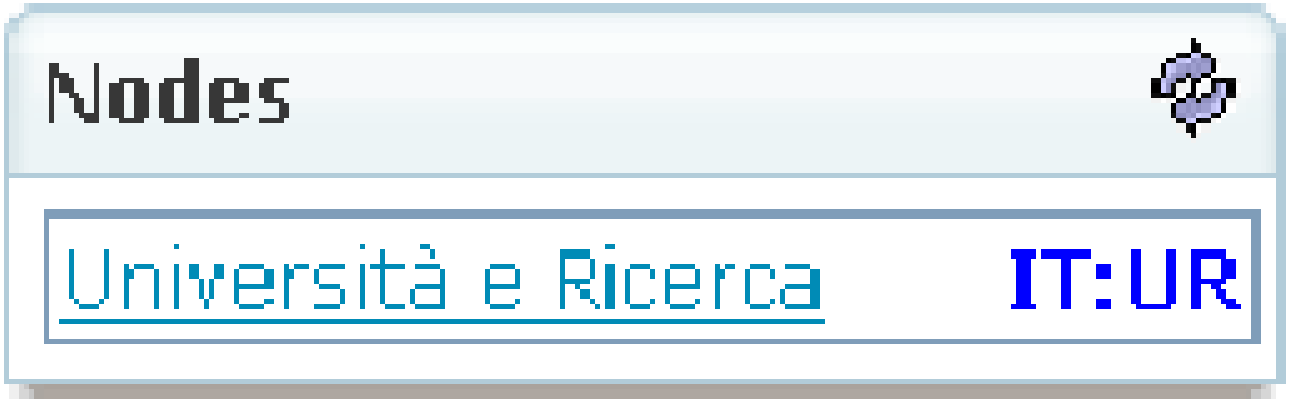
Lessons learned

- Distributed hierarchical approach is viable.
- PI uniqueness check can be performed by comparing the digital fingerprints of the resources with some limitations:
  - the fingerprint must be calculated at the digital repository level and included in the “harvestable” metadata;
  - MD5 fingerprint allows detecting only exact duplicates.

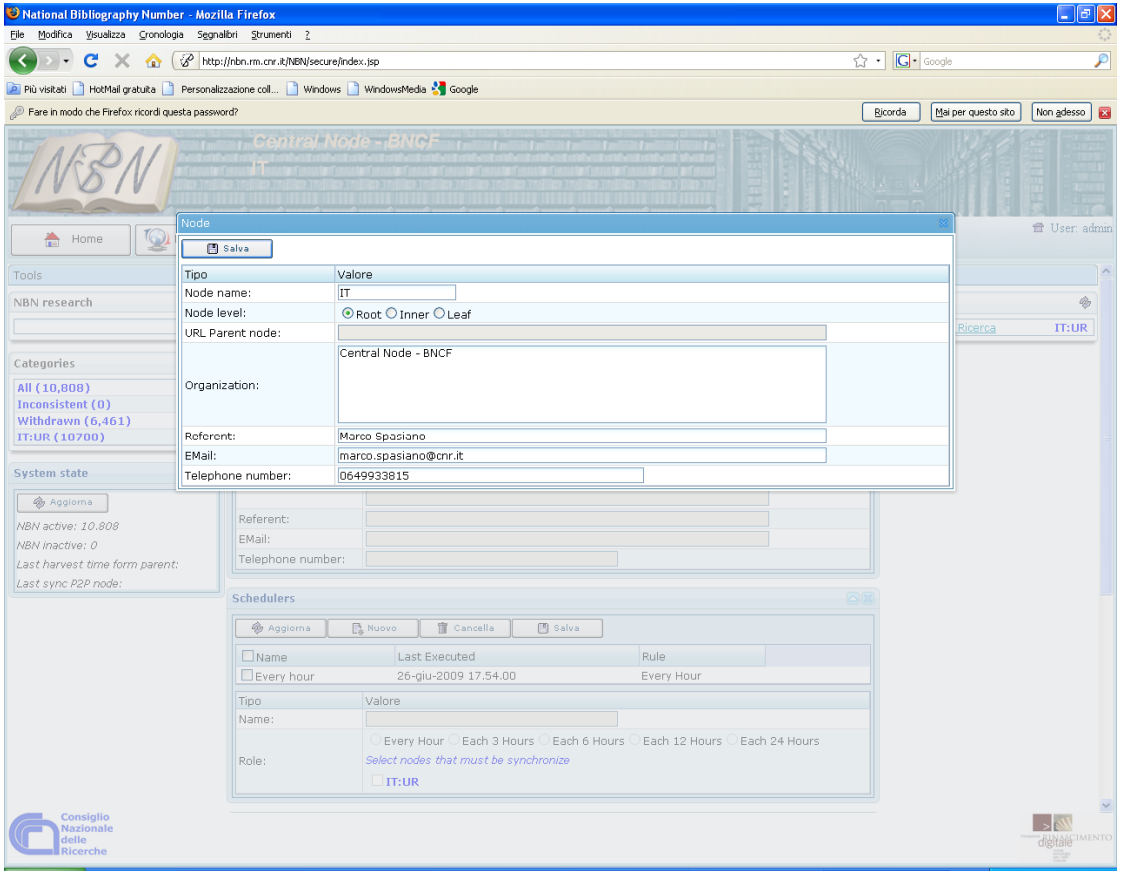
Future steps

- Software development …
- Implementation of new functions such as ‘accounting’, better duplicate detection, …
- Enlargement of the Italian network
- DOI interoperability
- Field experience of our software in other countries (Czeck REPUBLIC)
- Interoperability with NBN systems deployed in other countries for cross-domain name resolution.

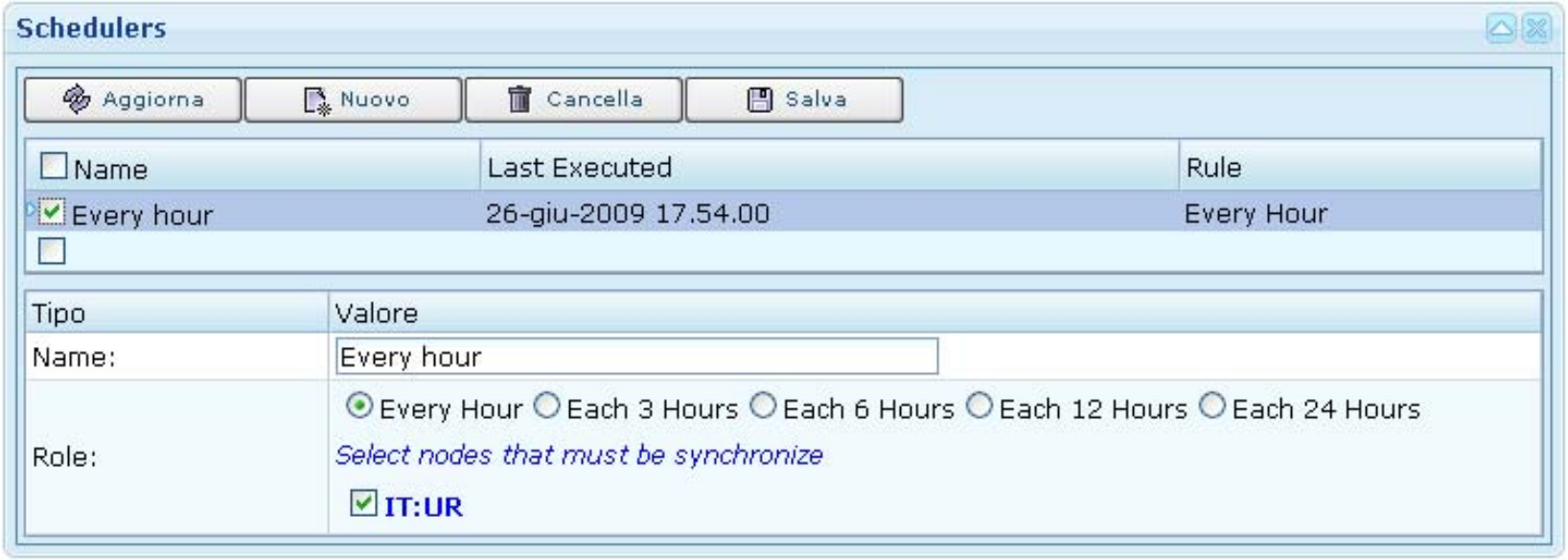
External node info



Node info


National Bibliography Number - Mozilla Firefox
Central Node - BNCF
IT
Node
Salva
Tipo
Valore
Node name:
IT
Node level:
Root Inner Leaf
URL Parent node:
Organization:
Central Node - BNCF
Referent:
Marco Spasiano
EMail:
marco.spasiano@cnr.it
Telephone number:
0649933815
Schedulers
Aggiorna
Nuovo
Cancella
Salva
Name
Last Executed
Rule
Every hour
26-giu-2009 17.54.00
Every Hour
Name:
Every Hour
Each 3 Hours
Each 6 Hours
Each 12 Hours
Each 24 Hours
Role:
Select nodes that must be synchronize
IT:UR

Schedule


Schedulers
Aggiorna
Nuovo
Cancella
Salva
Name
Last Executed
Rule
Every hour
26-giu-2009 17.54.00
Every Hour
Tipo
Valore
Name:
Every hour
Role:
Every Hour
Each 3 Hours
Each 6 Hours
Each 12 Hours
Each 24 Hours
Select nodes that must be synchronize
IT:UR

Repository listing

NBN	URL	Creation date	State	Flag
NBN:IT:UR:CNR:37	http://orca.cf.ac.uk/3786/	07/05/2009 00:50	active	withdrawn
NBN:IT:UR:UNIBO:	http://amsacta.cib.unibo.it/archive/00002445/01/self_evacuated_08.pdf	30/04/2009 13:23	active	consistent
NBN:IT:UR:UNIBO:	http://amsacta.cib.unibo.it/archive/00002446/01/Magi_ehb.pdf.pdf	30/04/2009 13:23	active	consistent
NBN:IT:UR:UNIBO:	http://amsacta.cib.unibo.it/archive/00002447/01/AdornoBerniniPellegrini3.pdf	30/04/2009 13:23	active	consistent
NBN:IT:UR:UNIBO:	http://amsacta.cib.unibo.it/archive/00002448/01/new-rigidity.pdf	30/04/2009 13:23	active	consistent
NBN:IT:UR:UNIBO:	http://amsacta.cib.unibo.it/archive/00002449/01/new-dynamics1.pdf	30/04/2009 13:23	active	consistent
NBN:IT:UR:UNIBO:	http://amsacta.cib.unibo.it/archive/00002455/01/Average_NT215.pdf	30/04/2009 13:23	active	consistent
NBN:IT:UR:UNIBO:	http://amsacta.cib.unibo.it/archive/00002457/01/supplemento_6_quaderni_scienza_comser	30/04/2009 13:23	active	consistent
NBN:IT:UR:UNIBO:	http://amsacta.cib.unibo.it/archive/00002460/01/FUNAIOLI_2007_versione_FINALE.pdf http://amsacta.cib.unibo.it/archive/00002460/02/funaioli_cop_acta.jpg	30/04/2009 13:23	active	consistent
NBN:IT:UR:UNIBO:	http://amsacta.cib.unibo.it/archive/00002461/01/Quaderni_2008_dualismomercatolavoro.pd	30/04/2009 13:23	active	consistent

1 / 1078