

Metadati

Editoria digitale

Maria A. Alberti
Informatica per la comunicazione digitale
AA 17/18

Organizzare l'informazione

I metadati sono ovunque, invisibili ma presenti

- Questa lezione ha un titolo, un docente, una durata
- Il principe, Nicolò Machiavelli
 - <https://archive.org/details/ilprincipe04machgoog>
- Un video su YouTube
 - Titolo, autore, data, durata, descrittori, categorie e licenze, commenti e video collegati
 - Metadati generati dall'autore (soggetto), scelti da una lista limitata, i commenti creati da altri, metadati generati automaticamente dal sistema

Metadati

Alcune categorie di informazioni sono comuni a diverse categorie di oggetti.

- Autore, titolo, data

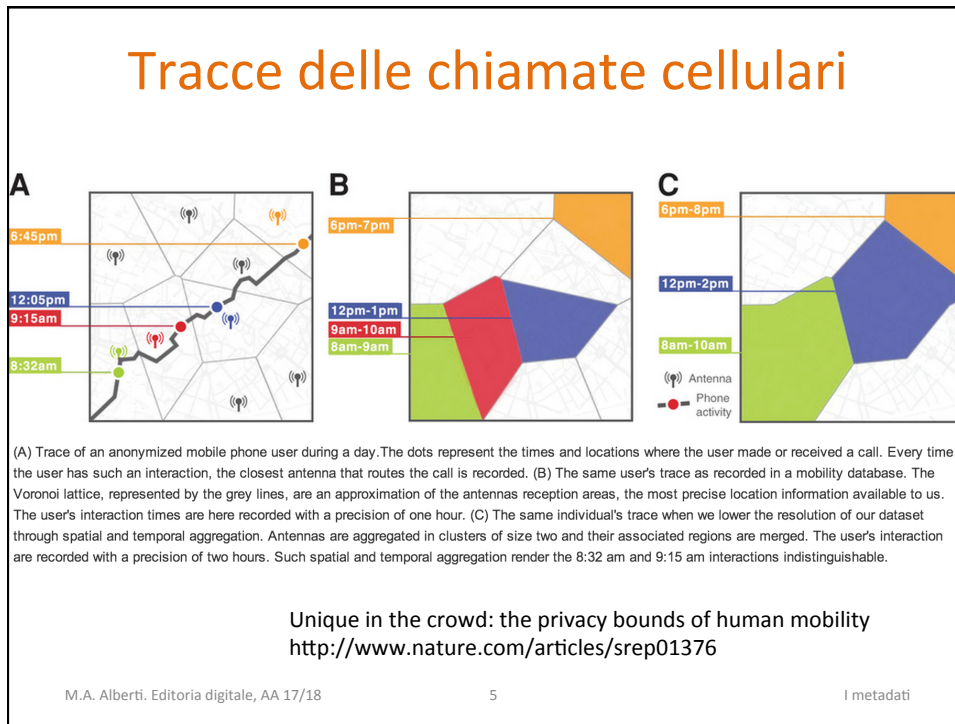
Altre sono specifiche del tipo dell'oggetto.

- Durata di un video, il numero delle pagine di un libro

Il caso delle telefonate NSA - 2013

National Security Agency - NSA raccoglie dati sulle telefonate

- Non il contenuto delle telefonate
- Ma i metadati di queste telefonate
 - Il numero da cui parte la telefonata
 - Il numero cui si telefona
 - Il ripetitore che trasmette la telefonata e che consente di definire l'area geografica
 - L'ora e la durata della telefonata



Metadati sono dati

- Metadati sono dati sui dati
- Metadati sono descrizioni di un qualunque oggetto
- Le descrizioni sono affermazioni su un oggetto, informazioni sull'oggetto
 - Il libro è rosso

```

graph LR
    S([Soggetto]) -- Predicato --> O([Oggetto])
            
```

M.A. Alberti. Editoria digitale, AA 17/18 6 I metadati

Perché descrivere?

Per potere catalogare, archiviare e poi ritrovare, quindi utilizzare

Necessità di punti di accesso per ritrovare i dati. I classici:

- Titolo, autore e soggetto

Metadati descrittivi

- I nomi delle strade, i numeri civici

Metadati amministrativi: informazioni su come maneggiare o gestire qualcosa

- Etichette di imballaggio

Etichette imballaggio



Cosa sono i dati e cosa i metadati?

- Il censimento raccoglie metadati descrittivi sulla famiglia
- Queste info rese pubbliche diventano dati
 - I metadati sono diventati dati
 - Dati sui metadati sono meta-metadati
- Conclusione: cosa siano i dati e cosa siano i metadati dipende dal punto di vista

Analisi del soggetto

In un libro il soggetto descrive il contenuto oltre al titolo, che non è sempre descrittivo

- Di che parla Moby Dick?
- Parla di balene? O di ossessione e vendetta?
- Come si determina cosa è significativo e caratteristico dell'oggetto da descrivere?
- Da dove vengono le parole con cui si descrive un oggetto?

Analisi del soggetto

Analisi del soggetto può essere problematica

Non sempre è possibile l'analisi del soggetto

La musica, ad esempio, spesso non ha un soggetto dichiarato

- Le quattro stagioni, Pierino e il lupo, Il volo del calabrone
- Il concerto per violino e orchestra, la IX sinfonia

Oggetto vs collezione

Una distinzione importante nella descrizione

- Item è un singolo oggetto
- Collezione è una raccolta di oggetti

Un articolo vs il giornale

<http://firstmonday.org/ojs/index.php/fm/index>

Una fotografia vs l'intera esibizione

<http://exhibitions.europeana.eu/exhibits/show/europe-america-en>

Library of Congress Subject Headings

- LCSH dal 1898 il prototipo di un sistema di metadati per la catalogazione del materiale librario
- Da non confondere con la classificazione per collocare gli oggetti sugli scaffali, Library of Congress Classification
 - fornisce l'identificativo alfa-numerico, *call number*, o *collocazione*, che viene posto sul dorso dei libri per poterli ritrovare negli scaffali
 - Definito da Putnam nel 1897

Library of Congress Classification



Call numbers

argomento

autore

Subclass QC	
QC1-999	Physics
QC1-75	General
QC81-114	Weights and measures
QC120-168.85	Descriptive and experimental mechanics
QC170-197	Atomic physics. Constitution and properties of matter Including molecular physics, relativity, quantum theory, and solid state physics
QC221-246	Acoustics. Sound
QC251-338.5	Heat
QC310.15-319	Thermodynamics
QC350-467	Optics. Light
QC450-467	Spectroscopy
QC474-496.9	Radiation physics (General)
QC501-766	Electricity and magnetism
QC501-(721)	Electricity
QC669-675.8	Electromagnetic theory
QC676-678.6	Radio waves (Theory)
QC701-715.4	Electric discharge
QC717.6-718.8	Plasma physics. Ionized gases
QC750-766	Magnetism
QC770-798	Nuclear and particle physics. Atomic energy. Radioactivity
QC793-793.5	Elementary particle physics
QC794.95-798	Radioactivity and radioactive substances
QC801-809	Geophysics. Cosmic physics
QC811-849	Geomagnetism
QC851-999	Meteorology. Climatology Including the earth's atmosphere
QC974.5-976	Meteorological optics
QC980-999	Climatology and weather
QC994.95-999	Weather forecasting

M.A. Alberti. Editoria digitale, AA 17/18

I metadati

Library of Congress Classification



M.A. Alberti. Editoria digitale, AA 17/18

16

I metadati

LCSH, dati sul soggetto del libro

The Getty Research Institute Publications Program

Thomas Gatchgens, *Director, Getty Research Institute*
Gail Feigenbaum, *Associate Director, Programs*
Julia Bloomfield, *Head, Publications Program*

Introduction to Metadata

Second Edition

Murtha Baca, *Series Editor*
Patrick E. Pardo, *Project Editor*
Sheila U. Berg, *Manuscript Editor*
Elizabeth Zozom, *Production Coordinator*
Designed by Hespeneide Design, Newbury Park, California
Printed and bound by Odyssey Press Inc.

Library of Congress Cataloging-in-Publication Data

Introduction to metadata / edited by Murtha Baca. — 2nd ed.
p. cm.
Includes bibliographical references.
ISBN 978-0-89236-896-9 (pbk.)
1. Database management. 2. Metadata. 3. World Wide Web.
4. Information organization. I. Baca, Murtha. II. Gety Research Institute.
QA76.9.D3f599 2008
025.3—dc22

2008007871

awkward locutions, it has been treated here throughout as singular.

This volume was published in September 2008. Updated print editions will be offered periodically. An online version of this edition is also available at: www.getty.edu/research/conducting_research/standards/intrometadata

Parole chiave tratte da LCSH

Monday, September 16, 13

M.A. Alberti. Editoria digitale, AA 17/18

17

I metadati

I volumi del LCSH



M.A. Alberti. Editoria digitale, AA 17/18

18

I metadati

Tesauro o vocabolario controllato

- Definisce le entità dell'universo che trattiamo e indica le parole che dobbiamo usare per parlarne
- LCSH fornisce la struttura di parole per definire i soggetti, che può essere modificata ed estesa, mediante un vocabolario controllato e strutturato
- Un tesauro e un dizionario differiscono
 - Dizionario: elenco di parole con il loro significato
 - Tesauro: elenco di parole e di relazioni tra loro

Il vocabolario controllato LCSH

- Un elenco strutturato di parole, raggruppate per semantica, con alcune relazioni
 - BT – broader term, il termine più generale
 - NT – narrower term, il termine più specifico
 - RT – related term, un termine associato
 - USE – rimando a un termine da usarsi
 - UF – use for, rimando a un termine da non usarsi
- Le relazioni BT e NT definiscono la gerarchia di tutte le entità possibili
- Le relazioni USE e UF stabiliscono i termini standard o canonici da usarsi per le entità

Il vocabolario controllato LCSH

BT Islands—Washington (D.C.) Columbia Plateau UF Channeled Scabland Columbia and Snake River Plateau Columbia River Plateau Columbia Plateau Scabland, Channeled BT Plateaus—Idaho Plateaus—Oregon Plateaus—Washington (State) Columbia Quarry Site (Ill.) (Not Subd Geog) BT Illinois—Antiquities Columbia Reservoir (Tenn.) BT Reservoirs—Tennessee Columbia River (Not Subd Geog) BT Rivers—British Columbia Rivers—Oregon Rivers—Washington (State) NT Priest Rapids (Wash. : Rapids) — Channels (Not Subd Geog) NT Hanford Reach (Wash.) — Power utilization NT Revelstoke Project Columbia River Gorge (Or. and Wash.) UJF Columbia Gorge (Or. and Wash.) BT Gorges—Oregon Gorges—Washington (State) Columbia River Plateau USE Columbia Plateau Columbia University — Buildings NT Arden House (Harriman, N.Y.) — Graduate students Columbia University. School of Law. Blackstone Moot Court Columbian (Express train) BT Railroads—United States—Express- trains	Columbus (Miss.) — History — Civil War, 1861-1865 Columbus & Greenville Railroad BT Railroads—United States Columbus and Xenia Railroad BT Railroads—United States Columbus Catholic Cemetery (Columbus, Ohio) UJF Catholic Cemetery (Columbus, Ohio) BT Cemeteries—Ohio Columbus Day (E120) BT Holidays Columbus Quincentenary, 1992-1993 (May Subd Geog) (E119.2) UJF Christopher Columbus Quincentenary, 1992-1993 Columbian Quincentenary, 1992-1993 Columbus, Christopher—Anniversaries, etc., 1992-1993 BT America—Discovery and exploration— Centennial celebrations, etc. Columbus test (BF698.8.C64) BT Picture interpretation tests Projective techniques Social maturity scales Columella auris (QL948) (QM507) UF Stapes, Non-mammalian BT Ear Ear ossicles Columella auris, Mammalian USE Stapes Columella 21	(TA681-TA683) UF Concrete columns BT Concrete construction — Fatigue Columns, Corinthian (NA2860) UF Corinthian columns BT Architecture, Greek NT Capitals (Architecture), Corinthian Columns, Doric (NA2860) UF Doric columns BT Architecture, Greek Columns, Ionic (NA2860) UF Ionic columns BT Architecture, Greek NT Capitals (Architecture), Ionic Columns, Iron and steel (May Subd Geog) (TA492.C7 (Testing)) (TH1610-TH1625 (Building)) UF Iron and steel columns Steel and iron columns BT Cast-iron Columns Iron, Structural Steel, Structural Columns, Packed USE Packed towers Columns, Plate USE Plate towers Columns, Salomónica USE Columns, Spiral Columns, Spiral (May Subd Geog) (NA2860) UF Barley-sugar columns Berninesque columns Columns, Barlesque
---	---	---

Monday, September 16, 13

I metadati

Columbia Reservoir (Tenn.)
 BT Reservoirs—Tennessee
Columbia River (Not Subd Geog)
 BT Rivers—British Columbia
 Rivers—Oregon
 Rivers—Washington (State)
 NT Priest Rapids (Wash. : Rapids)
 — Channels (Not Subd Geog)
 NT Hanford Reach (Wash.)
 — Power utilization
 NT Revelstoke Project
Columbia River Gorge (Or. and Wash.)
 UJF Columbia Gorge (Or. and Wash.)
 BT Gorges—Oregon
 Gorges—Washington (State)
Columbia River Plateau
 USE Columbia Plateau
Columbia University
 — Buildings
 NT Arden House (Harriman, N.Y.)
 — Graduate students
Columbia University. School of Law.
 Blackstone Moot Court
Columbian (Express train)
 BT Railroads—United States—Express-
 trains
Columbian Plateau
 USE Columbia Plateau
Columbian Quincentenary, 1992-1993

M.A. Alberti. Editoria digitale, AA 17/18

I metadati

National Library of Medicine - Medical Subject Headings

2014 MeSH

MeSH Descriptor Data

[Return to Entry Page](#)

Standard View. [Go to Concept View](#); [Go to Expanded Concept View](#)

MeSH Heading	Patella
Tree Number	A02.835.232.043.650.624
Annotation	/inj permitted: do not convert to KNEE INJURIES
Scope Note	The flat, triangular bone situated at the anterior part of the KNEE .
Allowable Qualifiers	AB AH BS CH CY DE EM EN GD IM IN IR ME MI PA PH PP PS RA RE RI SU TR UL US VI
Date of Entry	19990101
Unique ID	D010329

MeSH Tree Structures

Esistono altri vocabolari controllati con un obiettivo più specifico in campi più ristretti di quello universale considerato dal LCSH. Qui un esempio del campo medico non così ampio come LCSH ma va più in profondità. Breath vs depth

[Musculoskeletal System \[A02\]](#)

[Skeleton \[A02.835\]](#)

[Bone and Bones \[A02.835.232\]](#)

[Bones of Lower Extremity \[A02.835.232.043\]](#)

[Leg Bones \[A02.835.232.043.650\]](#)

[Femur \[A02.835.232.043.650.2471\] +](#)

[Fibula \[A02.835.232.043.650.321\]](#)

[Patella \[A02.835.232.043.650.624\]](#)

[Tibia \[A02.835.232.043.650.883\]](#)

23

I metadati

MeSH

Entità e relazioni

Term	Information Science
Scope note	Study of the gathering, organizing, storing, retrieving, and dissemination of information.
Broader term	(fields and disciplines)
Narrower term	Archival Science, Informatics, Information Architecture, Information Theory, Knowledge Management, etc.
Related terms	Cognitive Science, Computer Science, Cybernetics, Information Retrieval, Information Technology, Linguistics, etc.
Use for	Library and Information Science, LIS

Dal tesaurus di ASIST (American Society for Information Science and Technology)
Ogni gerarchia è soggettiva

M.A. Alberti. Editoria digitale, AA 17/18

24

I metadati

Classificazioni multiple

Ogni oggetto può essere classificato da diversi punti di vista e quindi essere descritto con diversi vocabolari controllati

- Un monumento classificato dal punto di vista della locazione geografica, dello stile architettonico, dell'uso, dei materiali usati...
- Un esempio di classificazione multipla è data dall'Art & Architecture Thesaurus del Paul Getty Institute che si trova online:

<http://www.getty.edu/vow/AATHierarchy?find=subjects&logic=AND¬e=&page=1&subjectid=300000000>

Ontologie

In Informatica: la rappresentazione formale di un insieme di concetti in un determinato ambito

- Una lista di parole e un insieme di relazioni più complesse di quelle nei vocabolari controllati

Ontologia della birra

- <http://www.cs.umd.edu/projects/plus/SHOE/onts/beer1.0.html>

Ontologia per descrivere le persone che usa inferenze per stabilire fatti nuovi implicati da quelli noti

- <http://www.cs.umd.edu/projects/plus/SHOE/onts/personal1.0.html>

Tesauri vs ontologie

- Le ontologie sono tesauri con relazioni più complesse e che hanno una logica interna
- Nelle ontologie entità, relazioni e inferenze consentono di dedurre fatti non esplicitamente asseriti
- Le ontologie possono quindi essere usate in algoritmi

Vocabolari non controllati

Ogni parola e ogni frase può essere usata per classificare

- Al contrario dei vocabolari controllati in cui le parole usate per descrivere sono limitate per esercitare il controllo sui metadati descrittivi

I tag delle reti sociali sono un esempio di vocabolario non controllato:

- *YouTube* non limita nell'uso di parole chiave
 - https://www.youtube.com/upload_defaults
- *FaceBook* limita a termini che devono già essere stati conati nell'universo FB
- *Twitter* usa gli hashtag, non tutti i tag sono descrittivi. A metà strada tra contenuto e metadato: i tag finiscono per avere un'area semantica di riferimento
- *Flicker* usa tag completamente liberi

Metadati per documenti librari

- Descrittivi – info sulla risorsa
 - Titolo, autore, formato
- Strutturali – info su come è organizzata
 - Una mappa della struttura: libro, capitoli, sezioni, pagine
- Amministrativi – info su come deve essere conservata
 - Diritto e conservazione
 - Copyrights, proprietà intellettuale, accesso, provenienza, proprietà, origine

Considerazioni generali

- Distinguiamo il **livello** dell'oggetto e quello della collezione
 - Qualche incertezza su cosa sia un oggetto singolo
- Scheda dei metadata e oggetto: che relazione? Dove è collocata?
 - *Embedded* metadata: come nell'intestazione di libro
 - *Linked* metadata: come nel catalogo delle biblioteche che fornisce un numero di catalogo che indica dove sia archiviato l'oggetto sugli scaffali

Metadati embedded

The Getty Research Institute Publications Program

Thomas Gachtgens, *Director, Getty Research Institute*
Gail Feigenbaum, *Associate Director, Programs*
Julia Bloomfield, *Head, Publications Program*

Introduction to Metadata

Second Edition

Murtha Baca, *Series Editor*
Patrick E. Pardo, *Project Editor*
Sheila U. Berg, *Manuscript Editor*
Elizabeth Zozom, *Production Coordinator*
Designed by Hespeneide Design, Newbury Park, California
Printed and bound by Odyssey Press Inc.

Library of Congress Cataloging-in-Publication Data

Introduction to metadata / edited by Murtha Baca. — 2nd ed.
p. cm.
Includes bibliographical references.
ISBN 978-0-89236-896-9 (pbk.)
1. Database management. 2. Metadata. 3. World Wide Web.
4. Information organization. I. Baca, Murtha. II. Getty Research Institute.
QA76.9.D3I599 2008
025.3—dc22

awkward locutions, it has been treated here throughout as singular.

This volume was published in September 2008. Updated print editions will be offered periodically. An online version of this edition is also available at: www.getty.edu/research/conducting_research/standards/intrometadata

Parole chiave tratte da LCSH

Monday, September 16, 13

M.A. Alberti. Editoria digitale, AA 17/18

31

I metadati

Metadati linked: schede di catalogo



M.A. Alberti. Editoria digitale, AA 17/18

32

I metadati

Audience

leggibilità - uomo vs macchina

- Per chi sono intesi i metadati?
- Il catalogo in biblioteca è per un lettore che deve trovare il libro
- Una scheda MARC (Machine Readable Cataloging) ha alcune parti leggibili ma altre concepite per essere elaborate automaticamente
- HTML separa ciò che è leggibile da un umano o da una macchina

Standard MARC

- Un formato di catalogazione introdotto negli anni '60 dall'US Library of Congress
 - Divenuto standard internazionale negli anni '70
 - Evoluto in una famiglia di standard
 - <http://www.loc.gov/marc/umb/>
- Una tipica scheda include:
 - La descrizione dell'oggetto
 - I campi principali, access point
 - Le parole chiave per soggetto
 - La classificazione e il numero di catalogazione
 - Molte altre informazioni

START OVER REGULAR DISPLAY RETURN TO BROWSING ANOTHER SEARCH

```
LEADER 00000cam 2200000 a 4500a
001 50906019
003 OCoLC
005 20030715092633.0
008 021023s2003 ilu b 001 0 eng
010 2002151683
015 GBA3-Y7095
020 0838908470
040 DLC|cDLC|dUKM|dc#P|dXFF|dKSU|dCoLC
049 KSUU
050 00 Z666.5|b.C37 2003
082 00 025.3|221
100 1 Caplan, Priscilla
245 10 Metadata fundamentals for all librarians /|cPriscilla
    Caplan
260 Chicago :|bAmerican Library Association,|c2003
300 ix, 192 p. ;|c28 cm
504 Includes bibliographical references and index
505 00 |tMetadata basics --|tSyntax, creation, and storage --
    |tVocabularies, classification, and identifiers --
    |tApproaches to interoperability --|tMetadata and the Web
    --|tLibrary cataloging --|gThe|tTEI header --|gThe|tDublin
    Core --|tArchival description and the EAD --|tMetadata for
    art and architecture --|tGILS and government information -
    -|tMetadata for education --|tONIX International --
    |tMetadata for geospatial and environmental resources --
    |gThe|tData Documentation Initiative --|tAdministrative
    metadata --|tStructural metadata --|tRights metadata
650 0 Metadata
650 0 Information organization
```

I metadati

Una scheda MARC

Parlando di metadati

Occore specificare

- Il livello della descrizione: item vs collection
 - Se si usano vocabolari controllati o no
- La locazione dei metadati: embedded vs linked
- Per chi sono intesi, chi li deve leggere: macchina vs uomo