

La conservazione delle risorse digitali

Giovanni Bergamin

Giovanni.bergamin@bncf.firenze.sbn.it

2003-02-14

10. Edizione

MASTER II DIREZIONE D  TIONE E BIBLIOTECA

Metadati

- “la parola più usata e abusata nel vocabolario semi-tecnico di questo inizio secolo” [McClennen]
- “dati che descrivono o identificano altri dati” [McClennen]
- si tratta di una *funzione* svolta da determinati *dati*



Metadati in quanto ...

- ***funzione svolta***

- non si tratta di una ***esclusiva*** dell'ambiente digitale (es. le schede di un catalogo, l'inserito settimanale dei programmi televisivi, elenco di file)

- ***termine***

- il contesto di riferimento è oggi quello delle risorse digitali in rete



Descrittivi e Gestionali

- metadati finalizzati al *recupero* della risorsa
metadati *descrittivi* = *informazioni che permettono*
 - *l'identificazione*
 - *la descrizione*
 - *la caratterizzazione*
 - *l'accesso tramite ad es. la responsabilità intellettuale, l'argomento ecc*
 - ★ Napster
 - ★ BNCF



Descrittivi e Gestionali

- metadati necessari alla *gestione* della risorsa
metadati *gestionali* = permettono la
pubblicazione, l'acquisizione, la vendita, la
*fruizione, la **conservazione** ecc.*

★ Napster

★ BNCF



... e **digital** (preservation)

- "Informazioni di tipo diverso
 - ★ (film, musica, testi, immagini ecc)
- possono essere tutte ridotte allo stesso codice di base,
 - ★ (convergenza)
- alle **lunghe catene di 0 e di 1** dell'informazione digitalizzata"
[Ciotti-Roncaglia - 2000]



lunghe catene di 0 e di 1

- Non sono legate ad uno specifico supporto
- E' sempre possibile "clonarle"
 - = copiare nella stessa sequenza i bit su un altro supporto
 - = creare altri "originali"

((Può accadere che un determinato prodotto digitale diffuso su un determinato supporto sia "protetto da copia"))



... e digital (preservation)

- se il digitale è indipendente dal supporto
- *allora* per conservare la "catena di 0 e di 1" è sufficiente
- *replicare* quella "catena"
 - ★ con regolarità (prima che ..)
 - ★ scegliendo supporti di destinazione aggiornati
 - ★ su più supporti
 - ★ con "qualità"



La fragilità *del supporto del digitale* può essere gestita - 1

- **basta**

- “rinfrescare” il supporto
 - strategia ben conosciuta (da tempo)
 - (assieme al backup - “salvaguardia” - dei dati)
- farsi carico dei costi (alti)



La fragilità *del supporto del digitale* può essere gestita - 2

- **“Domanda”**: Tra 10 o 20 o 100 anni come leggeremo i supporti informatici di ieri e di oggi?
- **“Risposta”**: probabilmente non li leggeremo più ma la “catena di 0 e 1” sarà disponibile su un nuovo supporto (esclusi i “protetti da copia”)



E' più difficile gestire la fragilità del digitale

Il digitale è intrinsecamente fragile perché:

- dipende da altre *catene di 0 e di 1*
- (software)

- dipende da dispositivi in grado di elaborare *quelle catene* e renderle significative (usabili/produttive)
- (hardware)



Questa presentazione ...

- È una risorsa digitale indipendente dal supporto (non *cambia* se la trasmetto per posta, su dischetto ecc)
- Ma dipende dalla disponibilità *nel tempo* di:
 1. Programma (Powerpoint [*versioni*])
 2. Sistema operativo (Windows [*versioni*])
 3. Personal computer Intel/AMD [*versioni*]



Il digitale è fragile perché ...

- “Data la continua evoluzione
- sia dell’hardware che del software
- può accadere che quell’insieme di strumenti non sia più disponibile
- nel giro di pochi anni dalla sua introduzione sul mercato (CPA/RLG 1996)”



Scelte tecnologiche

- La conservazione del digitale è un problema tecnologico (= di scelte tecnologiche)
- creato dalla tecnologia (= da scelte tecnologiche)
- può essere risolto solo da scelte tecnologiche "globali".



Meglio la “prevenzione”

Progetto ISO di *PDF/A*

- Definire un insieme di funzionalità di un documento PDF per la conservazione nel lungo termine
- Tra i divieti:
 - Crittografia
 - Metodi di compressione proprietari
 - Riferimenti esterni (font e file)
 - Programmi sepolti e font proprietari



Ma intanto occorre:

- Prendere in conto la realtà
- Trovare un linguaggio comune
- Trovare soluzioni sostenibili di *archiviazione*



OAIS - Open Archival Information System

Archivio [in questo contesto] =

- «un'organizzazione di soggetti e sistemi
- che hanno accettato la responsabilità di conservare l'informazione
- e di mantenerla disponibile per la comunità di riferimento»





Metadati descrittivi - Napster



The image shows a Windows XP desktop environment. A file explorer window titled 'traccia' is open, displaying the contents of the folder 'D:\My Music\traccia'. The folder name 'traccia' is prominently displayed in the left pane. The right pane shows a list of MP3 files with their names and file icons. A Winamp music player is overlaid on the bottom right of the file explorer window, showing the current track as '6. BEATLES - REVOL' with a progress bar and various playback controls.

Nome
Adriano Celentano - Preghero.mp3
Apache.mp3
Battisti Lucio - La luce dell'est.mp3
Beach Boys - Barbara Ann.mp3
Beatles - Imagine.mp3
Beatles - Revolution.mp3
Camaleonti - Applausi.mp3
Caterina Caselli - Il Volto Della Vita.mp3
Conte Paolo - Via con me (canta Roberto Benigni).mp3
Dik Dik - Sognando la California.mp3
Dire Straits-Tunnel of Love.mp3

WINAMP
00:18
<3:24> *** 6. BEATLES - REVOL
128 kbps 44 kHz mono stereo
SHUFFLE

Metadati descrittivi - BNCF



OPAC

PORTO SBN

BNCF
Marucelliana

Record selezionati: 13 - 14

[◀ Lista](#)



Eco, Umberto

Numele trandafirului / Umberto Eco ; traducere si postfata de Florin Chiritescu . Chisinau : Hyperion, 1992.

478 p. : 1 ritr. ; 21 cm.

Tit. orig.: [Il *nome della rosa](#)

ISBN 536801029X.

I. [Chiritescu, Florin](#)

Dedalus v. Eco, Umberto

Monografia - Materiale a Stampa - Pubbl. in: RO - Lingua: rum



Gli utenti registrati possono accedere ai servizi della biblioteca scegliendo il numero di inventario del documento desiderato.

Bibl. Nazionale Centrale di Firenze

Collocazione: GEN B15 05746

[Inventario:](#) CF005555284 1 v.

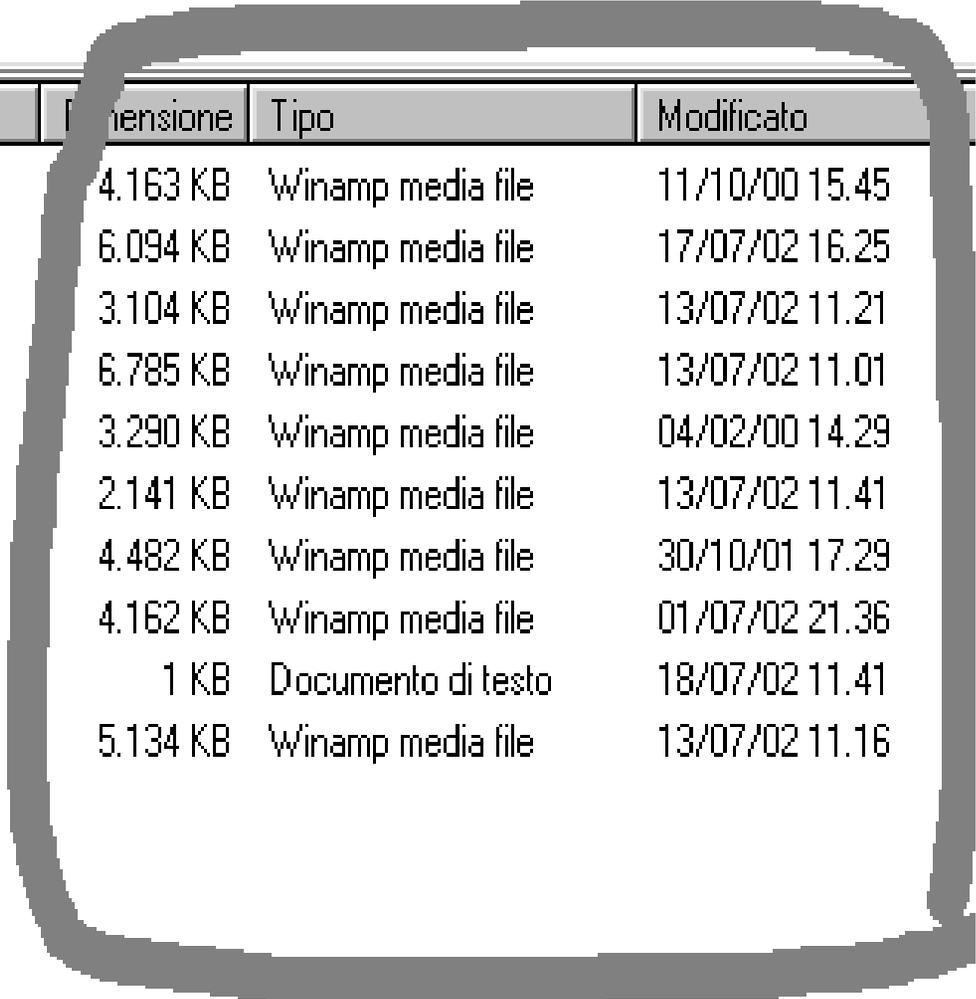


Metadati gestionali - Napster



Unità di rete	Disconnetti unità di rete	Taglia	Copia	Incolla	Annulla	Elimina	Proprietà	Visualizza
---------------	---------------------------	--------	-------	---------	---------	---------	-----------	------------

Nome	Dimensione	Tipo	Modificato
Alizee - Moi... Lolita.mp3	4.163 KB	Winamp media file	11/10/00 15.45
Dave%20Brubeck%20-%20Take%20Five%20(Live).mp3	6.094 KB	Winamp media file	17/07/02 16.25
Eric Clapton - Wonderful Tonight.mp3	3.104 KB	Winamp media file	13/07/02 11.21
Gabin - Doo Uap Doo Uap Doo Uap.mp3	6.785 KB	Winamp media file	13/07/02 11.01
Kaoma - Lambada.mp3	3.290 KB	Winamp media file	04/02/00 14.29
louis armstrong - what a wonderful world.mp3	2.141 KB	Winamp media file	13/07/02 11.41
Noir Desir - Le Vent Nous Portera.mp3	4.482 KB	Winamp media file	30/10/01 17.29
Red Hot Chili Peppers - By The Way.mp3	4.162 KB	Winamp media file	01/07/02 21.36
WONDERFUL%20TONIGHT.txt	1 KB	Documento di testo	18/07/02 11.41
Yu Yu - Mon Petit Garçon (radio edit) .mp3	5.134 KB	Winamp media file	13/07/02 11.16



Metadati descrittivi - BNCF



OPAC

POIO SBN

BNCF
Marucelliana

Record selezionati: 13 - 14

◀ Lista



Eco, Umberto

Numele trandafirului / Umberto Eco ; traducere si postfata de Florin Chiritescu . Chisinau : Hyperion, 1992.

478 p. : 1 ritr. ; 21 cm.

Tit. orig.: [Il *nome della rosa](#)

ISBN 536801029X.

I. [Chiritescu, Florin](#)

Dedalus v. Eco, Umberto

Monografia - Materiale a Stampa - Pubbl. in: RO - Lingua: rum



Gli utenti registrati possono accedere ai servizi della biblioteca scegliendo il numero di inventario del documento desiderato.

Bibl. Nazionale Centrale di Firenze

Collocazione: GEN B15 05746

[Inventario: CF005555284](#) 1 v.



Il modello OAIS ...

- Propone un modello di archiviazione *dei dati* che ha solide basi sui *metadati*
- vede la risorsa digitale **Content information** come inseparabilmente composta da:
- **Content data object** che consiste in una sequenza (stream) di bit o in un set di sequenze di bit [**dati**]
- e da ...



Risorsa = dati + metadati

e da ...

- **Representation information** metadati che traducono il Content data object in
- **conoscenza accessibile (per es. processabile da un computer)**
- **e dotata di significato (per es. comprensibile ad un essere umano)**



Ma per archiviare servono anche ...

- **Descriptive information:** i famosi *metadati descrittivi*
- **Packaging information:** danno informazioni relative al tipo di rapporto tra risorsa e supporti che la veicolano
- **Preservation Description information ...**



Preservation Description information ...

- Reference information (identificazione URN/URL)
- Context information
- Provenance information
- Fixity information (autenticità)



Fixity information (autenticità)

- L'impronta (p. es. md5)
 - è una nuova sequenza di bit
 - deriva dal doc. attraverso un algoritmo sw
 - ha lunghezza fissa (es. 128 bit in md5)
 - ogni documento la sua impronta (sensibile a cambiamenti anche minimi)
- PKI e firma elettronica
- Livelli di autenticità diversi



Res signata

- <<un cambiamento radicale nella concezione e nell'uso del *documento*, così come lo conosciamo da migliaia di anni, nella sua natura di *res signata*, cioè di una *cosa* che riporta dei segni, delle informazioni. [..]
- Il documento informatico è *immateriale*, dal momento che la sua esistenza non è legata a un determinato supporto fisico. [..]>>



Legislazione

- Art. 491-bis del codice penale, introdotto dalla legge 23 dicembre 1993, n. 547: "...per documento informatico si intende qualsunque supporto informatico contenente dati o informazioni aventi efficacia probatoria o programmi specificamente destinati ad elaborarli".
- Invece il DPR 10 novembre 1997, n. 513 afferma all'articolo 1 che il documento informatico è "la rappresentazione informatica di atti, fatti o dati giuridicamente rilevanti".



Digital preservation ... in pratica

- Ma è possibile/sostenibile registrare *tutti* i metadati appena esaminati per tutte le risorse che decidiamo di archiviare?

Principali ostacoli:

- Informazione proprietaria (per Linux è possibile, ma non ad es. per Windows XP)
- Può avere costi molto alti



Strategie

- Il passaggio a nuovi supporti è una pre-strategia (necessaria ma non sufficiente: comunque consolidata)
- **Migrazione ed emulazione** sono invece strategie che “funzionano”



La migrazione

- È da tempo praticata con successo soprattutto con i database
 - (i dati sono indipendenti dalle applicazioni
 - e quindi possono essere scaricati dal vecchio sistema e caricati su un nuovo sistema
 - le applicazioni sul nuovo ambiente vanno in pratica riscritte)



La migrazione - 2

- Richiede un livello dettagliato di metadati
- Può comportare perdita di informazioni



L'emulazione - 1

E' basata su due assunti:

- L'hardware (es il computer Intel) è perfettamente documentato
- In futuro esisterà qualcosa in grado di elaborare il digitale (computer)



L'emulazione - 2

- **Basta quindi archiviare**

- la documentazione dell'hardware
- il sistema operativo come "catena di 0 e 1"
- Il programma (Powerpoint) come "catena di 0 e 1"
- la risorsa stessa come "catena di 0 e 1"
- gli altri metadati (descrittivi, provenienza, autenticità ecc)



L'emulazione - 3

Per poi poter in futuro (tra 100 anni)

- Costruire un software che emuli l'hardware del 2002 (p. es Pentium Intel)
- Caricare il sistema operativo conservato come "catena di 0 e 1" (Windows XP)
- Caricare il programma come "catena di 0 e 1" (Powerpoint)
- Finalmente proiettare questa presentazione



OAIS

- **Cominciare *ora* con OAIS si può**
- **Molte (la maggioranza) delle risorse digitali in rete sono basate su formati standard e/o pubblici**
- **Il modello OAIS può aiutare**
 - nella comprensione del problema
 - a decidere quanti e quali metadati archiviare



Digitalizzazione e catalogazione

- la digitalizzazione come valorizzazione dell'oggetto analogico
- la digitalizzazione come riedizione dell'oggetto analogico



Digitalizzazione e catalogazione - 2

- "In the context of applying the ISBD(ER), all remote access electronic resources are considered to be published"
- Num. BNF de l'éd. de Paris : B. Grasset, 1914



856

MARC 21

856 41 \$u<http://www.jstor.org/journals/0277903x.html>

"In this case, the item represented by the bibliographic record is not electronic but an electronic version is available"

UNIMARC

856 4# \$u<http://lcweb.loc.gov/catdir/seminar.html>

---> non è prevista la versione digitale di una risorsa analogica



956

956 #0 \$a BNCf0001062723 (versione “parziale”)

956 #1 \$a BNCf0002373246 (versione “full”)

956 #2 \$a BNCf0000757777 (copia depositata)



Conservazione e digitalizzazione 1

Principalmente strumento di accesso

Alternativa al microfilm? (l'approccio ibrido)

In alcuni casi l'unica alternativa per la conservazione (es. disco a 78 giri)



Conservazione e digitalizzazione - 2

- Archival standard (l'originale non sarà più disponibile)
- Transmissive standard (qualità sufficiente da garantire l'uso che ne verrà fatto dalla maggior parte degli studiosi: l'originale dovrebbe beneficiare dal minore stress subito)



A livello ILS - 1

<< Should we implement separate systems for

Web archiving,

[digitization projects]

CD-ROM archiving,

and archiving of online journals?



A livello ILS - 2

- We realized that if we did, the process could go on forever; we would have hundreds of systems sitting next to each other.
- That would not be manageable. Therefore, we decided that we wanted to be able to treat all electronic publications in the same manner, regardless of type >>
- Un modello



Bibliografia -1

Ciotti, Fabio – Roncaglia, Gino. Il mondo digitale.
Roma: Laterza, 2000.

Cirotto, Gloria. Il modello OAIS,
<http://www.iccu.sbn.it/semimeta.htm>

*Gruppo di studio sugli standard e le applicazioni di
metadati nei beni culturali,*
<http://www.iccu.sbn.it/metadati.htm>

MAG Schema,
<http://www.bncf.firenze.sbn.it/progetti/mag/index.html>



Bibliografia - 2

Malinconico, Michael. S. *Biblioteche digitali : prospettive e sviluppo.* <<Bollettino AIB>>. 38(1998), 3, p.275-299 (reperibile anche in <http://www.aib.it/aib/boll/1998/98-3-275.htm>)

Metadata Encoding and Transmission Standard (METS) / maintained in the Network Development and MARC Standards Office of the Library of Congress,
<http://www.loc.gov/standards/mets/>



Bibliografia - 3

*Open Archival Information System (OAIS)
Resources / RLG,*

<http://www.rlg.org/longterm/oais.html>.

*Reference Model for an Open Archival
Information System (OAIS). Blue Book. Issue 1.
January 2002 (this recommendation has been
adopted as ISO 14721:2002),*

http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html



Bibliografia - 4

Robinson, Peter. *The digitization of primary textual sources*. - Oxford : OHC, 1993

Werf, Titia van der. Experience of the National Library of the Netherlands (2002)

<http://www.clir.org/pubs/reports/pub107/vanderwerf.html>

