

# Gli standard e i modelli di metadati: *Dublin Core, METS, MODS, MAG*

---

***Susanna Mornati, CILEA***

**Torino, 5 ottobre 2004**

ICCU - Progetto Centri E-Learning – Corso di Specializzazione in Esperti  
per la Documentazione Avanzata e la Promozione del Patrimonio Culturale

# Contenuti

---

- Definizione di metadati
- Impiego dei metadati
- Strutture, schemi, set di elementi
- Strumenti per la creazione ed il controllo
- Interoperabilità e scambio
- Futuro dei metadati

# Prima parte

---

## Definizione di metadati

# Cosa sono i metadati?

---

???

# Risposte della classe

---

1

2

3

# Dati o informazioni?

---

viola

AB713XY

112233

Username: **viola**

Password: **AB713XY**

Data di nascita: **112233**

# Dati o informazioni?

---

- ❑ I metadati contestualizzano i dati trasformandoli in informazioni
- ❑ I metadati sono informazioni sui dati
- ❑ Sono spesso definiti “dati sui dati” o “informazioni sulle informazioni” (dal greco “metà”: oltre, al di sopra, al di là)

# Dati o informazioni?

---

- ❑ I metadati sono una varietà di codici di riferimento
- ❑ La loro condivisione li rende comprensibili (importanza degli standard)
- ❑ Sono le nostre "etichette" per descrivere la realtà, spesso sottintese



# Metadati sottintesi

---

□ Sono quasi le 10.00

□ (ora del giorno)

□ Oggi è il 5

□ (giorno del mese)

□ A Torino ci sono 15 gradi

□ (Celsius di temperatura esterna)

# Esercitazione

---

- Individuare se e quali metadati sono presenti sul materiale distribuito

# Definizione di metadati

---

- Sono **informazioni strutturate** che descrivono, spiegano e collocano una **risorsa informativa**

# Quali metadati?

---

- Comunità diverse impiegano diversi vocabolari di metadati (ontologie):
  - Dublin Core
  - MARC
  - TEI
  - METS
  - MODS
  - ONIX
  - etc.

# Esempio: standard di metadati nelle biblioteche - 1

---

- Schemi formali per la catalogazione tradizionale, ad es.:
  - UNIMARC: <http://www.ifla.org/VI/3/p1996-1/sec-uni.htm>
  - MARC21: <http://www.loc.gov/marc/>
- Schemi formali per la classificazione semantica, ad es.:
  - CDD: <http://www.oclc.org/dewey/>
  - LCC: <http://www.loc.gov/catdir/cpsol/lcco/lcco.html>
  - PACS: <http://publish.aps.org/PACS/>
  - MSC: <http://www.ams.org/msc/>

# Esempio: standard di metadati nelle biblioteche - 2

---

- Set di regole per la catalogazione:
  - ISBD: <http://www.ifla.org/VI/3/nd1/isbdlist.htm>
  - RICA: <http://www.iccu.sbn.it/manucata.htm>
  - AACR2: <http://www.aacr2.org/>
- Schemi descrittivi per risorse non tradizionali:
  - Risorse elettroniche
  - Finding aids archivistici
  - Materiali audio-visivi e multimediali
  - Materiali didattici
  - Set di dati scientifici
  - ...

# Tipologie di metadati

---

- Metadati descrittivi
- Metadati strutturali
- Metadati amministrativi:
  - Metadati per il DRM
  - Metadati per la conservazione

# Metadati descrittivi

---

□ Descrivono una risorsa per scopi di presentazione (discovery), identificazione, recupero (retrieval), ad esempio:

- Autore
- Titolo
- Keywords
- Soggetto
- abstract



# Esempio di metadati descrittivi

(dal catalogo ALEPH: <http://opac.biblio.unimib.it/>)

---

LDR02602nam0-22003851--450-; 001000075967;  
00520021204150754.0; 010|a 88-781-2061-8; 035|a  
VIA0081576; 100|a 20021204d2000----|||y0itay50-----  
ba; 101||a ita; 102|a IT; 105|a ||||----0||||; 2001|a  
UNIMARC |f Antonio Scolari; 210|a Roma |c Associazione  
italiana biblioteche |d 2000; 215|a 110 p. |d 17 cm.;  
2252|a ET |e Enciclopedia tascabile |v 18; 410 0|1 2001 |a  
ET |e Enciclopedia tascabile |v 18; 619|a CATALOGAZIONE  
BIBLIOGRAFICA |x CODIFICA |x FORMATO UNIMARC |3  
MILC101687; 676|a 025.316 |v 21 |z ita |9 ANALISI E  
CONTROLLO BIBLIOGRAFICO. FORMATI DI  
REGISTRAZIONE PER CATALOGHI LEGGIBILI DALLA  
MACCHINA; 700 1|a Scolari, |b Antonio |c <1954- >; 801  
0|a IT |b SBN-MIL; 801 2|a IT |b Catalogo Collettivo  
Bicocca Insubria |g RICA; 950 0|a Insubria-Economia |c 1  
v. |d EVUFF 00057 |e EV 000007722500000000 1 v.; 950  
0|a Bicocca |c 1 v. |d BICDD. 025.30285 SCOA.UNI |e  
BI002000000009420010919/2000 1 v.

# Metadati strutturali

---

- Indicano come sono strutturati oggetti composti, ad esempio:
  - la sequenza delle pagine singole all'interno di un capitolo
  - la sequenza dei file in un oggetto digitale complesso

# Esempio di metadati strutturali

---

```
<structMap TYPE="logical">
<div ID="div1" LABEL="Oral History: Mayor Abraham Beame" TYPE="oral
  history">
<div ID="div1.1" LABEL="Interviewer Introduction" ORDER="1"> <fptr
  FILEID="FILE001"> <area FILEID="FILE001" BEGIN="INTVWBG"
  END="INTVWND" BETYPE="IDREF" /> </fptr> <fptr
  FILEID="FILE002"> <area FILEID="FILE002" BEGIN="00:00:00"
  END="00:01:47" BETYPE="TIME" /> </fptr> <fptr
  FILEID="FILE003"> <area FILEID="FILE003" BEGIN="00:00:00"
  END="00:01:47" BETYPE="TIME" /> </fptr> </div>
<div ID="div1.2" LABEL="Family History" ORDER="2"> <fptr
  FILEID="FILE001"> <area FILEID="FILE001" BEGIN="FHBG"
  END="FHND" BETYPE="IDREF" /> </fptr> <fptr FILEID="FILE002">
  <area FILEID="FILE002" BEGIN="00:01:48"END="00:06:17"
  BETYPE="TIME" /> </fptr> <fptr FILEID="FILE003"> <area
  FILEID="FILE003" BEGIN="00:01:48" END="00:06:17"
  BETYPE="TIME" /> </fptr> </div>
</div> </structMap>
(http://www.loc.gov/standards/mets/METSOverview.v2.html#structmap)
```

# Metadati amministrativi

---

- Forniscono informazioni utili alla gestione della risorsa descritta, ad esempio:
  - sulla creazione (es. la data)
  - tecniche (es. il tipo di file)
  - di profilo (es. le tipologie di destinatari della risorsa)

# Esempio di metadati amministrativi

---

**Type of Resource** software, multimedia

**Origin Information**

PlaceTerm:Washington, D.C

Publisher:Library of Congress

DateIssued:1998-]

DateIssued:Start (encoding="marc") 1998

DateIssued:End (encoding="marc") 9999

Issuance monographic

**Physical Description**

Form(marcform):Electronic

Form:Computer data and programs

LocationURL:<http://memory.loc.gov/ammem/dihtml/dihome.html>

**Record Information**

RecordContentSource:DLC

RecordCreationDate(encoding="marc"):980323

RecordChangeDate(encoding="iso8601"):20010614091341.0

RecordIdentifier:5004836

(<http://www.loc.gov/standards/mods/v3/mods98801326.html>)

# Metadati per il DRM

---

- Forniscono informazioni per la gestione elettronica dei diritti di proprietà intellettuale (Digital Rights Management), ad esempio:
  - Licenze
  - Autorizzazioni
  - Costi e modalità di accesso

# Esempio di metadati per il DRM

---

The rights depicted in this example grant the right to display the corresponding Content once, thus implementing the functionality to test-drive, i.e., preview, Content.

- ❑ `<o-ex:rights xmlns:o-ex=http://odrl.net/1.1/ODRL-EX xmlns:o-dd="http://odrl.net/1.1/ODRL-DD"`
- ❑ `xmlns:oma-dd=http://www.openmobilealliance.com/oma-dd`  
`xmlns:ds="http://www.w3.org/2000/09/xmldsig#"`
- ❑ `xmlns:xenc="http://www.w3.org/2001/04/xmlenc#">`
- ❑ `<o-ex:context><o-dd:version>2.0</o-dd:version><o-dd:uid>RightsObjectID</o-dd:uid></o-ex:context>`
- ❑ `<o-ex:agreement>`
- ❑ `<o-ex:asset><o-ex:context><o-dd:uid>ContentID</o-dd:uid></o-ex:context>`
- ❑ `<o-ex:digest><ds:DigestMethod`  
`ds:Algorithm="http://www.w3.org/2000/09/xmldsig#sha1"/>`
- ❑ `<ds:DigestValue>DCFHash</ds:DigestValue></o-ex:digest>`
- ❑ `<ds:KeyInfo><xenc:EncryptedKey>`
- ❑ `<xenc:EncryptionMethod Algorithm="http://www.w3.org/2001/04/xmlenc#kw-aes128"/>`
- ❑ `<xenc:CipherData><xenc:CipherValue>EncryptedCEK</xenc:CipherValue></xenc:CipherData></xenc:EncryptedKey>`
- ❑ `<ds:RetrievalMethod URI="REKReference"/></ds:KeyInfo></o-ex:asset>`
- ❑ `<o-ex:permission>`
- ❑ `<o-dd:display><o-ex:constraint><o-dd:count>1</o-dd:count></o-ex:constraint></o-dd:display>`
- ❑ `</o-ex:permission></o-ex:agreement></o-ex:rights>`  
(<http://xml.coverpages.org/OMADRM-RELv20-20040420.pdf>)

# Metadati per la conservazione

---

- Forniscono informazioni necessarie alla archiviazione e conservazione delle risorse, ad esempio:
  - Caratteristiche fisiche
  - Comportamento
  - Origine e modifiche



# Esempio di metadati per la conservazione

---

A behavior section can be used to associate executable behaviors with content in the METS object. A behavior section contains one or more <behavior> elements, each of which has an interface definition element that represents an abstract definition of the set of behaviors represented by a particular behavior section. A <behavior> also has a <mechanism> element which is used to point to a module of executable code that implements and runs the behavior defined abstractly by the interface definition. Digital object behaviors can be implemented as linkages to distributed web services.

```
<METS:behavior ID="DISS1.1" STRUCTID="S1.1" BTYPE="uva-
bdef:stdImage" CREATED="2002-05-25T08:32:00" LABEL="UVA Std
Image Disseminator" GROUPID="DISS1" ADMID="AUDREC1">
  <METS:interfaceDef LABEL="UVA Standard Image Behavior Definition"
  LOCTYPE="URN" xlink:href="uva-bdef:stdImage"/>
  <METS:mechanism LABEL="A NEW AND IMPROVED Image
  Mechanism" LOCTYPE="URN" xlink:href="uva-bmech:BETTER-
  imageMech"/></METS:behavior>
```

(<http://www.loc.gov/standards/mets/METSOverview.v2.html#behavior>)

# Livelli di descrizione – 1

---

- I metadati possono descrivere le risorse ad ogni livello di aggregazione (gerarchica):
  - Collezione
  - Risorsa in più parti
  - Risorsa singola
  - Componente di risorsa più estesa (es. pagina, illustrazione)

# Livelli di descrizione – 2

---

- I metadati possono anche descrivere le risorse ad ogni livello funzionale (es. FRBR: <http://www.ifla.org/VII/s13/frbr/frbr.htm>, primo gruppo di entità):
  - Opera (work)
  - Espressione (expression)
  - Manifestazione (manifestation)
  - Item (=)

# Embedded

---

- I metadati possono essere inclusi nell'oggetto digitale che descrivono, ad esempio:
  - nelle pagine web HTML
  - negli header dei file di immagine
  - nelle proprietà di un file PDF

# Esempio di pagina html con metadati embedded

---

```
❑ <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
❑         "http://www.w3.org/TR/html4/loose.dtd">
❑ <html>
❑ <head>
❑ <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
❑ <meta name="GENERATOR" content="Crimson Editor 3.60 a FreeWare
❑ Text Editor a FreeWare Text Editor">
❑ <meta name="keywords" content="OAI AEPIC open archives initiative
❑ electronic publishing e-publishing open access">
❑ <title>AEPIC Home Page</title>
❑ <link rel="stylesheet" type="text/css" href="CSS/aepic.css">
❑ <link href="http://www.cilea.it" rel="Copyright">
❑ <link href="agenda/agenda.htm" rel="Contents">
❑ <link href="index.htm" rel="Home">
❑ <link href="contatti/contatti.htm" rel="Author">
❑ <link href="documenti/documenti.htm" rel="Next">
❑ <link href="accessibilita.htm" rel="Help">
❑ </head>
❑ (http://www.aepic.it/)
```

# Esterni

---

- Oppure possono essere esterni all'oggetto e archiviati a parte, con link all'oggetto descritto, ad esempio:
  - in un database bibliografico
  - nel catalogo di una biblioteca

# Esempio di metadati esterni

---

- ❑ **ID Code:**775
- ❑ **Type:**Conference Paper
- ❑ **Deposited By:**[Mornati, Susanna](#)
- ❑ **Deposited On:**08 February 2004
- ❑ **Status:**Unpublished
- ❑ **Refereed:**No
- ❑ **Public Domain:**No
- ❑ **Authors:**Mornati, Susanna
- ❑ **Title:**Istanze "open" nella comunicazione scientifica: open archives
- ❑ **Subjects:**[E. Publishing and legal issues](#)
- ❑ **Year:**2003
- ❑ **Keywords:**open source, open access, open archives, copyleft, E-LIS, AEPIC
- ❑ Full text available as:
  - [PDF \(full text\)](#) - Requires [Adobe Acrobat Reader](#) or other PDF viewer. Language: it
  - [PDF \(slides\)](#) - Requires [Adobe Acrobat Reader](#) or other PDF viewer. Language: it

(<http://eprints.rclis.org/archive/00000775/fullmetadata.html>)

# Embedded vs. esterni - 1

---

- Vantaggi dell'embedding dei metadati:
  - archiviati con l'oggetto (non si perdono)
  - linking intrinseco
  - aggiornamento contestuale
- Ma non tutti gli oggetti possono includere metadati (ad es. una scultura)



# Embedded vs. esterni - 2

---

## □ Vantaggi dei metadati esterni:

- semplicità di gestione
- facilità di ricerca e recupero (inserimento in un database, con modalità di estrazione e indicizzazione)

# Ci siamo meritati un caffè?

---



# Seconda Parte

---

## Impiego dei metadati

# A cosa servono i metadati?

---

???

# Funzioni dei metadati

---

- Forniscono l'**identificazione** delle risorse che descrivono, facilitando la **presentazione** (discovery) e il **recupero**
- Consentono l'**organizzazione**, la **gestione**, la **localizzazione**, le **statistiche**
- Favoriscono l'**interoperabilità** e l'**integrazione** di risorse simili
- Facilitano l'**archiviazione** e la **conservazione**

# Esercitazione

---

- Individuare i metadati associabili ad un oggetto ed il possibile impiego

# Organizzazione delle risorse

---

- ❑ Liste e portali aggregano le risorse elettroniche sempre più numerose sul web
- ❑ È sempre più comune ed efficiente costruire liste dinamiche dei link
- ❑ I link a risorse esterne sono descritti in database ed associati a metadati

# Organizzare con i metadati

---

Es. lista risorse AEPIC

[www.aepic.it/risorse/risorse.htm](http://www.aepic.it/risorse/risorse.htm)

<http://bibliolinux.cilea.it:7000/xoops/html/>

Es. articoli Bollettino Cilea

<http://bibliolinux.cilea.it:7000/bollettino/>

<http://bibliolinux.cilea.it:7000/bollettino/rss.php>



# Interoperabilità: problemi

---

- Problemi di comunicazione fra sistemi con **diversità** di:
  - Hardware
  - Software
  - Struttura dei dati
  - Interfacce
  
- Lo scambio di dati comporta spesso **perdite** di contenuto e funzionalità.

# Interoperabilità: soluzioni

---

- **Protocolli** di trasferimento condivisi
- Schemi di metadati **definiti**
- **Crosswalks** (mappature) fra schemi di metadati diversi
  
- Due approcci all'interoperabilità (protocolli):
  - Ricerca attraverso i sistemi (es. Z39.50)
  - Harvesting dei metadati (es. OAI-PMH)

# Z39.50: ricerca cross-system

---

- Standard per l'information retrieval con architettura client/server
  
- Protocollo per la ricerca effettuata su sistemi diversi:
  - Mappatura delle possibilità di ricerca (indici)
  - Set comune di attributi (grammatica) che include le regole di ricerca
  
- Approfondimento: ANSI/NISO Z39.50-2003, <http://www.loc.gov/z3950/agency/>

# Z39.50: set di attributi (bib-1)

<http://lcweb.loc.gov/z3950/agency/defns/bib1.html>

---

- ❑ 1- use (punto d'accesso, es. titolo, data, ISBN)
- ❑ 2- relation (es. minore, uguale)
- ❑ 3- position (es. inizio, qualsiasi)
- ❑ 4- structure (es. frase, parola)
- ❑ 5- truncation (es. destra, sinistra)
- ❑ 6- completeness (es. la stringa occupa tutto il campo)

# Z39.50: Bath Profile

---

- Poiché il set di attributi bib-1 è molto complesso e nessun sistema lo applica integralmente, sono stati definiti profili con **attributi condivisi** per consentire il dialogo fra client e server
- Il profilo più utilizzato in campo bibliografico è il **Bath Profile**:  
<http://www.collectionscanada.ca/bath/tp-bath2-e.htm>

# OAI-PMH: harvesting

---

- I **data providers** traducono i propri metadati in un set comune per l'esportazione (DC simple)
- I **service providers** raccolgono i metadati dai set comuni e li indicizzano centralmente
  
- Approfondimento: OAI-PMH, Open Archives Initiative – Protocol for Metadata Harvesting, <http://www.openarchives.org/>

# OAI-PMH: Esempi

---

<http://rocky.dlib.vt.edu/~oai/cgi-bin/Explorer/2.0-1.45/testoai>

# Identificazione

---

- Per un oggetto digitale può essere il nome del file, unito al percorso che lo localizza:
  - URL (Uniform Resource Locator)
  - PURL <http://purl.org/> (Permanent URL)
  - DOI <http://www.doi.org/> (Digital Object Identifier)
  - gli stessi metadati combinati in una stringa identificativa (es. per ricerca con motore web)



# Archiviazione e conservazione

---

- L'evoluzione delle tecnologie decreta l'obsolescenza dei formati e delle piattaforme che li ospitano
- La migrazione (dei formati) e l'emulazione (dei comportamenti) sono due strategie per affrontare la sfida
- I metadati sono fondamentali per assicurare la sopravvivenza delle risorse digitali e la loro accessibilità in futuro:
  - documentano le caratteristiche fisiche del formato di un oggetto digitale
  - informano sul suo comportamento

# Schemi per la conservazione

---

- ISO Reference Model for an Open Archival Information System (OAIS):  
[http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)

Es. PDI (Preserving Digital Information):

**Content Information Type: Digital Library Collections**

**Reference: Bibliographic description, Persistent identifier**

**Provenance:**

- **For scanned collections: metadata about the digitisation process, pointer to master version**
- **For born-digital publications: pointer to the digital original**
- **Metadata about the preservation process: pointers to earlier versions of the collection item, change history**

**Context: Pointers to related documents in original environment at the time of publication**

**Fixity: Digital signature, Checksum, Authenticity indicator**

(<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf> , p.66)

# Terza parte

---

- Strutture, schemi, set di elementi

# Strutturare i metadati: schemi

---

- ❑ Gli schemi sono **insiemi di elementi** che descrivono un tipo particolare di risorsa
- ❑ Il **significato degli elementi** è noto come “semantica dello schema”
- ❑ I **valori dati agli elementi** costituiscono il contenuto

# Schemi di metadati: funzioni

---

- Specificano **nome** e **semantica** dei propri elementi
- Possono anche definire **regole** rispetto ai contenuti:
  - per formularli (es. come identificare il titolo)
  - per rappresentarli (es. lettera maiuscola per l'iniziale del titolo)
  - per controllarli (es. liste di autorità, vocabolari, tesauri)

# Schemi di metadati: sintassi

---

- Gli schemi possono definire regole sulla codifica gli elementi e dei contenuti (sintassi)
- Uno schema sintattico molto diffuso è l'XML (eXtensible Mark-up Language):
  - Forma estesa di HTML
  - Consente la definizione dei tag
  - Facilita lo scambio di informazione strutturata
- Approfondimento: XML (eXtensible Mark-up Language) è uno standard del consorzio W3C: <http://www.w3.org/XML/>

# Schemi più diffusi

---

- ❑ Dublin Core Metadata Element Set (**DC-MES**)
- ❑ Metadata Encoding and Transmission Standard (**METS**)
- ❑ Metadata Object Description Schema (**MODS**)
- ❑ ONline Information eXchange (**ONIX**)
- ❑ Molti altri (**TEI**, **LOM**, **EAD**, **DDI**, in Italia **MAG**)

# Dublin Core: origini

---

- ❑ Nato nel 1995 durante un workshop organizzato da OCLC e NCSA a Dublin (Ohio)
- ❑ Scopo originale: definire un set di elementi che i creatori di pagine web potessero impiegare per descrivere le proprie risorse



# Dublin Core: elementi

---

## □ 15 elementi:

- title
- creator
- subject
- description
- publisher
- contributor
- date
- type
- format
- identifier
- source
- language
- relation
- coverage
- rights

Esercizio: DC-assist

<http://www.ukoln.ac.uk/metadadata/dcassist/>

# Dublin Core: regole

---

- Tutti gli elementi sono ripetibili
- Tutti gli elementi sono opzionali
- Possono essere presentati in qualunque ordine
- Possono contenere valori controllati (es. type)
- Possono rientrare in application profiles (regole specifiche per dominio, es. istruzione, governo, etc.)

Traduzione italiana del Dublin Core:

<http://www.iccu.sbn.it/dublinco.html>

# Esempio di descrizione Dublin Core

---

Title="Unimarc"

Creator="Antonio Scolari"

Subject="metadati, MARC, catalogazione"

Description="Breve introduzione allo standard UNIMARC  
per la catalogazione bibliografica"

Publisher="Associazione Italiana Biblioteche"

Date="2000"

Type="text"

Identifier="ISBN:88-7812-061-8"

Language="it"

Rights="Copyright Associazione Italiana Biblioteche"

# DC fra minimalisti e strutturalisti

---

## □ Minimalisti:

- Mantenere pochi elementi
- Semantica e sintassi semplici

## □ Strutturalisti:

- Distinzioni semantiche più raffinate
- Estensioni per comunità

## □ Dublin Core simple

## □ Dublin Core qualified

# Dublin Core qualified

---

- Ogni elemento può essere dotato di qualificatori usati per rifinirlo, per restringerne l'ambito (es. `date.created`)
- Oppure può essere dotato di qualificatori che specificano uno schema di codifica:
  - es. data in ISO-8601:  
`2004-10-05T11.53CET`

# Dublin Core Metadata Initiative

---

- ❑ Mantiene il DC-MES (metadata element set)
- ❑ Incoraggia l'adozione di standard interoperabili di metadati
- ❑ Promuove lo sviluppo di vocabolari specializzati per sistemi di presentazione delle risorse
  
- ❑ Approfondimento: Sito web DCMI: <http://dublincore.org>

# Altri esempi di Dublin Core

---

<http://dublincore.org/documents/usageguide/elements.shtml>

<http://archive.dstc.edu.au/RDU/PICS/dc-in-rdf-ex.html>

[http://infoshare1.princeton.edu/katmandu/html/disc\\_09.html](http://infoshare1.princeton.edu/katmandu/html/disc_09.html)

<http://www.libraries.psu.edu/tas/jca/dublin/#examples>

DC-dot, generatore di dati Dublin Core per pagine web:

<http://www.ukoln.ac.uk/metadata/dcdot/>

# Metadata Encoding and Transmission Standard (METS)

---

- ❑ Struttura di dati standard per descrivere oggetti complessi in una biblioteca digitali
- ❑ Schema XML per descrivere gli oggetti, la struttura (dei files), i metadati amministrativi e tecnici
- ❑ Sviluppato dalla Digital Library Federation (DLF)



# METS: sezioni principali

---

- Header
- Metadati descrittivi (\*)
- Metadati amministrativi (\*)
- Lista dei files
- Mappa strutturale
- Link strutturali
- Comportamento

(\*) puntano a descrizioni esterne o estensioni, es.  
Dublin Core, MARCXML, MODS

# METS: un esempio

---

```
<metsHdr CREATEDATE="2003-07-04T15:00:00"
  RECORDSTATUS="Complete">
  <agent ROLE="CREATOR" TYPE="INDIVIDUAL"> <name>Jerome
    McDonough</name> </agent>
</metsHdr>
<dmdSec ID="dmd001"> <mdRef LOCTYPE="URN"
  MIMETYPE="application/xml" MDTYPE="EAD" LABEL="Berol Collection
  Finding Aid">urn:x-nyu:fales1735</mdRef>
</dmdSec>
<techMD ID="AMD001"> <mdWrap MIMETYPE="text/xml"
  MDTYPE="NISOIMG" LABEL="NISO Img. Data"> <xmlData>
  <niso:MIMETYPE>image/tiff</niso:MIMETYPE>
  <niso:Compression>LZW</niso:Compression>
</xmlData> </mdWrap> </techMD>
```

(<http://www.loc.gov/standards/mets/METSOverview.v2.html>)

# METS: estensioni e profili

---

- Prevede l'impiego di **estensioni** di metadati tecnici per ciascun tipo di risorsa digitale:
  - Es. MIX, Metadata for Images in XML, <http://www.loc.gov/standards/mix/>, schema (basato su NISO Z39.87)
- Prevede la creazione di **application profiles** per particolari oggetti specifici

# METS: altri esempi

---

- Overview e Tutorial
  - Schema
  - Documentazione
  - Diagramma della struttura
  - Profili
  - Registry di implementazione
- 
- Approfondimento: Sito METS:  
<http://www.loc.gov/standard/mets/>

# Metadata Object Description Schema (MODS)

---

- ❑ Deriva dal formato bibliografico MARC21 (subset di campi, anche riorganizzati)
- ❑ E' impiegato per l'importazione di record da MARC21 o per la creazione di nuovi record
- ❑ Utilizza tag basati sul linguaggio invece che numerici
- ❑ Si esprime in XML

# MODS: impieghi possibili

---

- Grazie alla sua flessibilità e all'impiego di XML, può essere utilizzato:
  - Come formato specificato da [Z39.50 Next Generation \(ZING\)](#)
  - Come schema di estensione per METS
  - Come set di metadati per l'harvesting
  - Per creare metadati in sintassi XML

# MODS: un esempio

---

```
<mods>
  <titleInfo>
    <title>UNIMARC</title>
  </titleInfo>
  <name type="personal">
    <namePart type="family">Scolari</namePart>
    <namePart type="given">Antonio</namePart>
    <role>
      <roleTerm authority="marcrelator">author</roleTerm>
    </role>
  </name>
  <originInfo>
    <dateIssued>2000</dateIssued>
    <publisher>Associazione Italiana Biblioteche</publisher>
  </originInfo>
  <identifier type="isbn">88-7812-061-8</identifier>
</mods>
```

# MODS: granularità

---

- La granularità delle componenti descrittive applicata ai diversi componenti di un oggetto si integra con la mappa strutturale di METS per gli oggetti digitali complessi
- Approfondimento: Sito MODS: <http://www.loc.gov/standards/mods/>
- Versione XML di MARC21 (MARCXML): <http://www.loc.gov/marc/marcxml.html>



# Metadati amministrativo-gestionali (MAG)

---

- Uno standard italiano (ICCU)
- Lo schema generale è composto da sei sezioni:
  - gen
  - bib
  - stru
  - img
  - ocr
  - doc
- Descrizione, documentazione, esempi, modelli:  
<http://www.iccu.sbn.it/schemag.htm>
- Esempio di software per la produzione di metadati in MAG: <http://www.metamag.it/>

# Altri schemi importanti - ONIX

---

**ONIX** (ONline Information eXchange)

<http://www.editeur.org/onix.html>

Standard internazionale per il commercio librario, creato da una collaborazione di editori europei e nordamericani

Contiene dati bibliografici, dati per la vendita e dati promozionali (immagini, recensioni)

Mappature verso MARC21 e UNIMARC rendono possibile la creazione di record bibliografici per i cataloghi delle biblioteche

# Altri schemi importanti - TEI

---

## TEI (Text Encoding Initiative)

<http://www.tei-c.org/>

Progetto internazionale per la codifica del mark-up di testi elettronici (narrativa, teatro, poesia) per supportare la ricerca nelle discipline umanistiche

Contiene una descrizione bibliografica della risorsa che può essere mappata in MARC per l'importazione in un catalogo di biblioteca

La DTD in SGML è molto complessa e ne viene spesso impiegato un subset noto come *TEI Lite*

# Altri schemi importanti - LOM

---

## LOM (Learning Object Metadata)

<http://ltsc.ieee.org/wg12/>

Standard per la descrizione di learning objects, consente uso e ri-uso di risorse per l'e-learning, l'insegnamento supportato dalla tecnologia (es. computer-based training e insegnamento a distanza)

Impiega attributi specifici in otto categorie: *general, lifecycle, technical, educational, rights, relation, annotation, classification*

[http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf)

# Altri schemi importanti - EAD

---

**EAD** (Encoded Archival Description)

<http://www.loc.gov/ead/>

Standard per la descrizione di finding aids per gli archivi, con header, descrizione della collezione e dei singoli record

# Altri schemi importanti - DDI

---

**DDI** (Data Documentation Initiative)

<http://www.icpsr.umich.edu/DDI/>

Standard per la descrizione di dataset  
per le scienze sociali e  
comportamentali

# Quarta parte

---

- Strumenti per la creazione ed il controllo

# Chi crea i metadati?

---

- Gli autori (pagine web, dataset scientifici, self-archiving)
- Tecnici specialisti (bibliotecari, archivisti)
- Produttori di database
- Editori (CIP, ONIX)
- Insegnanti (LOM)



# Strumenti per la creazione

---

- ❑ Commerciali e liberi
- ❑ Template: immissione dei dati attraverso campi facilitati, formattazione automatica
- ❑ Mark-up: per generare DTD, Document Type Definitions (XML, SGML)
- ❑ Estrazione: metadati generati automaticamente dall'analisi della risorsa (occorre revisione)
- ❑ Conversione: da un formato di metadati ad un altro (v. crosswalks)

# Criteri per la creazione

---

- Prestare attenzione agli errori di sintassi (elementi obbligatori, chiusura dei tag, terminologia coerente), utilizzare template
- Utilizzare metadati appropriati alla risorsa da descrivere, standard e interoperabili
- Utilizzare tesauri e liste di controllo (es. MIME Content Types per DC Format)
- Creare metadati persistenti
- Utilizzare strumenti di creazione o validazione
- Studiare manuali e linee guida

# Registri

---

- Possono applicarsi a vari livelli, dagli schemi ai profili, agli elementi, alle liste di valori per elemento
- Forniscono informazioni sulle definizioni
- Sono utili anche per le migrazioni e le conversioni

# Strumenti: esempi

---

- [Software per metadati](#)
- [RDF](#)
- [Dublin Core](#)
- [OAI](#)
- Registri: [DCMI Registry WG](#)

# Quinta parte

---

## Interoperabilità e scambio

# Perché tanti standard?

---

- ❑ Ogni schema ha un impiego distinto
- ❑ La stessa risorsa può essere descritta con schemi diversi a seconda del contesto (es. in MARC per biblioteca, in DC per repository)
- ❑ Esistono strumenti per l'integrazione di schemi diversi

# RDF – Resource Description Framework

---

- ❑ Sviluppata dal W3C, espressa in XML
- ❑ Specifica per la descrizione di risorse sul web
- ❑ Il namespace è un URL (punta alla risorsa remota che descrive lo schema utilizzato)
- ❑ Possono essere combinati namespace multipli (consentendo di utilizzare elementi da schemi diversi)

<http://www.w3.org/RDF/>

<http://www.w3.org/TR/REC-rdf-syntax/>

# RDF: un esempio

---

- ❑ `<?xml version="1.0"?>`
- ❑ `<!DOCTYPE rdf:RDF SYSTEM "http://purl.org/dc/schemas/dcmes-xml-20000714.dtd">`
- ❑ `<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntaxns#" xmlns:dc="http://purl.org/dc/elements/1.1/">`
- ❑ `<rdf:Description>`
- ❑ `<dc:title>UNIMARC</dc:title>`
- ❑ `<dc:creator>Scolari, Antonio</dc:creator>`
- ❑ `<dc:subject>schemi di catalogazione</dc:subject>`
- ❑ `<dc:description>Breve introduzione allo standard UNIMARC</dc:description>`
- ❑ `<dc:publisher>Associazione Italiana Biblioteche</dc:publisher>`
- ❑ `<dc:date>2000</dc:date>`
- ❑ `<dc:type>text</dc:type>`
- ❑ `</rdf:Description>`
- ❑ `</rdf:RDF>`



# Scambio e interoperabilità

---

- La migrazione, la conversione, l'importazione/esportazione di dati da un sistema all'altro, con diversi schemi di metadati, può essere molto complessa
- La ricerca fra dati provenienti da sistemi diversi può dare risultati incoerenti o incompleti

# Le mappature (crosswalks)

---

Queste operazioni sono facilitate dai crosswalks, che mappano:

- gli elementi
- la semantica
- la sintassi

fra schemi diversi di metadati, facilitando la costruzione di biblioteche virtuali

# Problemi per i crosswalks

---

- Differenze di **granularità** fra gli schemi (specialmente mappare schemi con meno elementi verso schemi più ricchi)
- **Compatibilità** fra le regole di contenuto per la compilazione degli elementi

# Crosswalks: esempi e bibliografia

---

- ❑ Dublin Core – UNIMARC:  
<http://www.iccu.sbn.it/dubluni.html>
- ❑ MARC21 – Dublin Core:  
<http://www.loc.gov/marc/marc2dc.html>
- ❑ Dublin Core – MARC21:  
<http://www.loc.gov/marc/dccross.html>
  
- ❑ All about crosswalks:  
[http://www.oclc.org/research/projects/mswitch/1\\_crosswalks.htm](http://www.oclc.org/research/projects/mswitch/1_crosswalks.htm)
- ❑ Issues in crosswalking content metadata standard:  
<http://www.niso.org/press/whitepapers/crswalk.html>
- ❑ Mapping between metadata formats:  
<http://www.ukoln.ac.uk/metadata/interoperability/>
- ❑ Metadata mappings:  
<http://libraries.mit.edu/guides/subjects/metadata/mappings.html>

# Sesta parte

---

## Futuro dei metadati

# Stato dell'arte

---

- I primi standard si sono concentrati sulla descrizione delle risorse
- Molti progetti in corso sono focalizzati sui metadati tecnici
- Lo sviluppo di metadati tende alla standardizzazione, ad esempio:
  - Dublin Core nasce nel 1995
  - Nel 2001 diventa standard ANSI/NISO (Z39.85)
  - Nel 2003 diventa standard internazionale (ISO 15836)

# Attività del W3C

---

- L'attività sui metadati è stata incorporata al Semantic Web (infrastruttura comune per la condivisione dei dati e il riuso attraverso diverse applicazioni e comunità)
- RDF è uno degli standard chiave per raggiungere questo obiettivo
- La chiave è l'interoperabilità

# In quale direzione?

---

- ❑ Creare nuova informazione non basta: occorre renderla accessibile
- ❑ I metadati sono la chiave per l'accesso alle risorse informative
- ❑ Metadati e Semantic Web sono il futuro dell'informazione



# Per saperne di più

---

- ❑ Digital libraries: metadata resouces  
<http://www.ifla.org/II/metadata.htm>
- ❑ Introduction to metadata: pathways to digital information  
[http://www.getty.edu/research/conducting\\_research/standards/intrometadata/index.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html)
- ❑ Metadata Information Clearinghouse Interactive (MICI) <http://www.metadainformation.org>
- ❑ Metadata resources  
<http://www.ukoln.ac.uk/metadata/resources>
- ❑ [http://www.chin.gc.ca/English/Standards/metadata\\_intro.html](http://www.chin.gc.ca/English/Standards/metadata_intro.html)
- ❑ <http://staff.library.mun.ca/staff/toolbox/standards.html>
- ❑ <http://www.iccu.sbn.it/metadati.htm>
- ❑ <http://xml.coverpages.org/>

# CREDITS ☺

---

Per parte della struttura e dei contenuti di questo corso sono debitrice nei confronti di:

- ☐ **Understanding Metadata**, NISO 2004,  
<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

Per preziosi consigli ed idee:

- ☐ **Antonella De Robbio**, Università di Padova,  
Last, but not least, per avere sempre creduto in me:
- ☐ **Valentina Comba**, Università di Bologna.

# Consigli ... ;-)

---

- ❑ Non diventate bibliotecari!  
<http://www.giampierorselli.it/pseudokkupati.htm>
- ❑ Imparate **XML** (e l'inglese)!  
<http://www.html.it/xml/guida/index.html>
- ❑ Siete voi ad usare il computer, non il computer ad usare voi!  
<http://www.studenti.math.unipd.it/manuali.html>

---

Grazie per l'attenzione!