



## Un sistema distribuido de Identificadores Permanentes para Bibliotecas Digitales

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

\*Fondazione Rinascimento Digitale

\*\*Consiglio Nazionale delle Ricerche

*Traducción de la Biblioteca Nacional de España*

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

*En este artículo presentamos una iniciativa italiana, en la que participan importantes instituciones de investigación y bibliotecas nacionales, cuyo objetivo es la implementación de una infraestructura de Identificadores Permanentes de NBN basada en una novedosa arquitectura de hard- y software. Describimos un modelo distribuido y jerárquico para la gestión de un espacio de nombres de NBN (NBN namespace) e ilustramos políticas de asignación y estrategias de resolución de identificadores basadas en mecanismos de reenvío de consultas (request forwarding). Comenzando por las motivaciones principales para la asignación de identificadores permanentes a objetos digitales, mostramos el estado de la cuestión en el ámbito de las tecnologías relativas a los PI, estándares e iniciativas e ilustramos otras implementaciones de los NBN. A continuación describimos la estructura y objetivos de nuestra iniciativa e ilustramos las funciones/las prestaciones/los servicios ya implementados en nuestro sistema así como los resultados de nuestras actividades de prueba. Finalmente mostramos algunas conclusiones e indicamos la dirección que seguirá nuestro trabajo en el futuro.*

### 1. Introducción

La existencia de referencias estables y certificadas para recursos en línea resulta crucial para las aplicaciones de bibliotecas digitales, no sólo para identificar un recurso de un modo fiable y certificado, sino también para garantizar el acceso continuado al mismo con el paso del tiempo. Iniciativas actuales como la European Digital Library (EDL) [1] y Europeana [2] muestran claramente la necesidad de un mecanismo certificado y estable de referencia a recursos digitales en los ámbitos cultural y científico. La falta de confianza en la fiabilidad de los recursos digitales obstaculiza el uso de la Biblioteca Digital como una plataforma de preservación, investigación, citación y diseminación de

contenidos digitales [15]. Una solución fiable es asociar a cada recurso digital de interés un Identificador Permanente que certifique su autenticidad y garantice su accesibilidad. De hecho, existen algunas propuestas tecnológicas [24], pero la situación actual muestra que no podemos esperar/imponer una única tecnología de PI o un único registro central para todo el mundo. Y lo que es más, diferentes comunidades de usuarios no suelen estar de acuerdo en los detalles relativos a qué debería esperarse de un identificador.

En el mundo bibliotecario, el National Bibliographical Number (NBN – RFC3188) ha sido definido y actualmente es fomentado por el CENL. Este formato estándar de identificador asume que las Bibliotecas Nacionales son responsables de los nombres de los registros nacionales. Las primeras implementaciones del NBN en Europa están ya disponibles en las Bibliotecas Nacionales de Alemania y Suecia.

En Italia actualmente estamos desarrollando una novedosa arquitectura de NBN con una importante participación de la comunidad científica liderada por el Consejo Nacional de Investigación (CNR) a través de su Biblioteca Central y su Servicio ITC (Instituto de Tecnologías de la Construcción). Hemos diseñado un sistema jerárquico y distribuido, similar al DNS<sup>1</sup>, para superar los puntos críticos de un sistema centralizado y para reducir los altos costes de gestión implicados por un único servicio de resolución. Nuestro enfoque implica un nodo central responsable del NBN:IT, el nivel superior del dominio italiano, y nodos de más bajo nivel responsables cada uno de gestionar uno de los subdominios italianos (NBN:IT:UR, NBN:IT:UR:CNR, NBN:IT:FRD, etc.). El número de niveles contenidos en esta jerarquía es virtualmente ilimitado. Sólo los nodos del nivel inferior recolectan metadatos de los repositorios existentes y crean identificadores de NBN. Los niveles superiores solamente recolectan nuevos registros de NBN de los nodos inferiores y los almacenan en sus bases de datos. De este modo, cada nodo mantiene todos los registros en su propio subdominio. Es fácil ver que, en este tipo de arquitectura, la responsabilidad de la creación/conversión/definición de los nombres se encuentra distribuida y la información acerca de identificadores permanentes se replica en múltiples páginas, logrando de este modo la redundancia y resistencia necesarias para implementar un servicio fiable.

Antes de describir nuestro sistema al detalle, en los próximos apartados vamos a mostrar un panorama de las tecnologías actualmente disponibles de identificadores permanentes.

## **2. Estándares de identificadores permanentes**

El hecho de que un Identificador Permanente se asocie a un recurso digital puede servir para certificar la autenticidad de su contenido, su procedencia, sus derechos de gestión y para proporcionar un localizador efectivo. La única garantía de la actual permanencia de los sistemas de identificadores es el compromiso mostrado por las asociaciones que asignan, gestionan y definen los identificadores [25],[26].

En la actualidad existen algunas soluciones tecnológicas pero no se ha alcanzado un acuerdo general entre las diferentes comunidades de usuarios. En las líneas que siguen ofrecemos una breve descripción de los más difundidos. Sólo el estándar NBN [3] será descrito con detalles en el siguiente apartado.

---

<sup>1</sup> DNS (Domain Name System) es la infraestructura que convierte nombres de servidores, como [www.ifla09.it](http://www.ifla09.it), en direcciones IP, tales como 150.146.1.5, en Internet.

El Document Object Identifier (DOI) es un sistema de orientación comercial adoptado ampliamente por la industria publicitaria que facilita herramientas administrativas y un Sistema de Gestión de Derechos Digitales (DRM).

El Archival Resource Key (ARK) es un estándar de identificación persistente basado en la URL que proporciona prestaciones no facilitadas por otros esquemas de PI. (por ejemplo, la capacidad de separar el identificador unívoco asociado a un recurso de las potencialmente múltiples direcciones que podrían servir como proxy al recurso final.

El Handle System [12], [26],[27] es una especificación tecnológica para asignar, gestionar y definir identificadores permanentes para objetos digitales y otros recursos en Internet. Los protocolos especificados permiten a un sistema distribuido de ordenadores almacenar identificadores de recursos digitales a los que otorgan la información necesaria para localizar estos recursos, acceder a ellos y hacer uso de los mismos. Esta información puede cambiarse en caso de ser necesario para reflejar el estado y/o localización actual del recurso identificado sin necesidad de cambiar el identificador.

Finalmente, la URL permanente (PURL) es simplemente una tabla de redireccionamiento de URLs. Es tarea del gestor del sistema implementar las políticas necesarias de autenticidad, derechos y fiabilidad. Junto a ello, el LCCN, el Número de Control de la Biblioteca del Congreso [14], es un sistema de identificadores permanentes con un servicio asociado de URLs permanentes similar al PURL pero con una política fiable en lo que hace la fiabilidad y estabilidad de los identificadores.

Este panorama muestra que no es posible imponer una tecnología única de identificadores permanentes y que el éxito del modelo que se lleve a cabo se relacionará con la credibilidad de la institución que la promueva. Es más, el nivel de detalle de los objetos a que deben asignarse dichos identificadores es muy diferente en cada sector de aplicación.

### **3. Panorama del NBN**

El Número Bibliográfico Nacional (NBN) [3] es un espacio de nombres de URNs que se halla bajo la responsabilidad de las Bibliotecas Nacionales. El espacio de nombres del NBN, como Identificador de Espacio de Nombres (NID), ha sido ya registrado y adoptado por Nordic Metadata Projects a petición del CDNL y del CENL. A diferencia de las URLs, las URNs no pueden accionarse directamente (los buscadores, generalmente, no saben qué hacer con ellas), ya que no tienen una infraestructura global que permita su definición (como los DNS que sostienen a las URLs). A pesar de que se han llevado a cabo diversas implementaciones, cada cual con su propia propuesta de definición mediante plug-ins y servidores Proxy, aún no se ha implementado una infraestructura que permita la definición a larga escala. Es más, cada dominio de una URN se encuentra aislado de otros sistemas y, en particular, el servicio de definición es específico (y diferente) para cada dominio.

Cada Biblioteca Nacional utiliza su serie de NBNs de modo independiente y los implementa mediante sistemas individuales, sin coordinarse con otras bibliotecas nacionales y sin utilizar formatos acordados de modo común. De hecho, muchas bibliotecas nacionales han desarrollado sus propios sistemas de NBNs para proyectos nacionales e internacionales de investigación; actualmente se utilizan muy diversas implementaciones, cada cual con diferentes metadatos y niveles de detalle.

Algunos ejemplos serían el proyecto DIVA [16], EPICUR [18], o ARK, de la Biblioteca Nacional Francesa [17].

Existen algunas iniciativas de importancia a nivel europeo como el proyecto TEL (The European Library), que se encuentra en proceso de implementar un sistema único basado en un espacio de nombres de NBN en el ámbito de la European Digital Library (EDL). La adopción de identificadores de NBN es necesaria para el desarrollo del “National Libraries Resolver Discovery Service” tal y como está descrito en el informe del grupo de trabajo de identificadores permanentes de la CNL [19].

En nuestra opinión, NBN es una tecnología creíble candidata a convertirse en una infraestructura abierta e internacional para identificadores, principalmente porque se basa en un estándar abierto y promueve la distribución de la responsabilidad para los distintos sub-espacios de nombres, permitiendo de este modo que cada institución mantenga el control sobre los identificadores permanentes asignados a sus recursos.

#### **4. La iniciativa NBN en Italia**

El proyecto de llevar a cabo un registro italiano de NBNs comenzó en 2007 como un proyecto de colaboración entre la “Fondazione Rinascimento Digitale” (FRD), la Biblioteca Digital de Florencia (BNCF), la Universidad de Milán (UNIMI) y el “Consorzio Interuniversitario Lombardo per l’elaborazione automatica” (CILEA). Tras un año de trabajo se presentó un primer prototipo que demostraba la viabilidad del modelo jerárquico. El prototipo hacía uso de algunas prestaciones de DSpace y de ARK y proporcionaba una interfaz web PHP básica para bibliotecarios y usuarios finales. La jerarquía estaba limitada a un máximo de dos niveles.

La segunda y actual fase de la iniciativa italiana de NBN se basa en una asociación diferente que incluye a la Agenzia Spaziale Italiana (ASI), el Consiglio Nazionale delle Ricerche (CNR), la Biblioteca Nazionale Centrale di Firenze (BNCF), la Biblioteca Nazionale Centrale di Roma (BNCR), el Istituto Centrale per il Catalogo Unico (ICCU), la Fondazione Rinascimento Digitale (FRD) y la Università di Milano (UniMi). A comienzos del 2009 el Consejo Nacional de Investigación (CNR) desarrolló un segundo prototipo basado en la tecnología de Java Enterprise en el cual se eliminaba la necesidad de DSpace y Ark así como el límite de la jerarquía establecido anteriormente en dos niveles, introduciendo además nuevas prestaciones.

Se han definido objetivos futuros para extender la funcionalidad e integrar el sistema en una infraestructura internacional. Con este fin, el grupo italiano está estableciendo actualmente colaboraciones internacionales.

A continuación se incluyen la descripción de objetivos, órganos de gobierno y política de licencias definidos en la iniciativa italiana.

##### Objetivos:

Las metas del proyecto son:

- crear un registro nacional de objetos digitales estable, fiable y certificado que pueda adoptarse en instituciones culturales y educativas;
- permitir un acceso más sencillo y amplio a los recursos digitales producidos por instituciones culturales italianas, incluyendo material digitalizado o aún no publicado;

- animar a que se adopten políticas de preservación a largo plazo haciendo los costes y las responsabilidades del servicio más sostenibles, al tiempo que se preservan el flujo institucional de procedimientos relativos la edición digital;
- implementar un nuevo servicio basado en URNs, similar a los sistemas de otros países pero con una arquitectura más avanzada que permita distribuir responsabilidades en el ámbito de la gestión de nombres;
- extender cuanto sea posible la adopción de la tecnología del NBN y la red de usuarios en Italia;
- desarrollar un servicio de resolución de dominio múltiple (por ejemplo NBN Italia y NBN Alemania, o bien NBN Italia y DOI) con un formato común de metadatos y una interfaz de usuario amigable (condición previa para un gestor global de enlaces);
- crear algunos mecanismos redundantes tanto para la duplicación de registros de nombres como, en algunos casos, también para los recursos digitales en sí.
- superar las limitaciones impuestas por un sistema centralizado y distribuir los altos costes de gestión implicados por la existencia de un único servicio, conservando al mismo tiempo un control de confianza.

#### Órganos de gobierno:

Para definir la organización y políticas del registro italiano se ha establecido un órgano de gobierno en que se hallan representados BNCF, BNCR, CNR, FRD y ICCU. Dicho órgano de gobierno define la estructura del primer nivel de la jerarquía del NBN italiano y las políticas relativas a gestión de infraestructuras, creación o eliminación de subdominios y asignación de identificadores permanentes.

#### Política de licencias de software:

Los miembros del proyecto están trabajando actualmente para definir una licencia de software que resulte apropiada a sus objetivos. En particular, para animar a que otros registros nacionales adopten nuestro software y para fomentar la creación de una comunidad de apoyo, se piensa que una licencia de acceso libre facilitaría estos aspectos.

### **5. El modelo de arquitectura distribuida**

La arquitectura propuesta (véase figura 1), comenzando por [22], [23] y teniendo en cuenta los requisitos estándar de las URN [20], [21], introduce algunos elementos que aportan flexibilidad y prestaciones adicionales. En el nivel más alto existe un nodo raíz, responsable del dominio del nivel superior (IT en nuestro caso). El nodo raíz delega la responsabilidad para los distintos dominios de segundo nivel (por ejemplo, IT:UR, IT:FRD, etc.) en las autoridades encargadas de dar nombres al segundo nivel. La responsabilidad de los subdominios puede seguir delegándose usando el número virtualmente ilimitado de subdominios (por ejemplo, IT:UR:CNR, IT:UR:UNIMI, etc.). En la base de esta jerarquía hay nodos externos, los únicos que recolectan metadatos de publicaciones de los actuales repositorios y asignan identificadores unívocos a los objetos digitales.

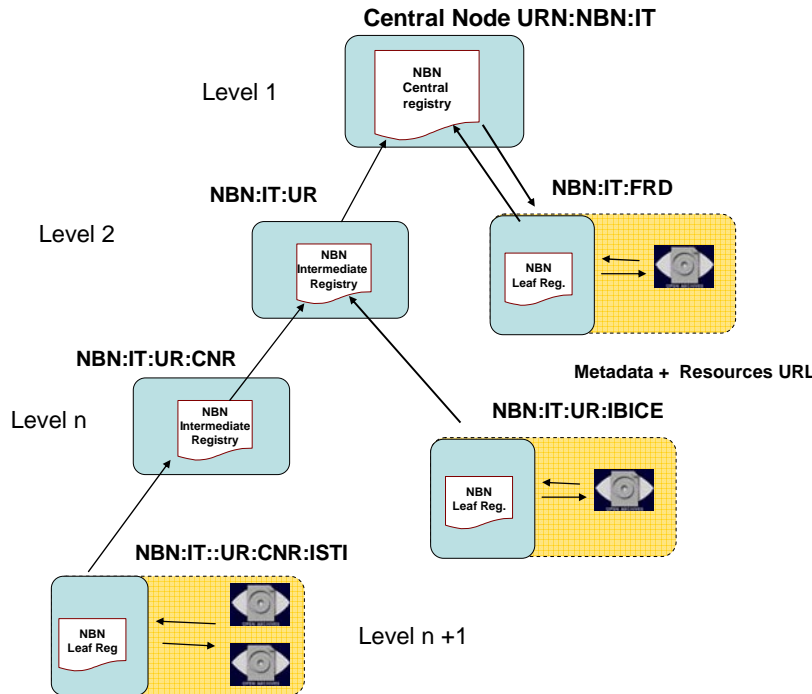
Cada agencia se adhiere a la política definida por el nodo padre y, en consecuencia, define las políticas a que deberán adherirse los nodos hijos.

Resulta fácil ver que este modelo de múltiples niveles distribuido y jerárquico implica que la responsabilidad en la creación y definición de identificadores permanentes puede delegarse de modo recursivo a autoridades encargadas de asignar nombres en niveles inferiores, cada una de las cuales gestionará un parte del espacio de nombres del dominio. Dada la similitud de los problemas a que nos enfrentamos, algunas de las ideas han sido tomadas prestadas del servicio de DNS.

En el marco de nuestra arquitectura cada nodo recolecta información acerca de identificadores permanentes de sus nodos hijos y es capaz de definir directamente todos los identificadores que no pertenezcan a su dominio. Esto implica que cada nodo puede definir cada ítem/objeto digital generado en el sub-espacio de nombres NBN:IT, tanto consultando sus propias tablas como interrogando otros nodos. En este último caso, el resultado de la interrogación se almacena localmente en la memoria caché para agilizar consultas subsecuentes relativas al mismo identificador.

Esta redundancia de de puntos de acceso al servicio y de ubicaciones de los almacenes de información aumenta la fiabilidad de toda la estructura eliminando posibles SPOFs<sup>2</sup>. Junto a ello, la fiabilidad aumenta al crecer el número de instituciones que participan en el proyecto.

En nuestra opinión, una arquitectura distribuida aumenta asimismo la extensibilidad y el rendimiento, al tiempo que mantiene inalterados los flujos de publicación definidos por los distintos repositorios.



## 5.1. Políticas de actuación

<sup>2</sup> Siglas de Single Point of Failure. En informática, cualquier componente de un sistema que, tras un fallo en su funcionamiento, ocasiona un fallo global en el sistema completo. *Nota del traductor.*

La fiabilidad y consistencia de una infraestructura de NBN distribuida sólo puede garantizarse definiendo y aplicando políticas efectivas de actuación. Con este fin, los órganos de gobierno del NBN italiano van a publicar unas políticas generales que deberán ser firmadas por todos los organismos participantes.

Hemos llevado a cabo un análisis inicial para detectar problemas y otros asuntos que deberían ser tenidos en cuenta. En nuestra opinión, cada organismo implicado debería satisfacer una serie de requerimientos, técnicos y organizativos, así como comprometerse a respetar las directrices.

#### Requisitos organizativos

Cada organismo que participe en el proyecto debe presentar a una persona de referencia del ámbito administrativo que sea responsable del cumplimiento de las políticas relativas a los procedimientos de registro y definición así como de las relaciones con los organismos de nivel superior e inferior. También debe presentar a una persona de referencia del ámbito técnico que se responsabilice de la infraestructura de hardware, software y red.

#### Requisitos técnicos

El hardware que aloje un registro/gestor de enlaces de NBN debe albergarse en una infraestructura gestionada para alojarlo, con conexión eléctrica ininterrumpida y una conexión a la red de alta velocidad. Los organismos que no cuenten con una torre de servidores interna pueden subcontratar servicios de alojamiento de un proveedor externo que cumpla con los requisitos técnicos.

La arquitectura de hardware debe ser redundante para garantizar que no haya ni un solo SPOF. En nuestra opinión también sería útil para identificar y supervisar algunos indicadores simples del nivel de servicio, tales como la respuesta del servicio de tiempo en tiempo, así como para definir los umbrales que cada organismo debería respetar. El encargado de la actualización de cada dominio podrá supervisar a sus sub-dominios hijos y notificarles si se han producido violaciones en el nivel de servicio. Las políticas definirán también cómo se actuará ante estas violaciones.

#### Directrices

Las políticas de actuación definirán las reglas relativas a:

- creación de identificadores permanentes bien formados;
- identificación de los recursos digitales que “merecen” un identificador permanente;
- identificación del nivel de detalle de cada recurso de cara a la asignación de identificadores permanentes (artículo, sección de un artículo, libro, capítulo de un libro, etc.)
- inventariado de repositorios con la finalidad de evaluar sus puntos fuertes y débiles (la herramienta Drambora podría ser útil en este sentido).

## **5.2. Elementos de la arquitectura**

Todos los elementos que componen la arquitectura de nuestro sistema incluyen una interfaz web que permita a los operadores llevar a cabo tareas administrativas y a los usuarios finales a definir nombres así como a acceder a metadatos y contenidos. A continuación sigue una descripción de los diferentes componentes:

## **1. Nodo central**

En el punto más alto de la jerarquía se encuentra un nodo central que mantiene el registro central en que se almacenan todos los nombres de NBNs generados por cualquiera de los nodos externos. El nodo central puede asignar subdominios a instituciones que hayan llevado a cabo el procedimiento de registro; puede resolver consultas de usuarios directamente o reenviarlas al organismo indicado de nivel inferior; comprueba las referencias de NBNs recolectadas de registros de subdominios en relación con el cumplimiento de políticas y de univocidad. El nodo central apoya la gestión de una Agencia Central, responsable de la calidad y fiabilidad de las respuestas ofrecidas por la infraestructura completa a las consultas de los usuarios.

## **2. Nodo interno**

La arquitectura define el nodo interno para gestionar subdominios específicos (como NBN:IT:UR en la figura 1) que controlan otros dominios de nivel inferior (como NBN:IT:UR:CNR). Los organismos internos pueden definir sus propias políticas en lo que hace a generar nombres de NBNs o registrar instituciones en relación a dominios específicos (por ejemplo Patrimonio Cultural, Ciencias, Radio y Televisión, etc.). El nodo interno recolecta los registros de los NBNs de los nodos internos y externos que se encuentran bajo su responsabilidad y crea marcas similares a las descritas por el nodo central pero para un grupo más pequeño de recursos. El nodo interno no puede generar nombres de NBNs pero puede definirlos directamente o reenviando las consultas a los nodos pertinentes.

## **3. Nodo externo**

Los nodos externos son responsables de los subdominios del nivel inferior, que son asignados a los organismos que gestionan las bibliotecas digitales propiamente dichas. Recolectan recursos de los repositorios que se encuentran bajo su responsabilidad y generan nombres para los recursos a petición del operador. Estos nodos también pueden definir todos los nombres de NBNs.

## **6. Funciones que hemos puesto en marcha**

En este apartado haremos unas breves descripciones de las principales funciones que hemos puesto en marcha en nuestro software.

### **6.1. Registrar un nuevo nodo**

El proceso de registro de un nuevo nodo es un paso importante en lo que hace a la fiabilidad y credibilidad de la red completa. Es importante definir estrictos procedimientos de autenticación *off-line* (basados en la interacción humana) y *on-line* (basados en la interacción de máquinas) para las instituciones que deseen participar. El consorcio italiano de NBN está desarrollando actualmente estos procedimientos. A día de hoy el software lleva a cabo los siguientes pasos:

- 1) La primera vez que es ejecutado, el nodo debe haber sido configurado como nodo interno o externo y el operador debe insertar la URL del nodo padre además de otra información adicional, así como el correo electrónico del administrador y una descripción de la institución que lo gestiona.
- 2) El software envía esta información al nodo padre a través de un servicio web.
- 3) El nuevo nodo espera la respuesta del nodo central que asigna un sub-espacio de nombres al nuevo nodo. Durante este tiempo de espera, los nodos externos están

habilitados para registrar repositorios OAI y para recolectar registros OAI, pero no pueden generar NBNs.

- 4) Cuando el nodo es reconocido por el nodo padre ya puede llevar a cabo todas las acciones asociadas a su rol.

## **6.2. Gestionar un registro de nodos externos de NBNs**

Cada nodo externo es responsable de recolectar metadatos de publicaciones de uno o más repositorios y de generar identificadores permanentes de NBNs. La recolección se lleva a cabo mediante el protocolo OAI-PMH [5] y permite reunir metadatos en los formatos DC [6], METS [7] y MPEG21 [8], tal y como se presenta en [9].

El administrador del nodo externo puede gestionar la lista de repositorios e iniciar manualmente el procedimiento de recolección, que reúne metadatos de todas las publicaciones depositadas allí tras la última “recolección”. Hecho esto, puede seleccionar nuevas publicaciones a las que asignar identificadores permanentes. Una vez generados, los IP están en un estado “inconsistente”, es decir, que el objeto digital al que han sido asignados puede tener otro IP en un subdominio diferente. Los nodos de niveles superiores recolectan periódicamente metadatos, que son transmitidos por toda la jerarquía hasta que alcanzan el nodo central, que controla la univocidad de los IP (véase la siguiente sección). Si el recurso digital ya tiene asociado un IP, se pide al nodo externo que borre el registro; en caso contrario, el estatus cambia a “consistente”. La comprobación de la univocidad del PI se lleva a cabo mediante la comparación del hashCode<sup>3</sup> de los recursos.

## **6.3. Sincronización entre registros de NBNs**

Como se indicó anteriormente, el nodo central y sus nodos internos recolectan periódicamente nuevos registros de NBNs de sus nodos hijos y ponen al día sus archivos. Los administradores de nodos pueden configurar este procedimiento de sincronización automatizada desde la terminal administrativa definiendo uno o más programadores automáticos de tareas. El administrador, para cada uno de ellos, debe indicar un nombre, el intervalo de sincronización y los registros del subdominio que deben ser recolectados. Un nodo sólo puede recolectar a sus nodos hijos.

El nodo central también verifica si los NBNs recolectados cumplen las políticas definidas, lleva a cabo comprobaciones de univocidad y notifica a los nodos externos si los nuevos registros deben conservarse o borrarse y si su estado debe cambiar a “consistente”.

## **6.4. Gestión de nombres de NBNs**

Puede consultarse a cada nodo que forme parte de la infraestructura del NBN para definir un identificador de NBNs. Si el nodo interrogado es capaz de hacerlo directamente, responde inmediatamente al usuario; de otro modo, analiza la URN para identificar el nodo externo responsable del subdominio asociado. Si el nodo interrogado conoce la dirección del nodo externo, trata de contactarlo a través de un servicio web

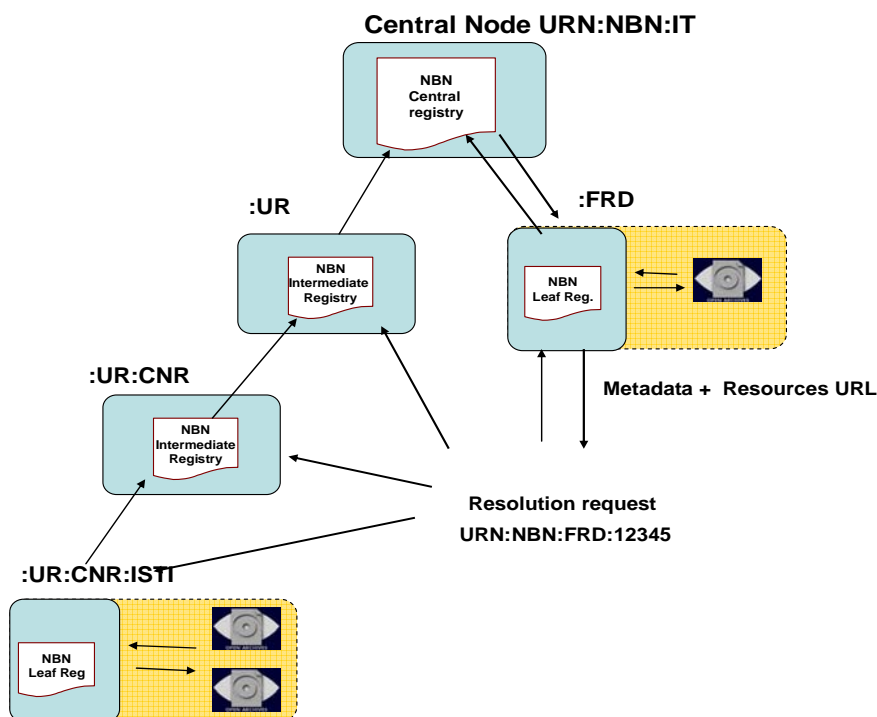
---

<sup>3</sup> Durante el procedimiento de recolección, los nodos externos traen los archivos que contienen los recursos digitales y usa su corriente binaria como introducción de datos (input) para una función de hash. La función de hash devuelve, para cada archivo, el denominado código de hash, que puede considerarse una especie de huella dactilar del archivo. Si dos archivos tienen el mismo código de hash hay una probabilidad muy alta de que su contenido sea idéntico. La probabilidad depende de la función de hash que se haya adoptado.

particular. Si el nodo externo está en funcionamiento, el nombre es interpretado y se pasa la respuesta al usuario; en caso contrario, el nodo interrogado escala la jerarquía hasta encontrar un servidor que sea capaz de definir el nombre.

El procedimiento que acabamos de describir es factible porque cada nodo conserva información acerca de los otros nodos (al menos acerca de algunos de ellos) y por ello es capaz de reenviar peticiones. A día de hoy, es tarea del administrador insertar manualmente esta información en la configuración de los nodos, pero se están estudiando mecanismos automáticos para retransmitir información topológica de la infraestructura en intervalos fijados.

El usuario puede enviar una petición de gestión bien usando un formulario disponible en la interfaz de los nodos, bien especificando una URL http particular en el buzón del buscador. Hemos desarrollado asimismo un plugin de Mozilla que permite que el buscador acepte URNs de NBN.



## 6.5. Reubicación de objetos digitales

Cuando se transfiere el contenido de una biblioteca digital (o parte de él) a otra biblioteca digital, surge un problema para la infraestructura encargada de los NBNs, ya que la permanencia del identificador es un requisito obligatorio del mismo. En estos casos, la comparación con el DNS no aporta ninguna ayuda. En este sentido, deben considerarse dos posibles situaciones:

- la biblioteca digital desaparece y ninguna otra está legitimada (al menos temporalmente) para albergar sus objetos digitales;
- uno o varios objetos digitales son transferidos de una biblioteca a otra que pertenece a un subdominio diferente.

En el primer caso, nuestro software guarda el registro del NBN y cambia su status a “eliminado”. El nombre aún puede resolverse, pero no es posible acceder al recurso. Si ese documento es absorbido de modo subsecuente por otra biblioteca y su registro es

recolectado por un nodo externo, la infraestructura detecta mediante una comparación de hash que el objeto ya posee un identificador permanente de NBN y no crea uno nuevo. Cuando esto ocurre, no sólo puede resolverse el antiguo nombre sino que también puede accederse de nuevo al recurso. En el segundo caso, el IP no cambia. El archivo que generó en primer lugar el identificador simplemente redirige las consultas al nuevo encargado de la actualización. En ambos casos nuestra infraestructura conserva y resuelve el IP de NBN asignado en un primer momento, logrando de este modo que este sea permanente en la práctica.

## **7. Actividades de prueba**

Después de desarrollar un primer prototipo, se han establecido acuerdos de colaboración con diferentes instituciones dedicadas a la investigación para crear una comunidad en que se encuentren representados tanto los usuarios finales como los encargados de desarrollar el software. Actualmente hay muchas instituciones implicadas en la definición de los requisitos de los usuarios y otras muchas han declarado estar disponibles para unirse a la red de NBN. Estas instituciones son: el grupo GRIDS (ISS, INAF, INFN, INGV, ASI, ENEA, INOA, APAT, Università degli Studi di Pisa, Università degli Studi di Roma “La Sapienza”), Universidad de Florencia, Florence University Press.

Se ha puesto en marcha un primer banco de pruebas en que los usuarios pueden ejecutar casos de prueba y dar su opinión al respecto a los diseñadores del sistema en forma de notificaciones de errores y defectos, peticiones de cambios o mejoras y nuevos requisitos. Por otra parte, los diseñadores del sistema están llevando a cabo pruebas técnicas para evaluar el funcionamiento, la extensibilidad y la fiabilidad de la infraestructura y llevar a cabo las acciones necesarias para satisfacer las indicaciones de los usuarios.

Dicho banco de pruebas está configurado del siguiente modo:

- a) un nodo central en la BNCF, responsable del subdominio italiano (NBN:IT),
- b) un nodo interno de segundo nivel en el CNR, responsable del subdominio “Universidad e Investigación” o “University and Research” (NBN:IT:UR),
- c) un nodo externo de segundo nivel en el FRD responsable del subdominio local NBN:IT:UR:FRD,
- d) un nodo externo de tercer nivel en la UNIMI, responsable del subdominio local NBN:IT:UR:UNIMI,
- e) un nodo externo de tercer nivel en el CNR, responsable del subdominio local NBN:IT:UR:CNR.

El nodo interno de segundo nivel ubicado en el CNR tiene el objetivo de desarrollar el Registro Nacional de Universidades e Investigación. Actualmente agrega los registros creados por los nodos externos de la UNIMI y el CNR para los recursos almacenados en un repositorio local Dspace. Se ha llevado a cabo un primer conjunto de pruebas para verificar las funcionalidades y el comportamiento del sistema en un espacio distribuido usando distintos conjuntos de metadatos. El funcionamiento no ha sido el aspecto principal a tener en cuenta en esta fase, motivo por el cual los servidores utilizados para armar la infraestructura no son especialmente potentes ni están totalmente actualizados.

Las primeras reacciones de los usuarios han sido positivas en lo que hace a las funcionalidades relativas al registro y resolución. El sistema recolecta recursos, asigna NBNs y facilita el acceso a metadatos y documentos tal y como se había planeado. En lo relativo a la detección de duplicados mediante comparaciones de hash, se ha indicado que dicho mecanismo funciona únicamente si los archivos objeto de la comparación son idénticos, pero falla si difieren aunque sea en un solo bit. Se ha comentado asimismo que, actualmente, no es posible representar en el identificador la relación “parte de” entre dos objetos digitales (es decir, la que indica cuándo uno es parte de otro). Por lo tanto, si deseamos asignar identificadores tanto a un documento completo como a partes del mismo (por ejemplo, una imagen), a día de hoy no existe un modo acordado para representar esta relación de inclusión en la parte final del identificador permanente. Finalmente, diversos usuarios han expresado la necesidad de que se desarrollen servicios de más alto nivel, principalmente la posibilidad de elaborar informes acerca del número de publicaciones albergadas en un subdominio en un periodo de tiempo determinado. Este problema está estrechamente relacionado con la detección de duplicados. Si este último no se soluciona, las estadísticas de contabilidad de los recursos pueden verse afectadas por errores cuyo impacto no puede aún calcularse en este momento.

Se han llevado a cabo algunas mediciones para probar el comportamiento del sistema cuando se llevan a cabo recolecciones masivas en diferentes conjuntos de metadatos. Concretamente, hemos recolectado 1000 recursos digitales en cada prueba, usando metadatos de Dublin Core y METS. Un nodo externo recolectando información de una biblioteca digital necesita unos 25 minutos para esa tarea usando Dublin Core y 11 minutos usando METS. Los valores absolutos no son relevantes en este contexto porque los servidores no eran ni recientes ni potentes y no se llevaron a cabo adaptaciones a tal efecto en ningún componente de software o hardware. El aspecto más relevante es el hecho de que el tiempo requerido para el procedimiento si usamos METS es menos de la mitad del necesario cuando se usa Dublin Core. Esto se debe a que METS incluye por defecto el código de hash MD5 que se utiliza para realizar comprobaciones de univocidad, lo que evita que el sistema tenga que calcularlas. Es importante resaltar que debería especificarse el MD5 también en Dublin Core usando un campo personalizado (custom field).

El tiempo medio que necesita un nodo interno para recolectar 1000 registros de un nodo hijo se ha medido en unos 40 minutos. Este valor se ve fuertemente influido por el control de univocidad, que depende del rendimiento del DBMS subyacente. Esto puede mejorarse definiendo índices para los campos involucrados y, obviamente, usando una configuración de hard- y software de alto rendimiento.

## **8. Conclusiones y actuaciones futuras**

En esta presentación hemos descrito una nueva aplicación de software para una infraestructura de registro y gestión de enlaces de NBN jerárquica y distribuida. Los primeros exámenes que se han llevado a cabo al respecto han mostrado la viabilidad de este modelo. Junto a ello, el grupo de instituciones que apoya el proyecto incluye tanto importantes organismos de investigación italianos como bibliotecas nacionales.

Los principales problemas técnicos con que nos enfrentamos actualmente tienen que ver con la garantía de que cada identificador sea unívoco. La solución propuesta, usar códigos de hash MD5, resuelve en parte el asunto pero plantea problemas de

rendimiento y no es aplicable a casos en que el mismo contenido se vea representado en distintos formatos. Una solución más amplia deberá abarcar probablemente la comparación de un grupo de metadatos definido estrictamente. Esto implica que deban definirse reglas estrictas y responsabilidades claras en lo relativo a la incorporación de datos a las bibliotecas digitales.

Desde un punto de vista político, el objetivo a corto plazo es aumentar el grupo de instituciones que apoyan el proyecto para crear un primer núcleo de una infraestructura nacional de NBN digna de crédito. A mayor escala, se han establecido contactos con el consorcio de Intercambio de Conocimiento (Knowledge Exchange) para cooperar en el desarrollo de un gestor de enlaces de NBN europeo de nivel superior. En los próximos meses se considerará la adopción de nuestro software como gestor de nodos de nivel superior.

En nuestra opinión, también es importante identificar servicios de valor añadido de alto nivel (tales como contabilización de recursos digitales) que puedan llevarse a cabo en la parte superior de la infraestructura. Esto probablemente favorecería la difusión de los identificadores permanentes NBN.

Desde un punto de vista técnico, los próximos pasos a seguir incluyen el análisis del rendimiento y la puesta a punto (del sistema) para definir los requisitos de hardware necesarios para crear una infraestructura que pueda garantizar los niveles de servicio necesarios.

Se aumentará el ámbito de análisis del banco de pruebas para incluir un nodo externo instalado en la Universidad de Bolonia que recolectará registros del repositorio del proyecto “Magazzini digitali”. La meta de este proyecto es lograr que la biblioteca digital de la BNCF sea capaz de recolectar tesis doctorales del repositorio de e-prints de la Universidad de Bolonia, con la finalidad de cumplir con ellas el depósito legal. En este caso, los recursos ya tienen un nombre de NBN. Se creará un nuevo registro de NBN en nuestro archivo usando el identificador existente, al que se asociará una nueva URL asignada por el depósito legal en la BNCF.

Se ha establecido asimismo un grupo de investigación para examinar de modo exhaustivo el problema de los duplicados y sus posibles soluciones. En este ámbito, quizá otros códigos de hash distintos del MD5 proporcionarían un mejor rendimiento en lo que hace a las operaciones de comparación. El mismo grupo se encargará del problema de la representación de la relación “parte de”.

Finalmente, planeamos investigar posibles vías para establecer conexiones permanentes y fiables entre los NBNs y otros identificadores permanentes como el DOI que pudieran favorecer la implementación de un gestor de enlaces global de estándares múltiples.

### **Agradecimientos**

Queremos agradecer a Luciana Truffelli de la Agenzia Spaziale Italiana (ASI) por su contribución a las actividades descritas en el presente artículo.

## Bibliografía

- [1] European Digital Library  
<http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edlproject/>
- [2] Europeana [www.europeana.eu/](http://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. Diciembre 2004, volumen 10, número 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
- [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](http://www.handle.net)
- [13] Persistent URL <http://purl.oclc.org>
- [14] Library of Congress Control Number <http://www.loc.gov/marc/lccn.html>
- [15] David Giarretta, número 1, volumen 2 | 2007 The CASPAR Approach to Digital Preservation 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 en 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](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 Identifiers, 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](http://en.wikipedia.org/wiki/Handle_System)