

Pubblicazioni on line: Digital preservation e identificatori persistenti persistenti

Il National Bibliographic Number (NBN)



La Fondazione Rinascimento Digitale FRD

La Fondazione Rinascimento Digitale opera in tutti i settori collegati alla produzione, alla conservazione e alla diffusione delle memorie digitali attraverso:

- ricerca e sperimentazione
- conservazione e restauro delle memorie digitali
- utilizzazione, con particolare riferimento ad applicazioni fortemente innovative
- trasferimento delle conoscenze, formazione e consulenza

Le attività della Fondazione si sviluppano su quattro specifici filoni:

- a) produzione di contenuti digitali
- b) gestione e accessibilità delle risorse digitali
- c) preservazione delle memorie digitali
- d) e-services per applicazioni culturali e scientifiche

I principali progetti

Le tematiche affrontate spaziano dalla produzione di contenuti digitali alla gestione e accessibilità delle risorse digitali, dalla preservazione delle memorie digitali allo studio di e-services per applicazioni culturali e scientifiche.

- Accesso alle biblioteche digitali
- Archivio Integrato - Pinakes 3
- Archivi sonori e della Musica
- Digital Libraries Applications
- Digital Preservation Europe (DPE)
- Digitalizzazione di documentazione archivistica
- Magazzini Digitali
- **National Bibliographic Number**
- Opere GDS Uffici in digitale ←
- OptiMA
- Studio dell'impatto delle sorgenti luminose
- VIDI-Video

Perché servono i Persistent Identifiers

Le risorse su Internet hanno una vita molto corta e la loro **Identificazione/reperibilità** porta problemi molto complessi che influenzano negativamente le attività di ricerca e di citazione delle risorse scientifiche/culturali da parte delle comunità di utenti.

Qualora l'indirizzo web del testo dovesse cambiare per motivi di ordine tecnico e/o amministrativo – allora qualsiasi citazione in rete collegata con un link di quel testo smetterebbe di essere valida. Infatti cliccando sul link che rinvia a quell'indirizzo web, otterremmo in risposta il messaggio di errore “HTTP 404 - File not found”.

Statistiche

- 16% links after 6 months (T.Dowling)
- 50% after 24 months (British Library)
- 87% since 1998 (OCLC) are **broken links**
- 13% of article publications: after 27 months not available ([Science](#))

I principali sistemi di PI

a) **URN** (Uniform Resource Name)

- URN:namespace_identifier: namespace_specific_string
- non è direttamente azionabile (URN:NBN)

b) **PURL** (Persistent Uniform Resource Locator)

- solo risoluzione uno-a-uno

c) **Handle system**

- Architettura distribuita Global registry + Local registries
- open e gratuita

d) **DOI** (Digital Object Identifier)

- non solo identificatori ma anche metadati
- risoluzione 1-n + collezione di risorse (CROSSREF)

e) **ARK** (Archival Resource Key)

- nomi gerarchici risolti attraverso la ricerca in un file
- risorsa, metadati, politica dell'organizzazione

f) **InfoURI**

- solo identificazione
- nessuna risoluzione e persistenza

g) **OpenURL**

- link azionabile che trasporta metadati su un oggetto
- meccanismo di risoluzione 1-n

PIF (Persistent Identifiers Interoperability Framework)

FP7 call 3 challenge 4.3 (Digital Library)

Persistent Identifiers (PI) of Internet resources is a crucial issue for all the information society. In particular, in the cultural/scientific digital libraries instability of URLs reduces trustability and use of Internet as a platform for research. Many technologies have been proposed to support PI but we have identified some weak points in the current scenario:

- *we cannot expect/impose a global unique technology*
- *trustability / longevity of the institution offering the service is crucial*
- *granularity of what a PI must point to is widely different in each area*
- *user interface must be unique for different PI systems*

To overcome these weaknesses, an interoperability framework is needed.

Il progetto NBN - Obiettivo

- a) Sperimentare una nuova tecnologia orientata alla comunità dei Beni Culturali basata su **Uniform Resource Name** (URN) utilizzando il **National Bibliographic Number** (NBN) come namespace di riferimento.
- b) Sviluppare un prototipo open source per l'implementazione di un servizio di identificazione delle opere culturali digitali su Internet,
- c) Facilitare l'accesso alle risorse da parte dei cittadini e sensibilizzare/vincolare le istituzioni culturali ad applicare programmi di **digital preservation** per le proprie risorse digitali.
- d) Incentivare attività di ricerca e sviluppo in questo settore promuovendo lo sviluppo delle **EDL** (European Digital Library) e l'interoperabilità tra gli attuali sistemi di identificazione. In particolare, sperimentare la interoperabilità con altri NBN nazionali.

Biblioteca Nazionale di Firenze, Florence University Press
CILEA, Università di Milano

National Bibliographic Number

L'iniziativa per la definizione di un identificativo persistente per risorse digitali che accomuni i vari identificativi usati dalle biblioteche nazionali, nasce dal **CDNL** (the Conference of Directors of National Libraries) che ha assunto la responsabilità della gestione del namespace **NBN** per le biblioteche nazionali ed ha indicato la Library of Congress come detentore del registro internazionale.

Il **National Bibliographic Number** è stato inizialmente inteso come “un nome generico che si riferisce ad un gruppo di sistemi di identificazione utilizzati dalle Biblioteche Nazionali e solo da loro per l'identificazione di pubblicazioni depositate a cui manca un identificativo o metadati descrittivi che descrivono la risorsa” (RFC 3188 Cap 3.1).

Il progetto NBN – User requirements

- **Gestione decentralizzata:** Ogni istituzione gestisce i propri nomi associati alle risorse e si assume la responsabilità di mantenere il registro di nomi creati del sub-domain e il servizio di risoluzione.
- **Risoluzione gerarchica:** Se l'istituzione a cui si fa richiesta di risoluzione non possiede la risorsa, chiama il nodo centrale il quale potrà rispondere direttamente alla richiesta o inoltrare la richiesta all'istituzione responsabile per quel nome. **Peer to Peer:** I nodi periferici possono inoltrare direttamente la richiesta di risoluzione all'istituzione competente senza passare dal nodo centrale.
- **Back-up centrale:** Il nodo centrale con un protocollo di harvesting raccoglie sia tutti i nomi generati dai sub-domain, sia copia delle risorse digitali.

Il progetto NBN – User requirements

- **Associazione multipla:** Ogni risorsa ha un solo NBN associato. Se viene generato un NBN da istituzioni differenti per una copia di una risorsa che ha già un nome, il nodo centrale gestirà le “**collisions**” tramite un sistema di segnalazione in modo che le copie possono far riferimento al NBN già presente tramite una associazione multipla URN-URLs
- **Aggiornamento inter-livello:** La sincronizzazione dei nuovi record NBN generati dai periferici verso il registro centrale avviene tramite OAI-PMH. Attualmente è previsto un polling ad intervalli prefissati, ma è allo studio un protocollo di segnalazione per azioni PUSH dai nodi periferici verso il nodo centrale.
- **Metadati del registro:** Scelta di non duplicare i metadati descrittivi delle risorse definendo per il registro un set di metadati esclusivamente per la gestione del record come l’hash MD5 per identificare eventuali copie della risorsa o lo storico delle modifiche al record (lo spostamento della risorsa che produce un cambio nel campo URL del record).

