

Interoperabilità delle biblioteche digitali

Università Roma Tre – 9/10 settembre 2006

Identificativi per le risorse digitali

Zeno Tajoli – tajoli@cilea.it



Sommario

- ❑ Perché gli identificativi
- ❑ La base: l'URI
- ❑ Alcuni identificativi (usati e non)
- ❑ Come li trasportiamo ?
- ❑ Il software OS disponibile
- ❑ C'è una sfida in corso ?

Perchè gli identificativi

- E' proprio quello ?
- E' sempre quello ?
- Posso integrare questo con quello ?
- E' mio, è tuo, è suo ?
- Ci posso accedere ?

Perchè gli identificativi

- Bisogna identificare con sicurezza
- Al livello che serve
 - Articolo
 - xxx.doc
 - Xxx.pdf
 - Immagine di un saggio
 - Sinfonia
 - Movimento
 - Etc.

Perchè gli identificativi

- La macchina vuole sintesi, possibilmente numerica.
- Per il commesso di libreria:
 - I promessi sposi e Storia della colonna infame, di A. Manzoni. Zanichelli 1987
- Per un web service
 - ISBN 88-08-04380-0

Perchè gli identificativi

- Una gestione economica il più svicolata possibile
 - Scelta della granularità
 - Scelta delle opzioni di accesso
 - Differenziazione dei titolari e delle percentuali
 - Accesso a diversi regimi di protezione

La base: l'URI

- ❑ Uniform Resource Identifier
- ❑ Lo schema generale
- ❑ Definito dall' [RFC 3986](#)
- ❑ Aggiornato nel gennaio 2005
- ❑ Uno spazio in cui altri si inseriscono
- ❑ `scheme ":" hier-part ["?" query] ["#" fragment.]`

La base: l'URI

- ❑ info:lccn/n78890351
- ❑ <http://purl.org/dc/terms/audience>
- ❑ <http://purl.org/poi/rdn.ac.uk/12-34>
- ❑ sbn://bid/MIL0378495
- ❑ doi:10.1000/182
- ❑ <http://hdl.handle.net/1814/2131>

Alcuni identificativi (usati e non)

□ Handle

- RFCs: [3650](#), [3651](#) e [3652](#).
- Forma: <Handle Naming Authority> "/" <Handle Local Name>
(<http://hdl.handle.net/1889/475>)
- Servizio strutturato a due livelli, mondiale e locale
- Identificatore "opaco"
- Si parla di Handle System più che del singolo handle
- Gestito dalla CNRI, base di partenza <http://www.handle.net/>
- Da poco richiede un pagamento a chi li crea

Alcuni identificativi (usati e non)

□ **DOI**

- Il DOI è un specifico Handle
- ANSI/NISO Z39.84-2005 è solo la sintassi formale
- Sintassi: <DIR>.<REG>/<DSS> (10.1000/123456)
- Semantica proprietaria dei metadati dal indecs framework
- Sito di riferimento: <http://www.doi.org/>
- Forte interesse sulla gestione dell' IPR
- Sistema funzionante attraverso una rete di agenzie
- Coordinamento attraverso la "International DOI Foundation"
- Agenzia più nota: CrossRef
- Mix di servizi gratuiti e a pagamento

Alcuni identificativi (usati e non)



OAI Identifier

- Strettamente collegato all'Open Access Movement
- Sintassi: oai:<nome di dominio>:<id locale>
(oai:eprints.rclis.org:789)
- Rif: [Specification and XML Schema for the OAI Identifier Format](#)
- Non obbligatorio per aderire all'OAI-PMH, solo consigliato
- Risoluzione locale
- Nessuna risoluzione a livello globale
- Metadati (Dublin Core) finalizzati all'identificazione
- Creati come sottoprodotto degli Open Archives
- Uso completamente gratuito

Alcuni identificativi (usati e non)



Purl

- URL con alcune caratteristiche speciali
- Sviluppato da OCLC durante la discussione sugli URN
- Schema: <protocol>://<resolver>/<name>
(<http://purl.oclc.org/keith/home>)
- Sito di riferimento: <http://purl.oclc.org/>
- Funziona sia localmente che globalmente
- I resolver sono accessibili con nomi di dominio
- Attualmente strettamente legato a OCLC
- Alla base c'è la funzionalità di *redirect* dell' http
- Il focus è messo sulla persistenza, su evitare 404 error.

Alcuni identificativi (usati e non)

□ Tinyurl

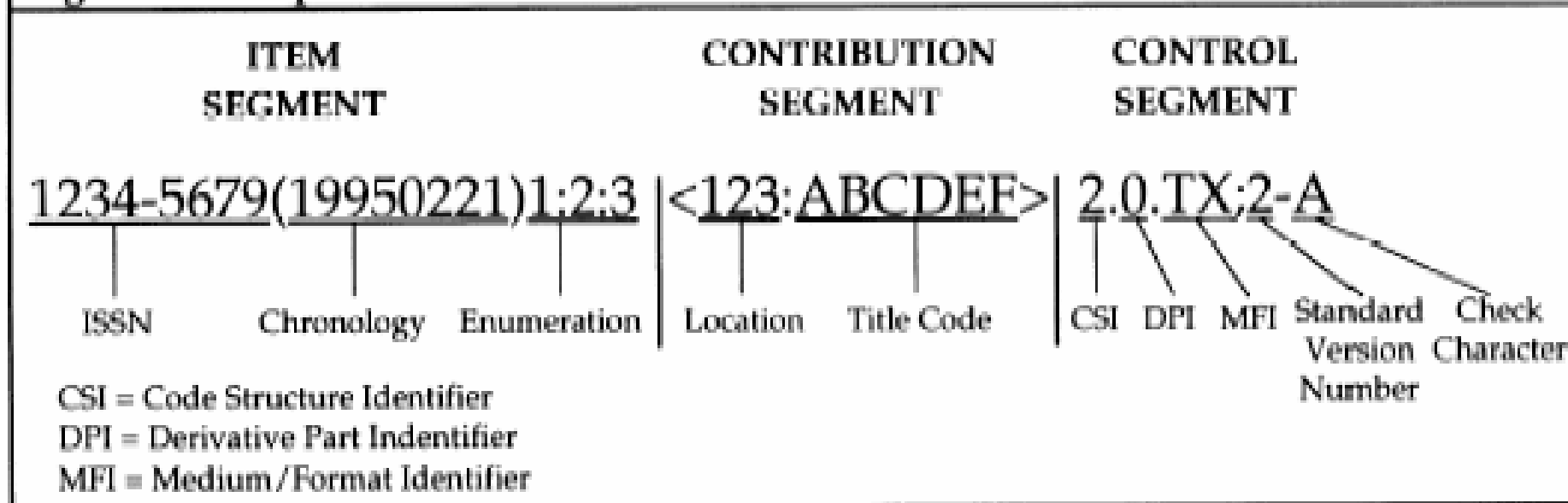
- Fatto da una azienda, Gilby Productions
- Sintassi: `http://tinyurl.com/<stringa>`
- Riferimento: <http://tinyurl.com/>
- Si basa unicamente sul redirect
- Serve per abbreviare gli URL
- In se stesso è un URL
- Non serve per la persistenza
- Utile per i client di posta elettronica
- Diffuso via informale

Alcuni identificativi (usati e non)

□ Sici

- Riferimento: ANSI/NISO Z39.56-1996

Figure 2: Example of a CSI-2



Alcuni identificativi (usati e non)

□ Sici

- I suoi componenti hanno un loro significato
- Uso specifico fascicoli e articoli di periodici
- Nato nell'ambito della documentazione scientifica
- Calcolabile autonomamente
- Molto legato all' ISSN
- Manca completamente di un risolutore
- Può essere riutilizzato da un software tipo link resolver

Come li trasportiamo ?

- Protocollo info:
- http generico
- OpenUrl (http standardizzato)

Come li trasportiamo ?

□ Info:

- Schema: info ":" namespace "/" identifier ["#" fragment]
(info:lccn/n78089035)
- Definito come schema URI in [RFC 4452](#)
- Riferimento: <http://www.loc.gov/standards/uri/info.html>
- Attualmente non utilizzabile direttamente.
- Suo uso già compreso nell' OpenUrl ver. 1.0

Come li trasportiamo ?

□ http generico

- Soluzione proprietaria
- Da implementare di volta in volta
- Bisogna ripetere l'analisi della situazione ogni volta
- Veloce da fare
- Difficile interoperabilità e persistenza
- Esempio: Tinyurl

Come li trasportiamo ?

□ **OpenUrl 1.0**

- Protocollo su http con forme standardizzate
- ANSI/NISO Z39.88-2004
- Riferimento: <http://openurl.info/registry>
- Una cornice (framework) per il trasporto
- La semantica dei dati sta in una serie di accordi

Come li trasportiamo ?

- ❑ Standard omnicomprensivo (OpenUrl 0.1)
- ❑ Standard più astratto (OpenUrl 1.0)
- ❑ Due tipi di formattazione : con chiave/valore e con XML
- ❑ Tra i mezzi che **possono** essere usati ci sono gli identificativi espressi come URI
- ❑ Esempi:
 - &rft_id=info:doi/10.1126/science.275.5304.1320
 - &rft_id=info:pmid/9036860
- ❑ Un identificativo semplifica notevolmente il lavoro del resolver
- ❑ ... Ma non sempre c'è l'identificativo

Come li trasportiamo ?

- Principali semantiche definite finora:
 - info:ofi/fmt:kev:mtx:book Key/Encoded-Value Metadata Format for Books
 - info:ofi/fmt:kev:mtx:dc Use Key/Encoded-Value Metadata Format for Dublin Core (Trial)
 - info:ofi/fmt:kev:mtx:journal Key/Encoded-Value Metadata Format For Journals
 - info:ofi/fmt:xml:xsd:journal XML Metadata Format for Journals
- Si usano quando non ci sono gli identificativi

Il software OS disponibile

Handle

- CNRI offre il resolver locale scritto in Java
- Il resolver si aspetta i dati in un DB relazionale
- Dà per scontato di appoggiarsi al risolutore globale della CNRI
- Uso intranet solo con forti modifiche
- Dspace usa questo resolver per aver gli Handle

Il software OS disponibile

□ DOI

- Una serie di utilità per facilitarne l'uso
- Estensioni per i browser
- Integrazione in Google Toolbar
- CrossRef offre alcune interfacce a corredo con a corredo del software e il sorgente relativo
- Anche un servizio di risoluzione a uso sporadico

Il software OS disponibile

- Eprints/Fedora/OJS/OCS
 - Gestiscono oai-identifier
 - Il nome di dominio lo basano sull'indirizzo web del sistema
 - Non esiste un generatore come software indipendente
 - Esisteva un risolutore globale gestito da OCLC

Il software OS disponibile

□ Purl

- Sorgente del server locale in C e perl
- Servizio globale offerto da OCLC
- Sviluppo gestito totalmente da OCLC

□ TinyUrl

- Servizio di web service per [sviluppatori](#)

Il software OS disponibile

□ OpenUrl Resolvers:

- La suite [reSearcher](#) della Simon Fraser
- [OpenUrl module](#) di OCLC
- [OpenResolver](#) di UKLON
- I moduli offerti dalla [Openly](#)
- [Olinks](#) della Ohio University

□ In realtà il vero problema non è il software ma la knowledge base

C'è una sfida in corso ?

- Attualmente il DOI ha avuto un grande successo.
- La risposta definitiva nell'ambito documentazione ?
- E nel campo musica/video ?
- Meglio un servizio centralizzato (DOI/Handle) o uno decentralizzato (OAI-Identifier) ?
- Forse vale la pena di rifare un resolver globale per gli OAI-Identifier.

Links

- ❑ RFC 3986 <<http://www.ietf.org/rfc/rfc3986.txt>>
- ❑ RFC 3650 <<http://www.ietf.org/rfc/rfc3650.txt>>
- ❑ RFC 3651 <<http://www.ietf.org/rfc/rfc3651.txt>>
- ❑ RFC 3652 <<http://www.ietf.org/rfc/rfc3652.txt>>
- ❑ <http://www.doi.org/>
- ❑ OAI Identifier <<http://www.openarchives.org/OAI/2.0/guidelines-oai-identifier.htm>>
- ❑ <http://purl.oclc.org/>
- ❑ <http://tinyurl.com/>
- ❑ <http://www.loc.gov/standards/uri/info.html>
- ❑ RFC 4452 <<http://www.ietf.org/rfc/rfc4452.txt>>
- ❑ <http://openurl.info/registry>
- ❑ http://tinyurl.co.uk/xml_interface/
- ❑ reSearcher <<http://researcher.sfu.ca/index.php/plain/development>>
- ❑ OpenUrl module <<http://www.oclc.org/research/software/openurl/default.htm>>
- ❑ OpenResolver <<ftp://ftp.ukoln.ac.uk/metadata/tools/openresolver/>>
- ❑ Openly <<http://www.openly.com/>>
- ❑ Olinks <<http://olinks.sourceforge.net/>>