



Consiglio Nazionale delle Ricerche



# **Identificativi Persistenti: standard, policy, infrastrutture**

**Roberto Puccinelli, Massimiliano Saccone, Luciana Trufelli**

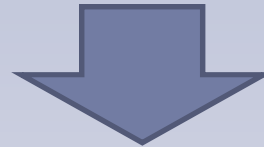
# Sommario

- **Contesto e problematiche**
- **Introduzione agli identificativi persistenti**
- **Standard per l'identificazione persistente**
  - **PURL**
  - **ARK**
  - **Handle**
  - **DOI**
  - **NBN**
- **Esperienze dell'Amministrazione Centrale  
CNR**



# Contesto

**Esplosione della comunicazione e  
dell'informazione in rete**



**Indiscutibili vantaggi**

**ma anche**

**diversi problemi che solo ora cominciano a  
essere percepiti dal pubblico non  
specialistico.**



## Problemi e sfide

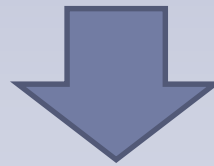
- **Information overloading**
- **Multi version control**
- **Digital preservation**
- **Interoperabilità tra i diversi sistemi informativi**

**L'identificazione persistente delle risorse informative è un fattore abilitante per tutti questi ambiti**



# Information overloading

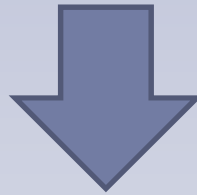
**Troppa informazione non ritagliata sulle esigenze delle singole classi di utenti**



**L'identificazione persistente associata a meccanismi di reperimento della risorsa e dei metadati associati fornisce uno strumento per la selezione di contenuti informativi**

# Multiversion control

**Più versioni della stessa risorsa digitale**



**Un opportuno meccanismo di identificazione persistente può fornire un collegamento tra le diverse versioni, tracciandone al contempo la diversità.**

# Digital preservation

**La conservazione delle risorse digitali,  
specie nel lungo termine,  
non può prescindere da meccanismi di  
identificazione persistente**

# Interoperabilità

**Collegamento tra diversi sistemi possibile attraverso:**

- Authority file condivisi**
- Identificazione persistente**
- ...**



## Altri ambiti

- **Certificazione dei contenuti e della provenienza.**
- **Accounting delle risorse informative.**

Es.: valutazione della ricerca



## Identificazione, certificazione, localizzazione e conservazione permanente delle risorse informative

Le risorse informative su Internet hanno generalmente una **vita molto breve** e la loro Identificazione/reperibilità permanente pone problemi complessi che riguardano non solo questioni tecnologiche, ma anche e soprattutto questioni **politico-organizzative**, che implicano la gestione, la selezione, il reperimento e la preservazione delle risorse informative.

Questi problemi influenzano negativamente le attività di **ricerca e di citazione delle risorse informative** scientifiche/culturali da parte delle diverse comunità di utenti.

Qualora l'indirizzo web della risorsa informativa dovesse cambiare per motivi di ordine tecnico e/o amministrativo – allora qualsiasi citazione in rete collegata con un link alla risorsa smetterebbe di essere valida. Infatti, cliccando sul link che rinvia all'indirizzo web, otterremmo in risposta il messaggio di errore "**HTTP 404 - File not found**" e, in assenza di ulteriori informazioni, non saremo più in grado di reperire alcuna informazione sulla risorsa.

### 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)



## Possibili Soluzioni

**Una delle soluzioni più affidabili per risolvere questi problemi è quella di associare un Persistent Identifier (PI) alla risorsa digitale, che rimarrà perennemente associato alla risorsa stessa indipendentemente da dove essa sia collocata e/o localizzata.**

**In altri termini, per gestire e rendere effettivamente fruibili le risorse informative distribuite in rete e per garantirne la preservazione duratura, è necessario disporre di sistemi che consentano di identificare in maniera, chiara, univoca, non ambigua, ma soprattutto stabile nel tempo tali risorse.**



# Identificatori Persistenti

- **Definizione:** Un Identificatore Persistente (PI) è una sequenza di caratteri alfanumerici associata in modo univoco e permanente ad una qualsiasi risorsa digitale (articolo, libro, file multimediale, etc.), che ne consente l'identificazione (problema dell'overload informativo - mancanza di trasparenza da parte dei gestori dei motori di ricerca nei criteri di organizzazione e di ordinamento delle risorse informative).
- **Scopo:** l'associazione di un Identificatore Persistente (PI) ad una risorsa digitale (con i relativi metadati) può essere usata per certificarne - in modo univoco e permanente - l'autenticità, la provenienza ed i diritti di autore, per garantirne la localizzazione persistente e la *long term preservation* (problema della "volatilità" delle informazioni sul Web).
- **Affidabilità:** l'effettiva persistenza dei sistemi di identificazione **può essere conseguita solo attraverso accordi vincolanti (Policy)**, sottoscritti dalle organizzazioni che assegnano, gestiscono e conservano gli identificativi, fornendo, inoltre, servizi di risoluzione che consentano il reperimento/identificazione della risorsa e/o dei metadati ad essa associati indipendentemente da eventuali cambiamenti di localizzazione.
- **Comunità di utenti:** le comunità di utenti interessate all'uso di identificatori persistenti attualmente adottano soluzioni basate su standard differenti → Digital Object Identifier (DOI)(usato prevalentemente nell'ambito dell'editoria commerciale), Archival Resource Key (ARK), Handle System (usato anche nell'ambito dell'Open Archives Initiative), URN (Uniform Resource Name), PURL, etc. (problema dell'interoperabilità tra sistemi e servizi identificativi).



# Uniform Resource Identifiers

Un URI (Uniform Resource Identifier) è una **stringa che identifica univocamente una risorsa generica** (una risorsa direttamente accessibile in rete, una risorsa non direttamente accessibile in rete, un concetto astratto), **non fornendo necessariamente accesso alla risorsa identificata.**

Un URI è in pratica una stringa avente la seguente **sintassi:**

**URI = scheme “:” scheme-specific-part**

Ogni URI inizia con la parte **scheme** contenente il nome dello specifico URI scheme (lo **schema che definisce le specifiche per assegnare gli identificatori al suo interno**) seguito dai “:” e dalla parte specifica dello schema definita nell’ambito dell’URI scheme di riferimento.

# Uniform Resource Identifiers

La sintassi è dunque organizzata in modo gerarchico, con i vari componenti disposti in ordine decrescente di significato da sinistra a destra:

URI = scheme “:” [“//” authority] path [“?” query] [“#” fragment]

Gli elementi significativi, oltre allo scheme, sono l'**authority** (elemento gerarchico opzionale che identifica il **naming authority al quale è affidato il namespace definito dal resto dell'URI**), il **path** (elemento obbligatorio contenente i dati organizzati generalmente in forma gerarchica, che insieme ai dati contenuti nella componente query permettono di identificare la risorsa), il **componente opzionale query**, che contiene dati identificativi non gerarchici, il **componente opzionale fragment**, che permette un'identificazione indiretta di una risorsa secondaria per mezzo del riferimento alla risorsa primaria e di ulteriori dati identificativi (la risorsa secondaria può essere una porzione o sottogruppo della risorsa primaria o un'altra risorsa definita o descritta dalla risorsa primaria).



# URL

Un URI può essere teoricamente classificato come URL o come URN.

Un URL (**Uniform Resource Locator**) è un URI che, oltre a identificare una risorsa, fornisce i mezzi per agire su o per ottenere una rappresentazione della risorsa, descrivendo il suo meccanismo di accesso primario o la sua "ubicazione" ("location") in una rete. Per esempio, l'URL `http://www.cnr.it` è un URI che identifica una risorsa (l'homepage del CNR) e lascia intendere che una rappresentazione di tale risorsa (il codice HTML della versione corrente di tale homepage) è ottenibile via HTTP da un host di rete chiamato `www.cnr.it`.

Gli URL sono utilizzati per localizzare una risorsa fornendo una identificazione astratta della localizzazione della risorsa. **La commistione tra identificazione e localizzazione, che è alla base di questo standard, pone diversi problemi.**

# Limiti degli URL

Ristrutturazioni tecniche e amministrative rendono inevitabile cambiare, prima o poi, i nomi di dominio dei server sui quali sono allocate le risorse digitali, inoltre, può sorgere la necessità di trasferire le risorse da un server a un altro con un altro nome di dominio. In tutti questi casi, nel momento in cui si cambia l'URL, i documenti vengono di fatto ridenominati. L'uso dell'URL non può essere considerato un approccio affidabile per risolvere le questioni legate alla persistenza della risorsa informativa, a causa dell'**instabilità strutturale dei link** (ad esempio domini non più disponibili) e delle risorse collegate (rilocalizzazione o aggiornamento). Gli URL comunemente intesi sono quindi inerentemente non persistenti.

Il corrente utilizzo dell'indirizzo URL accresce il rischio di non recuperare le risorse informative o di sottoutilizzarle. Nel settore dei Beni Culturali e della Ricerca è essenziale non soltanto identificare una risorsa, ma anche garantire un accesso continuo nel tempo nonché una certificazione dell'autenticità e dell'integrità della risorsa.





# Uniform Resource Name

Un URN (Uniform Resource Name) è un **URI che identifica in modo univoco, globale e persistente una risorsa mediante un "nome" in un particolare dominio di nomi ("namespace"), indipendentemente dalla sua localizzazione.**

Un URN può essere usato per parlare di una risorsa senza far riferimento alla sua ubicazione o come ottenerne una rappresentazione. Per esempio, l'URN `urn:isbn:0-395-36341-1` è un URI che consente di individuare univocamente un libro mediante il suo nome `0-395-36341-1` nel namespace dei codici ISBN, ma non suggerisce dove e come possiamo ottenere una copia di tale libro.

Lo scopo di URN è quello di fornire un identificatore (nome) globalmente univoco usato per identificare una risorsa ed eventualmente accedere ai suoi metadati e/o alla risorsa stessa.

# Gli URN come identificatori persistenti

Gli URN sono utilizzati come identificatori persistenti della risorsa, indipendenti dalla localizzazione e sono progettati per semplificare la mappatura di altri namespace (che condividono le proprietà degli URN).

La sintassi degli URN fornisce un mezzo per codificare dati con caratteri in una forma che può essere veicolata tramite i protocolli esistenti.

Allo stato attuale, la mancanza di un qualsiasi supporto infrastrutturale a livello internazionale in questo ambito determina che **ogni implementazione dello standard URN comporti lo sviluppo di un proprio sistema di risoluzione.**



## Persistent identifier: classificazione operativa

Si può tentare, quindi, una **classificazione** degli identificatori persistenti, non in base allo standard a cui si conformano (URI, URN, ecc.), ma piuttosto **secondo il sistema di risoluzione che utilizzano** (il sistema attraverso cui l'identificatore si configura come l'input (richiesta) di un servizio in rete, per ricevere di ritorno uno specifico output (risorsa, metadati, risorsa e metadati, ecc)).

Su questa base possiamo concretamente distinguerli in due principali tipologie: **identificatori basati sul tradizionale sistema dei DNS** (Domain Name System); **identificatori basati su un name system alternativo al DNS** (I'Handle System).



## Identificatori DNS-based: PURL

Tra gli identificatori persistenti che si basano sul tradizionale DNS, il più noto è senz'altro il PURL (**Persistent Uniform Resource Locator**) **sviluppato dall'OCLC** (Online Computer Library Center).

Nella sostanza, il PURL è un tradizionale indirizzo web. **Quello che il PURL offre in più, rispetto ad un comune URL, è un servizio sistematico di reindirizzamento.**

# Identificatori DNS-based: PURL

**Funzionalmente un PURL è un URL che punta a un servizio intermedio di risoluzione.** Il PURL, in risposta alle richieste del browser, invia un normale URL che può poi essere utilizzato per accedere alla risorsa remota in modo tradizionale. Nel linguaggio web questo è un comune servizio di redirect. Qui non c'è niente di nuovo: il redirect è un servizio già previsto dallo standard HTTP e comunemente usato. Ciò che differenzia un PURL da un comune indirizzo WEB, è l'uso sistematico del servizio di *redirect* al fine di garantire la persistenza dell'identificatore (o almeno fornire un servizio che assomigli il più possibile ad un servizio di identificazione persistente). L'URL associata al PURL può cambiare, il PURL no. La persistenza consiste unicamente in questo servizio, in quanto i PURL di per sé sono a tutti gli effetti URL.

# Sintassi del PURL

A differenza del comune URL, il PURL è composto di tre parti: 1) l'identificatore di protocollo; 2) l'indirizzo di resolver, 3) il nome della risorsa.

<http://purl.oclc.org/3214..>

Da notare che l'indirizzo del resolver è rappresentato dal nome del dominio (ovvero dall'indirizzo IP) del resolver PURL: questa parte del PURL continua ad essere risolta dal tradizionale DNS.



## Creare un PURL

Per creare un PURL occorre in primo luogo essere un **utente registrato di un resolver PURL**, ad esempio, quello dell'OCLC. Ci si registra, alla stessa stregua di tanti altri siti web, immettendo nella form opportuna i propri dati e creando un proprio identificativo e una password. Un utente registrato può quindi creare i propri PURL purché rispetti **determinate condizioni**:

- **In primo luogo che esista il dominio top-level dell'indirizzo web che si vuol rendere persistente.**
- **La parte del PURL che costituisce il nome della risorsa viene decisa dal creatore del PURL. La componente relativa all'indirizzo del resolver, no.**
- **I nomi possono essere arbitrari e non è necessario che vi sia alcuna correlazione tra il nome all'interno del PURL e l'URL associato con esso.**

# Utilizzo e limiti dei PURL

Gli utenti possono assegnare un PURL a ogni risorsa discreta per la quale si desidera un accesso affidabile nel tempo. Ad esempio una homepage, un periodico on-line, un singolo articolo, uno scritto, sono tutti oggetti digitali suscettibili di ricevere un identificatore PURL. Invece, **le risorse che non vivono isolatamente**, come le sezioni di un documento, ad esempio i grafici o gli schemi che non avrebbero senso al di fuori del documento che li contiene, **non sono adatte ad essere identificate tramite un PURL.**

**I PURL non vengono aggiornati automaticamente quando l'URL associato ad essi varia.** L'aggiornamento delle informazioni presenti nel server PURL **è un compito di cui deve farsi carico un *maintainer*.** La manutenzione del PURL viene svolta connettendosi al resolver PURL e quindi compilando l'opportuna form di manutenzione. Solo i *maintainer* autorizzati possono modificare un PURL. **Se un *maintainer* dovesse eliminare il PURL, il resolver, invece di fornire una destinazione, fornirebbe la storia di quel PURL (gli indirizzi associati ad esso fino all'ultimo giorno in cui era operativo).**





## Identificatori basati sul tradizionale sistema dei DNS (Domain Name System)

Un altro persistent identifier basato sul sistema DNS, abbastanza diffuso soprattutto nell'ambito delle digital libraries, è L'ARK (**Archival Resource Key**). Si tratta di uno schema di identificazione per l'accesso persistente agli oggetti digitali (incluse immagini, testi, data set, e aiuti alla ricerca) **sviluppato dalla National Library of Medicine, attualmente utilizzato dalla California Digital Library (CDL) per le proprie collezioni digitali.** L'ARK è l'unico identificatore che antepone gli aspetti procedurali del servizio di identificazione permanente, agli altri aspetti più prettamente tecnologici: **un principio fondamentale dell'ARK è che la persistenza è una pura materia di servizio e che non è inerente all'oggetto né viene conferita a questo da una particolare sintassi di identificazione.**

In concreto l'ARK è un normale URL, caratterizzato da una forma particolare, che gli conferisce alcune delle condizioni necessarie per l'identificazione persistente.

# Sintassi dell'ARK

[<name mapping authority>/]ark:/<name assigning authority number>/<nome della risorsa><qualificatore della risorsa>

**Name Mapping Authority:** indirizzo web convenzionale dell'host che risolverà, in maniera tradizionale, l'identificatore.

**Name Assigning Authority Number:** Identificativo numerico dell'autorità che assegna l'ark.



## Caratteristiche dell'ARK

La particolarità dell'ARK sta nel **connettere tre cose: l'oggetto identificato, i suoi metadati e l'impegno del gestore dell'identificatore circa la sua persistenza.**

Quando un utente inserisce l'ARK nel campo degli indirizzi del browser, l'identificatore conduce l'utente all'oggetto identificato. Lo stesso ARK, seguito da un **singolo punto interrogativo (?)** restituisce un **breve record di metadati** che è leggibile tanto dall'uomo che dalla macchina. Quando l'ARK è seguito da **due punti interrogativi (??)** i metadati in risposta contengono **l'assunzione di responsabilità del gestore.** Questo ci conduce alla enunciazione dei tre requisiti che deve possedere un ARK. Il primo requisito risiede proprio nel legame che l'identificatore stabilisce tra un oggetto e un'assunzione di responsabilità per la manutenzione di quell'oggetto. Il secondo requisito è che l'ARK deve fornire all'utente un link dall'oggetto a una descrizione di questo. A questo scopo l'ARK si avvale del **set di metadati METS** (Metadata Encoding and Transmission Standard) mantenuto dalla Library of Congress. Il terzo requisito è quello di garantire l'accesso permanente agli oggetti identificati.

# Limiti dell'ARK

- L'ARK trova la sua principale limitazione nel fatto che **l'accesso è basato esclusivamente sulla risoluzione tramite il tradizionale sistema DNS**. In altri termini: Il fatto di premettere «http://» e un NMA a un ARK, è un modo per creare un identificatore azionabile tramite un metodo che è esso stesso temporaneo. Ipotizzando che un giorno non sia più disponibile l'infrastruttura che supporta il recupero dell'informazione tramite HTTP, l'ARK dovrà essere allora convertito in qualche nuovo tipo di identificatore azionabile.



## Identificatori basati su un name system alternativo al DNS (l'Handle System)

Una delle strade più promettenti per risolvere il problema dell'identificazione persistente degli oggetti digitali è, probabilmente, quella che consiste nel creare sistemi di risoluzione, non esclusivamente basati sui DNS. Il principale esponente di questo approccio alternativo è rappresentato dall'Handle System.

Questo sistema è stato inizialmente concepito e sviluppato presso la Corporation for National Research Initiatives (CNRI) come parte di una più vasta architettura globale per gli oggetti digitali.



## Identificatori basati su un name system alternativo al DNS (l'Handle System)

L'Handle System è un servizio che utilizza un suo **specifico protocollo** per creare, aggiornare, mantenere e rendere accessibile in modo sicuro un **database distribuito per l'associazione permanente tra nomi univoci e oggetti digitali**. È stato disegnato per rendere possibili servizi di trasmissione sicura di informazione e condivisione di risorse su network quali appunto Internet.

Le **applicazioni** dell'Handle System possono includere **servizi di metadati** per le pubblicazioni digitali, **servizi di gestione di identità** per le identità virtuali e/o ogni altra applicazione che richieda la risoluzione e/o l'amministrazione di identificatori univoci globali.



## Identificatori basati su un name system alternativo al DNS (l'Handle System)

Naturalmente l'Handle System, come tutte le realizzazioni concrete, è l'espressione di un compromesso tra varie e diverse esigenze tecniche ed organizzative.

Come configurazione globale del servizio è stato adottato un **modello gerarchico**. Il top-level del sistema è rappresentato da un singolo servizio handle, meglio conosciuto come **Global Handle Registry** (GHR). I livelli inferiori sono rappresentati da tutti gli altri servizi handle, meglio noti come **Local Handle Services** (LHS).

Il GHR è il gestore unico di tutte le naming authorities appartenenti al sistema. I singoli LHS operano al servizio di una data naming authority.

# Sintassi dell'Handle

Ogni handle consiste di **due parti**: la sua naming authority, altrimenti nota come prefisso, e un nome locale univoco sotto la naming authority, meglio noto come suffisso. L'insieme è preceduto da un'etichetta che, nel caso di un handle tradizionale, è «hdl:»

**<hdl> :**      **<naming Authority>**      **“/” <local name>**  
(etichetta)              (prefisso)                              (suffisso)

La naming authority identifica l'unità amministrativa responsabile per la creazione e l'amministrazione dell'handle associato. **Le naming authority sotto lo Handle System hanno struttura gerarchica ma, a differenza del DNS, le handle naming authorities sono costruite da sinistra verso destra.**





## Funzionamento dell'Handle

Operativamente il sistema funziona così: **un utente che debba risolvere un indirizzo handle, interroga con il suo client il GHR; questi invia in risposta l'indirizzo dell'LHS responsabile per la risoluzione di quel dato handle; il client interroga allora LHS in questione; questi a sua volta invia in risposta l'indirizzo della risorsa digitale associata all'handle.** Un aspetto importante dell'Handle System è la sua architettura distribuita. L'Handle System, nel suo insieme, consiste di vari **servizi** handle individuali. Ognuno di questi servizi può consistere di uno o più **siti di servizio**. Ogni sito di servizio, a sua volta, può consistere di uno o più **server** handle. Tutte le richieste handle, indirizzate ad un dato sito di servizio, possono essere ridistribuite, eventualmente, tra i vari server handle. Non ci sono limiti al numero di servizi handle o al numero di siti che possono essere allestiti per ogni servizio o al numero di server per ogni sito.

La comunicazione tra le varie componenti del sistema avviene mediante uno specifico protocollo (lo **Handle System Protocol**), più efficiente sotto molti aspetti rispetto al tradizionale protocollo DNS. In primo luogo le applicazioni DNS non reggono bene le situazioni in cui grandi quantità di dati sono associate con un particolare nome DNS. Inoltre, i nomi DNS sono gestiti dall'amministratore di zona di uno specifico livello della gerarchia DNS. Non sono previste modalità per implementare una struttura amministrativa finalizzata specificatamente alla gestione dei nomi. Solo l'amministratore del network può creare o maneggiare i nomi DNS. **Ogni handle, invece, ha il proprio amministratore distinto dall'amministratore del server.** Ancora, il protocollo dello Handle System include delle **opzioni di sicurezza che garantiscono la riservatezza e l'integrità dei dati durante la trasmissione.** Nel DNS questo non c'è.



## Identificatori basati su un name system alternativo al DNS (l'Handle System)

Lo Handle System fornisce un name service globale a scopi generali che permette di risolvere e amministrare in maniera sicura gli handle sulla rete.

Come è noto, dal punto di vista di un utente finale, per risolvere un URL basta scriverlo dentro lo spazio degli indirizzi del browser.

Anche gli handle si risolvono nello stesso modo. Ma prima occorre adattare il proprio browser Microsoft Internet Explorer o Netscape, scaricando dal sito del CNRI un **apposito programma**. Questo si chiama Handle Resolver e va installato, con una procedura estremamente semplice, sul proprio PC. Un browser così adattato può risolvere gli handle direttamente.

**<hdl> : <naming Authority> "/" <local name>**

Non disponendo di un browser adattato, si possono comunque risolvere gli handle utilizzando il normale protocollo HTTP e un **server proxy**. Il CNRI, ad esempio, rende disponibile un server proxy all'indirizzo <http://hdl.handle.net/>. Risolvere l'handle utilizzando un server proxy consiste nel costruire una normale URL costituita dall'indirizzo del server proxy seguito dall'handle da risolvere.

**[<local name>](http://hdl.handle.net/<naming Authority>)**



## Identificatori basati su un name system alternativo al DNS (I'Handle System)

La differenza tra le diverse modalità d'accesso risiede nella persistenza: se dovesse cambiare la localizzazione del documento, l'handle, a condizione che le informazioni nell'LHS vengano aggiornate, continuerà a operare. L'URL no. L'unica limitazione per l'handle sta nell'indirizzo del server proxy che, essendo un normale indirizzo DNS, può variare anch'esso. In tal caso anche gli indirizzi non saranno più validi.



## Digital Object Identifier

Attualmente l'implementazione più importante dell'Handle System è rappresentata dal Digital Object Identifier (DOI). Il DOI System è gestito dall'**International Doi Foundation**, un consorzio “aperto” che comprende partner sia commerciali che non-commerciali (Joint Information Systems Committee (JISC)(UK), CENL (Conference of European National Librarians)).

L'International DOI Foundation (IDF), pur non essendo una fondazione a scopo di lucro è di fatto sotto il controllo dei principali editori commerciali internazionali del settore STM (il suo CdA è composto principalmente da questi ultimi – Elsevier, John Wiley, Springer, ecc.).

All'IDF fanno capo svariate **agenzie regionali** di registrazione DOI che sinora hanno complessivamente assegnato alcune decine di milioni di DOI negli Stati Uniti (**Crossref** – Agenzia controllata da editori accademici e commerciali), in Europa (**mEDRA** - Agenzia controllata dall'Associazione Italiana Editori e dal CINECA) e in estremo oriente.



# Digital Object Identifier

Il DOI è a tutti gli effetti un handle e l'IDF una *handle naming authority*.

Un DOI si riconosce per la **prima parte del prefisso che è sempre 10** (che identifica l'identificatore DOI); ad esempio: **doi:10.1045/xxxx** è un DOI. La seconda parte del prefisso (1045) è l'identificativo dello specifico registrante responsabile di quel particolare DOI; il suffisso è il nome della risorsa identificata in maniera permanente da quel DOI (xxxx).

**Il nome può essere rappresentato da qualsiasi sequenza di caratteri.** Si possono usare per i nomi anche sistemi di codifica già esistenti. Ad esempio:

**doi:10.1000/ISBN1-900512-44-0**

è un DOI valido che incorpora un codice ISBN.

Anche un DOI può essere risolto direttamente utilizzando lo stesso Handle Resolver indicato prima, oppure via proxy-server, utilizzando il sistema allestito dall'IDF. Ad esempio.

**<http://dx.doi.org/10.1045/xxxx>**



## Digital Object Identifier

Una caratteristica importante del DOI è quella di essere un **sistema a pagamento** (gestito in massima parte da editori commerciali).

L'assegnazione del DOI richiede un pagamento e l'impegno a seguire standard e regole definite.

La DOI Foundation è un'organizzazione non-profit, non un operatore commerciale, tuttavia, **il sistema ha dei costi che vanno coperti.**

**La persistenza è una funzione delle organizzazioni, non una tecnologia: per supportare un sistema di identificazione persistente, occorre che esista una organizzazione persistente (“Filosofia” DOI).**



## Digital Object Identifier

**Per registrare un DOI occorre rivolgersi ad una apposita agenzia.** La più importante in assoluto è certamente CrossRef. Questa agenzia consente di assegnare i DOI ai seguenti materiali:

- periodici/articoli;
- libri: collana, titolo, capitolo, voce;
- atti di convegni: titoli multi - volume, titoli, elaborati;
- componenti: parti di articoli di periodici (tabelle, grafici, illustrazioni, ecc.); voci/capitoli di libri e di atti di convegni comprese le figure, le tavole e gli aggiornamenti, ecc.

Una delle caratteristiche principali del DOI è la sua **granularità identificativa** che può variare a seconda delle diverse necessità degli utilizzatori.

In pratica il DOI può identificare qualsiasi “frazione significativa di proprietà intellettuale” (dalla rivista, all’articolo, alla singola immagine, tabella, ecc. contenuta all’interno di una risorsa informativa).



## Digital Object Identifier

Naturalmente, quando un DOI viene assegnato ad un dato oggetto, nel database della agenzia di assegnazione deve essere inserita una descrizione di quell'oggetto.

Il DOI Handbook, a questo riguardo, prescrive un set minimo di metadati denominato **DOI Kernel Metadata Declaration**. Ogni risorsa identificata con il DOI deve essere descritta secondo le specifiche di questo insieme minimo di metadati.

Oltre al kernel, **le agenzie di registrazione possono implementare ulteriori schemi di metadati**, soprattutto a fini di interoperabilità tra piattaforme e servizi differenti.

**mEDRA**, a questo riguardo, si avvale di un set di metadati, denominato **ONIX** (Online Information Exchange), che è uno standard internazionale per rappresentare le **informazioni commerciali** su libri, riviste e video in formato elettronico. Questo set è mantenuto da **EDItEUR**, un gruppo internazionale che coordina lo sviluppo e la promozione degli standard per il commercio elettronico nell'ambito dell'editoria libraria e dei periodici.





## Lo standard NBN

- In questo contesto l' *Internet Engineering Task Force (IETF)* ha pubblicato lo standard *NBN (National Bibliography Number)*, che fornisce indicazioni sul formato degli identificatori e sulle relative modalità di gestione.
- NBN è uno *standard internazionale aperto* ed ogni paese ha un proprio dominio assegnato in modo esclusivo (es. IT per l'Italia, FR per la Francia).
- Lo standard NBN (RFC 3188, basato su URN, RFC 2141) prevede che la Biblioteca Nazionale gestisca il dominio affidato al proprio paese da accordi internazionali che ne riconoscono la responsabilità. Il registro internazionale dei domini, compreso quello NBN:IT per l'Italia, è mantenuto dalla Library of Congress. Ogni Stato è responsabile dell'implementazione del proprio registro e del servizio di risoluzione nazionale.



- **Diffusione:** l'adozione di questo standard è promossa dalla Conference of Directors of National Libraries (CDNL) e dalla Conference of European National Libraries (CENL).
- **Soluzioni adottate in altri paesi:** gestione centralizzata a livello nazionale di attribuzione del codice NBN (Germania, Finlandia, etc.).
- **Criticità:**
  - **scarsa diffusione** a causa dell'assegnazione centralizzata del codice;
  - **assenza di strumenti avanzati** per l'assegnazione del codice e il recupero della risorse.



## ...domini multilivello



Possibilità di adottare un approccio gerarchico distribuito per la registrazione/risoluzione (DNS-like)



# Bibliografia

Altma, Micah, King, Gary, *A Proposed Standard for the Scholarly Citation of Quantitative Data*, "D-Lib Magazine", 13, n. 3-4 (2007).

Bayer-Schur, Barbara, Brétel, Foudil, Bulatovic, Natasa, et al., *PEER: D2.2 Final report on the provision of usage data and manuscript deposit procedures for publishers and repository managers*, eContentplus, 2009

Berners-Lee, Tim, Fielding, Roy T., Masinter, Larry, *RFC 2396. Uniform Resource Identifiers (URI): Generic Syntax*, Internet Society, 1998, <<http://www.ietf.org/rfc/rfc2396.txt> >, data dell'ultima consultazione 9 novembre 2009.

Bellini, Emanuele , Cirinnà, Chiara, Lancia, Maurizio, Lunghi, Maurizio, Puccinelli, Roberto, Saccone, Massimiliano, Sebastiani, Brunella, Spasiano, Marco, *Persistent Identifier Distributed System for Digital Libraries*, in *World Library and Information Congress: 75. IFLA General Conference and Assembly, Milan, 23 – 27 August 2009*, Milano, IFLA, 2009.

Bellini, Emanuele, Cirinnà, Chiara, Lunghi, Maurizio, *Gli identificatori persistenti per i beni culturali*,

<[http://www.rinascimentodigitale.it/documenti/dpe/it\\_persistent\\_identifiers\\_for\\_cultural.pdf](http://www.rinascimentodigitale.it/documenti/dpe/it_persistent_identifiers_for_cultural.pdf)>

Bellini, Emanuele, Cirinnà, Chiara, Lunghi, Maurizio, Damiani, Ernesto, Fugazza, Cristiano, *Persistent Identifiers distributed system for cultural heritage digital objects*, in *IPRES2008 conference, 5. International Conference on Preservation of Digital Objects, The British Library, London. 29 – 30 September*, p. 242-249, <<http://www.bl.uk/ipres2008/ipres2008-proceedings.pdf>>, data dell'ultima consultazione 9 novembre 2009.

Dack, Diana, *Persistence is a Virtue*, "National Library of Australia Staff Papers", 2009, <<https://www.nla.gov.au/openpublish/index.php/nlas/article/viewArticle/1320/1606>>, data dell'ultima consultazione 9 novembre 2009.



# Bibliografia

Bellini, Emanuele, Damiani, Ernesto, Fugazza, Cristiano, Lunghi, Maurizio, *Semantics-aware Resolution of Multi-part Persistent Identifiers*, in *Emerging Technologies and Information Systems for the Knowledge Society. 1. World Summit on the Knowledge Society, WSKS 2008, Athens, Greece, September 24-26, 2008. Proceedings*, Berlin, Springer, 2008, p. 413-422.

Dobratz, Susanne, Scholze, Frank, *DINI institutional repository certification and beyond*, "Library Hi Tech", 24, n. 4 (2006), p. 583-594.

Gomes, Daniel, Silva, Mario J., *Modelling information persistence on the web*, in *Proceedings of the 6. international conference on Web engineering, Palo Alto, California, July 10 - 14, 2006, "International Conference On Web Engineering"*, 155 (2006).

Hakala, Juha, *RFC 3188. Using National Bibliography Numbers as Uniform Resource Names*, Internet Society, 2001, <<http://tools.ietf.org/html/rfc3188>>, data dell'ultima consultazione 9 novembre 2009.

Kunze, John, *ARK (Archival Resource Key)*, 2007, <<http://www.cdlib.org/inside/diglib/ark/>>, data dell'ultima consultazione 9 novembre 2009.

Library of Congress, *Relationship Between URNs, Handles, and PURLs*, Library of Congress - National Digital Library Program, 1997, <<http://lcweb2.loc.gov/ammem/award/docs/PURL-handle.html>>, data dell'ultima consultazione 9 novembre 2009.

Morgan, Hilary, *Persistent Identification of Digital Resources: Environmental Scan*, Library and Archives Canada , 2008, <[http://www.carl-abrc.ca/projects/nmrdi/Alouette-PersistentID\\_Scan-e.pdf](http://www.carl-abrc.ca/projects/nmrdi/Alouette-PersistentID_Scan-e.pdf) >, data dell'ultima consultazione 9 novembre 2009.

Paskin, Norman, *Digital Object Identifiers for scientific data*, "Data Science Journal", 4 (2005), p. 12-20.

*RFC 3305. Report from the Joint W3C/IETF URI Planning Interest Group: Uniform resource Identifiers (URIs), URLs., and Uniform Resource Names (URNs): Clarifications and Recommendations.*, Michael Mealling , Ray Denenberg Editors, Internet Society, 2002, <<http://www.ietf.org/rfc/rfc3305.txt>> data dell'ultima consultazione 9 novembre 2009

# Bibliografia



Schroeder, Kathrin, *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, n. 3 (2006) p. 75 – 87.

*Semantic Interoperability in the European Digital Library: Proceedings of the First International Workshop, SIEDL 2008, Tenerife, June 2, 2008*, Stefanos Kollias, Jill Cousins (eds), Europeana, 2008, <<http://image.ntua.gr/swamm2006/SIEDLproceedings.pdf>>, data dell'ultima consultazione 9 novembre 2009.

Sun, Sam, Lannom, Larry, Boesch, Brian, *RFC 3650. Handle System Overview*, Internet Society, 2003, <<http://www.ietf.org/rfc/rfc3650.txt>>, data dell'ultima consultazione 9 novembre 2009.

Sun, Sam, Reilly, Sean, Lannom, Larry, *RFC 3651. Handle System Namespace and Service Definition*, Internet Society, 2003, <<http://www.ietf.org/rfc/rfc3651.txt>>, data dell'ultima consultazione 9 novembre 2009.

Sun, Sam, Reilly, Sean, Lannom, Larry, Petrone, Jason, *RFC 3652. Handle System Protocol (ver 2.1) Specification*, Internet Society, 2003, <<http://www.ietf.org/rfc/rfc3652.txt>.<http://www.cnri.reston.va.us/k-w.html>>, data dell'ultima consultazione 9 novembre 2009.

Vitiello, Giuseppe, *L'identificazione degli identificatori*, "Biblioteche Oggi", n. 2 (2004), p. 67-80.

Mosley, Ray, *Federal Register 2.0: Opening a Window onto the Inner Workings of Government*, 2009, <<http://www.whitehouse.gov/blog/Federal-Register-20-Opening-a-Window-onto-the-Inner-Workings-of-Government>>, data dell'ultima consultazione 9 novembre 2009.

# Lista dei siti consultati

**CASPAR - Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval**, <<http://www.casparpreserves.eu/>>, data dell'ultima consultazione 9 novembre 2009.

**Corporation for National Research Initiatives**, <<http://www.cnri.reston.va.us/>>, data dell'ultima consultazione 9 novembre 2009.

**CrossRef.org**, <<http://www.crossref.org/>>, data dell'ultima consultazione 9 novembre 2009.

**D4Science**, <<http://www.d4science.eu/>>, data dell'ultima consultazione 9 novembre 2009.

**Digital Object Identifier (DOI®) System**, <<http://www.doi.org/>>, data dell'ultima consultazione 9 novembre 2009.

**Europeana**, <<http://www.europeana.eu/portal/>>, data dell'ultima consultazione 9 novembre 2009.

**Handle Net - The Handle System**, <<http://www.handle.net/index.html>>, data dell'ultima consultazione 9 novembre 2009.

**Knowledge Exchange**, <<http://www.knowledge-exchange.info/>>, data dell'ultima consultazione 9 novembre 2009.

**mEDRA**, <<http://www.medra.org/>>, data dell'ultima consultazione 9 novembre 2009.

**Minerva Europe**, <<http://www.minervaeurope.org/>>, data dell'ultima consultazione 9 novembre 2009.

**PARSE.Insight**, <<http://www.parse-insight.eu/>>, data dell'ultima consultazione 9 novembre 2009.

**PEER (Publishing and the Ecology of European Research)**, <<http://www.peerproject.eu/>>, data dell'ultima consultazione 9 novembre 2009.

**PURL Homepage**, <<http://purl.oclc.org/docs/index.html>>, data dell'ultima consultazione 9