

Flussi di Gestione Documentale

Paolo Ceravolo

paolo.ceravolo@unimi.it

Editoria Digitale

- Nella produzione editoriale possono esistere diversi flussi di creazione e gestione dei documenti
- Differiscono in base a
 - Flusso di lavoro
 - Qualità contenuti
 - Qualità formati, es. citazioni
 - Tecnologie
 - Creazione
 - Condivisione
 - Produzione

Metadati

- I metadati sono ovunque, a volte invisibili ma presenti
- Questa lezione ha un titolo, un docente, una durata, dei CdL
- Un video su YouTube
 - Titolo, autore, data, durata, categorie e licenze
 - Metadati generati dall'autore, scelti da una lista limitata, i commenti creati da altri, metadati generati automaticamente dal sistema
- Una risorsa di archivio
 - *Il principe, Nicolò Machiavelli*
 - <https://archive.org/details/ilprincipe04machgoog>

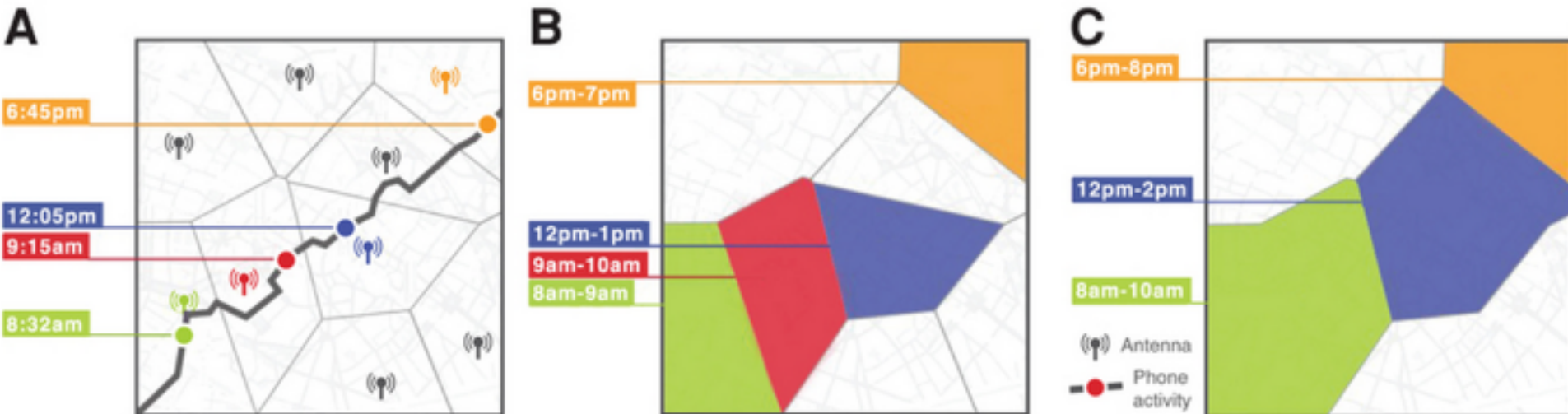
- Alcuni descrittori sono comuni a diverse categorie di oggetti.
 - Autore, titolo, data
- Altri sono specifiche del tipo dell'oggetto.
 - Durata di un video, il numero delle pagine di un libro
- Alcuni sono stabili nel tempo altri variano
 - Autore, soggetto vs commenti, voci correlate
- Alcuni valgono per ogni tipo di interazione altri dipendono dall'interazione
 - Titolo del corso vs CdL
- Alcuni dipendono dal processo di raccolta
 - Tabulati telefonici

- Nei tabulati telefonici non si raccoglie il contenuto delle telefonate
- Si raccolgono 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

IL CASO DEI TABULATI TELEFONICI



UNIVERSITÀ
DEGLI STUDI
DI MILANO



(A) Trace of an anonymized mobile phone user during a day. The dots represent the times and locations where the user made or received a call. Every time the user has such an interaction, the closest antenna that routes the call is recorded. (B) The same user's trace as recorded in a mobility database. The Voronoi lattice, represented by the grey lines, are an approximation of the antennas reception areas, the most precise location information available to us. The user's interaction times are here recorded with a precision of one hour. (C) The same individual's trace when we lower the resolution of our dataset through spatial and temporal aggregation. Antennas are aggregated in clusters of size two and their associated regions are merged. The user's interaction times are recorded with a precision of two hours. Such spatial and temporal aggregation render the 8:32 am and 9:15 am interactions indistinguishable.

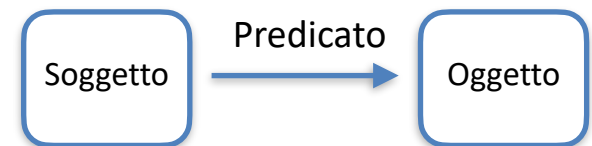
Unique in the crowd: the privacy bounds of human mobility
<http://www.nature.com/articles/srep01376>



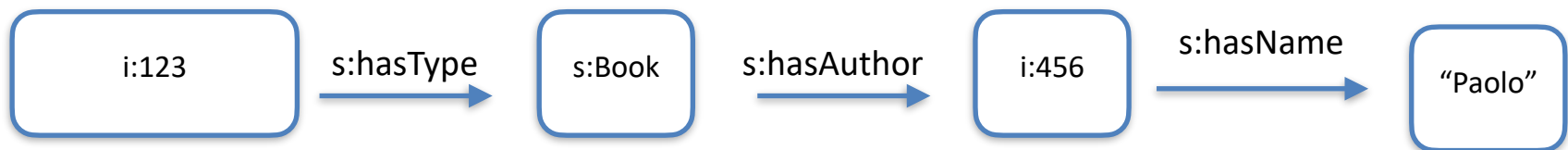
UNIVERSITÀ
DEGLI STUDI
DI MILANO

SESAR LAB

- Metadati sono dati sui dati
- Metadati sono descrizioni di un oggetto
- Le descrizioni sono affermazioni su un oggetto, informazioni sull'oggetto
- Siccome possiamo collegare oggetti a risorse informative che li descrivono e identificano possiamo dire che i metadati possono riferirsi a qualsiasi oggetto
 - Questa risorsa è un libro
 - Questo libro ha un autore
 - Questo autore ha un nome



- Metadati sono dati sui dati
- Metadati sono descrizioni di un oggetto
- Le descrizioni sono affermazioni su un oggetto, informazioni sull'oggetto
- Siccome possiamo collegare oggetti a risorse informative che li descrivono e identificano possiamo dire che i metadati possono riferirsi a qualsiasi oggetto



COSA SONO I DATI E COSA I METADATI?



UNIVERSITÀ
DEGLI STUDI
DI MILANO

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



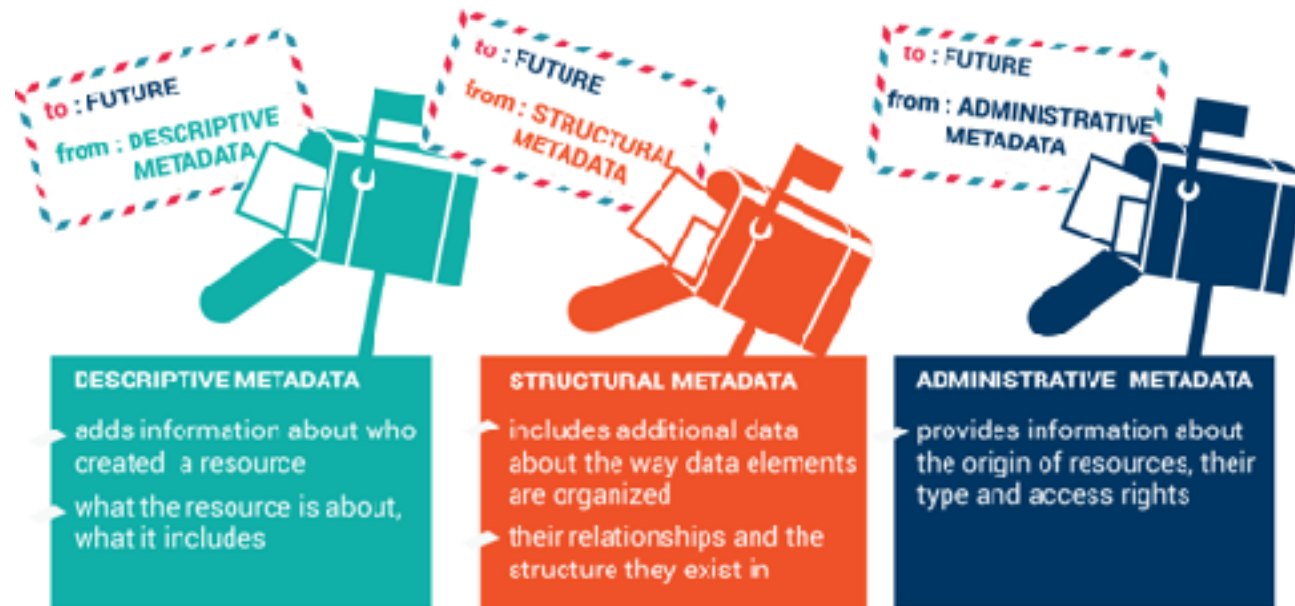
UNIVERSITÀ
DEGLI STUDI
DI MILANO

 SESAR LAB

- Per potere catalogare e concettualizzare i dati
 - Necessità di punti di accesso per ritrovare i dati
 - Necessità di un vocabolario condiviso per rendere interoperabile lo scambio di informazioni e per attivare processi di automazione
- <https://www.crossref.org/documentation/>
- Per potere disporre l'uso dei dati
 - Necessità di gestire il ciclo di vita di un dato: revisioni, versioning, autorizzazioni, diritti d'uso
 - Necessità di strumenti di validazione per verificare l'integrità del dato, la sua compatibilità con altri dati, l'autenticità

- Ci sono diverse tipologie di metadati
 - Descrittivi
 - Strutturali
 - Amministrativi
 - Tecnici
 - Utilizzo

Esempi



- 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>

- 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>

<i>Standard di strutturazione dei dati</i>	MARC (Machine-Readable Cataloging) Format, Encoded Archival Description (EAD), BIBFRAME (Bibliographic Framework), Dublin Core Metadata Element Set , Categories for the Description of Works of Art, VRA Core
<i>Standard di valori dei dati (vocabolari controllati, thesauri, liste controllate)</i>	Library of Congress Subject Headings, VIAF , ORCID , Name Authority File, and Thesaurus for Graphic Materials; Getty Art & Architecture Thesaurus, Dewey Decimal Classification (DDC), and Thesaurus of Geographic Names; ICONCLASS;
<i>Standard di organizzazione sintattica dei dati (regole e codici di catalogazione)</i>	Anglo-American Cataloguing Rules, Resource Description and Access, International Standard Bibliographic Description, Cataloging Cultural Objects, Describing Archives: A Content Standard
<i>Formato dei dati/standard tecnici di interscambio</i>	Resource Description Framework, MARC21, MARCXML, EAD XML DTD, METS, BIBFRAME, LIDO XML, Simple Dublin Core XML, Qualified Dublin Core XML, VRA Core 4.0 XML

- Un formato di catalogazione introdotto negli anni '60 dal 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

STANDARD MARC



UNIVERSITÀ
DEGLI STUDI
DI MILANO

START
OVER

REGULAR
DISPLAY

RETURN TO
BROWSE

ANOTHER
SEARCH

(Search History)

```
LEADER 00000cam 2200000 a 4500a
001 50906019
003 oCoLC
005 20030715092633.0
008 021023s2003 ilu b 001 0 eng
010 2002151683
015 GBA3-Y7D95
020 0838908470
040 DLC|cDLC|dUKM|dc#P|dXFF|dKSU|dOCoLC
049 KSUU
050 00 Z666.L5|b.C37 2003
062 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
```



UNIVERSITÀ
DEGLI STUDI
DI MILANO

 SESAR LAB

- Dublin, Ohio; quartier generale di OCLC – Online Computer Library Center
- Non profit org., proprietari del sistema decimale Dewey; FirstSearch e WorldCat, portali a diversi cataloghi
- Workshop, marzo 1995. Come descrivere, organizzare e dare accesso all'informazione attraverso la rete
- Mosaic, 1993: quindi il contesto web era poco conosciuto
- Definire uno standard per descrivere oggetti su scala web: a livello core cioè il minimo insieme di descrittori necessario per descrivere qualunque risorsa in rete

SCOPO DEL DUBLIN CORE



UNIVERSITÀ
DEGLI STUDI
DI MILANO

- Semplicità
 - Basso costo di adozione tanto da non avere scuse
 - Basso costo per implementarla in applicativi di ricerca che la usano
 - Solo 15 elementi di metadati
- Semantica condivisa
 - Generalista e di conseguenza con perdita di specializzazione ma incremento di usabilità
- Estendibile
 - Anche per ovviare alla perdita di specializzazione
- Internazionale
 - Con possibilità di traduzione in più lingue, ma per lo più sviluppata in lingua inglese



UNIVERSITÀ
DEGLI STUDI
DI MILANO

 SESAR LAB

GLI ELEMENTI DI DUBLIN CORE



UNIVERSITÀ
DEGLI STUDI
DI MILANO

- Contributor
- Publisher
- Coverage
- Relation
- Creator
- Rights
- Date
- Source
- Description
- Subject
- Format
- Title
- Identifier
- Type
- Language

<http://dublincore.org/documents/dces/>



UNIVERSITÀ
DEGLI STUDI
DI MILANO

 SESAR LAB

- Quando si descrive qualcosa si formula una frase. Uno schema di metadati controlla il tipo di frasi che si possono formulare e come
- Un vocabolario controllato per controllare i termini che si possono usare nella frase
 - Elemento
 - La categoria di frase che si può usare in uno schema di metadati per descrivere un oggetto
 - DC usa solo 15 elementi per formulare frasi ben formate
 - Valore
 - Le informazioni, i parametri che si passano all'elemento
 - DC spesso suggerisce di usare specifici vocabolari controllati
- Record
 - L'insieme delle coppie Elemento/Valore, cioè le frasi che si possono usare per descrivere le risorse



Title: Monna Lisa

Creator: Leonardo da Vinci

Subject: Lisa del Giocondo

Description: ritratto di donna

Date: c. 1503-1506

....

....

*Le coppie elemento/valore sono
ripetibili e in qualunque ordine*

- Includere solo informazioni rilevanti – semplificare (principio dumb-down)
- Subject non è rilevante per la descrizione di uno spartito?
Language non è rilevante per descrivere un dipinto?
Tralasciamo...
- Per ogni risorsa ci deve essere uno e un solo record che la descrive
- No allo stesso record per due risorse
- No a due record per la stessa risorsa



Title: Monna Lisa

Creator: Leonardo da Vinci

Subject: Lisa del Giocondo

Description: ritratto di donna

Publisher:

Contributor:

Date: c. 1503-1506

Type:

Format: olio su legno di pioppo

Identifier:

Source:

Language:

Relation:

Coverage:

Rights: Musée du Louvre


```
<html>

  <head>

    <title>Record del dipinto di Leonardo</title>

    <meta name="DC.creator" content="Leonardo, da Vinci,
    1452-1519">

    <meta name="DC.title" content="Monna Lisa">

    <meta name="DC.subject" content="Lisa del Giocondo">

    <meta name="DC.date" content="c. 1503-1506">

    <meta name="DC.format" content="olio su legno">

  </head>

  <body> contenuto della pagina web </body>

</html>
```

ESEMPIO



UNIVERSITÀ
DEGLI STUDI
DI MILANO



Title: Monna Lisa
Creator: Leonardo da Vinci
Subject: Lisa del Giocondo
Description: ritratto di donna
Date: c. 1503-1506
Format: olio su legno di pioppo
Rights: Musée du Louvre

Title: Monna Lisa
Creator: chi ha creato img digitale
Subject: Lisa del Giocondo
Description: ritratto di donna
Contributor: Leonardo da Vinci
Publisher: Musée du Louvre
Type: image
Date: 2014
Format: JPEG
Identifier: URI dell'immagine
Relation: URI ad altra risorsa
Rights: Musée du Louvre

**Un record per l'originale, un record
per l'immagine digitale e la relazione
tra i due. L'immagine digitale è una
rappresentazione dell'originale**



UNIVERSITÀ
DEGLI STUDI
DI MILANO

 SESAR LAB

- I metadati possono essere gestiti con vocabolari che variano per grado di specificazione
- **Vocabolari non controllati**
 - Definiti liberamente dagli utenti.
 - Flessibili ma possono causare inconsistenze terminologiche.
- **Vocabolari controllati**
 - Insiemi di termini predefiniti, spesso organizzati gerarchicamente.
 - Garantiscono coerenza e interoperabilità.
- **Ontologie**
 - Strutture formali che definiscono concetti, relazioni e proprietà.
 - Alto grado di specificazione, utilizzate in ambiti complessi come il web semantico.

- Ogni parola e ogni frase può essere usata per classificare
- Al contrario dei vocabolari controllati in cui le parole usate per descrivere sono limitate
- I tag delle reti sociali sono un esempio di vocabolario non controllato:
 - YouTube non limita nell'uso di parole chiave
 - 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
 - Flickr usa tag completamente liberi

► Metadati di YouTube

Video details

Title (required) 
Change everything here!

Description 
And here!

9/5000

Thumbnail

Select or upload a picture that shows what's in your video. A good thumbnail stands out and draws viewers' attention.

[Learn more](#)

- 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 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

- 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

IL VOCABOLARIO CONTROLLATO LCSH



UNIVERSITÀ
DEGLI STUDI
DI MILANO

BT Islands—Washington (D.C.)
Columbia Plateau
UF Channeled Scabland
Columbia and Snake River Plateau
Columbia River Plateau
Columbian 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.)
UF 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-

— Landmark of Christopher Columbus
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)
UF Catholic Cemetery (Columbus, Ohio)
BT Cemeteries—Ohio
Columbus Day
(*E120*)
BT Holidays
Columbus Quincentenary, 1992-1993
(*May Subd Geog*)
(*E119.2*)
UF 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

(*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
Berniniesque columns
Columns, Barley-sugar

Monday, September 16, 13




UNIVERSITÀ
DEGLI STUDI
DI MILANO

SESAR LAB

- Ad esempio i metadati di Amazon sono basati sullo standard [ONIX](#)

Product details

Publisher : The MIT Press; 4th edition (April 5, 2022)
Language : English
Hardcover : 1312 pages
ISBN-10 : 026204630X
ISBN-13 : 978-0262046305
Item Weight : 4.8 pounds
Dimensions : 8.25 x 2.01 x 9.31 inches
Best Sellers Rank: #11,808 in Books (See Top 100 in Books)
#1 in [Programming Algorithms](#)
#1 in [Computer Algorithms](#)
#3 in [Computer Programming Languages](#)
Customer Reviews: ★★★★★  335 ratings

book metadata on amazon.com

Product Details

ISBN: 9780262046305
Binding: Hardcover
Publication date: 04/05/2022
Publisher: MIT Press
Language: English
Pages: 1312
Height: 2.00IN
Width: 8.20IN
LCCN: 2021037280
Illustration: Yes
Author: Charles E. Leiserson
Author: Thomas H. Cormen
Author: Ronald L. Rivest

book metadata on powells.com

- ONIX (Online Information Exchange) è uno standard internazionale per lo scambio di informazioni di prodotti editoriali
 - Descrizione di libri, eBook e audiolibri
 - Gestione di cataloghi editoriali
 - Automazione del flusso di informazioni lungo la catena del valore editoriale. Es software [ONIXEDIT](#)
- Riduce errori manuali nella gestione dei metadati
- Supporta una vasta gamma di informazioni, inclusi dettagli sui prezzi e sulla disponibilità
- Gestisce metadati in più lingue e valuta per il commercio globale

- Si tratta di un vocabolario XML, [esempio1](#) [esempio2](#)
- La struttura di un messaggio ONIX è organizzata secondo le seguenti componenti:
- Header: informazioni sul mittente e il destinatario del messaggio.
- Product: dettagli sul prodotto editoriale
 - DescriptiveDetail: titolo, autore, ISBN, lingua, numero di pagine
 - CollateralDetail: immagini e materiali promozionali
 - PublishingDetail: editore e data di pubblicazione
 - ProductSupply: prezzi, disponibilità e condizioni di vendita

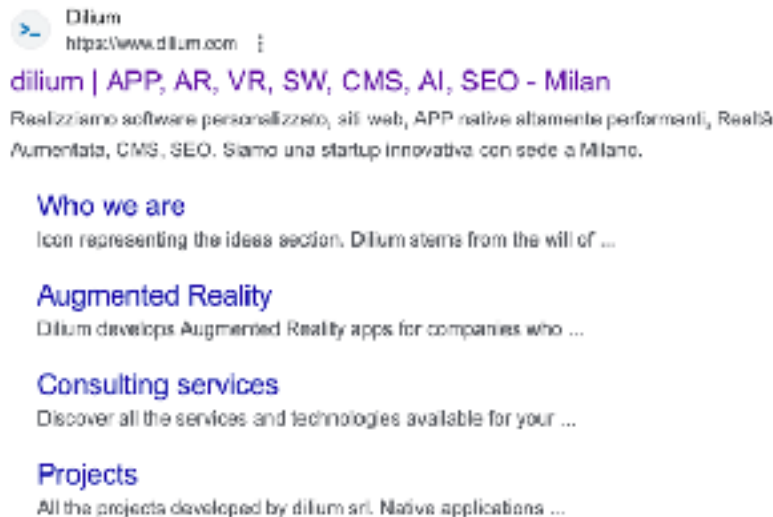
- Diversi motori di ricerca come Google, Bing, Yahoo sono in grado di interpretare le descrizioni di alcune tipologie di contenuti se seguono lo schema definito da schema.org
- Queste descrizioni possono essere associate ad una pagina HTML usando diversi formati, il più diffuso sta diventando JSON-LD

<https://developers.google.com/search/docs/guides/intro-structured-data>

<https://json-ld.org/playground/>

- In questo modo i motori di ricerca possono costruire strutture dati più significative come Breadcrumbs, Sitelinks Search Box, Rich Card, Rich Snippets o (Google Knowledge Graph)

- Ad esempio le informazioni relative alle diverse aree di un sito sono costruite grazie ai metadati



- È possibile verificare la struttura dei metadati utilizzando [Rich Results Test](#) di Google o lo [Schema Markup Validator](#) di schema.org
- È possibile verificare l'impatto sulle poiché di SEO utilizzando Google [Search Console](#)

- Nell'ingegneria della conoscenza, la rappresentazione formale di un insieme di concetti in un determinato ambito
- Un insieme di termini di relazioni più complesse di quelle nei vocabolari controllati
 - Sussunzione: super- o sub-class
 - Meronimia: part-of
 - Istanziamento: type-of
 - Relazioni: simmetria, transitivi, riflessività
- 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>

- Le ontologie sono tesauro con relazioni formali: esprimono sottoinsiemi della logica del primo ordine
- Nelle ontologie entità, relazioni e inferenze consentono di dedurre fatti non esplicitamente asseriti
- Le ontologie possono quindi essere usate in algoritmi, alcuni linguaggi
 - OWL <https://www.w3.org/TR/owl-ref/>
 - SWRL <https://www.w3.org/Submission/SWRL/>
- Alcuni formati di serializzazione:
 - RDF <https://www.w3.org/TR/rdf11-concepts/>
 - Turtle <https://www.w3.org/TR/turtle/>

ESEMPIO DI CONCETTUALIZZAZIONE



UNIVERSITÀ
DEGLI STUDI
DI MILANO



ESEMPIO DI CONCETTUALIZZAZIONE



UNIVERSITÀ
DEGLI STUDI
DI MILANO



Fabio Fazio



Luciana Littizetto

ESEMPIO DI CONCETTUALIZZAZIONE



UNIVERSITÀ
DEGLI STUDI
DI MILANO

Maschio



Fabio Fazio

Femmina



Luciana Littizetto

ESEMPIO DI CONCETTUALIZZAZIONE



UNIVERSITÀ
DEGLI STUDI
DI MILANO

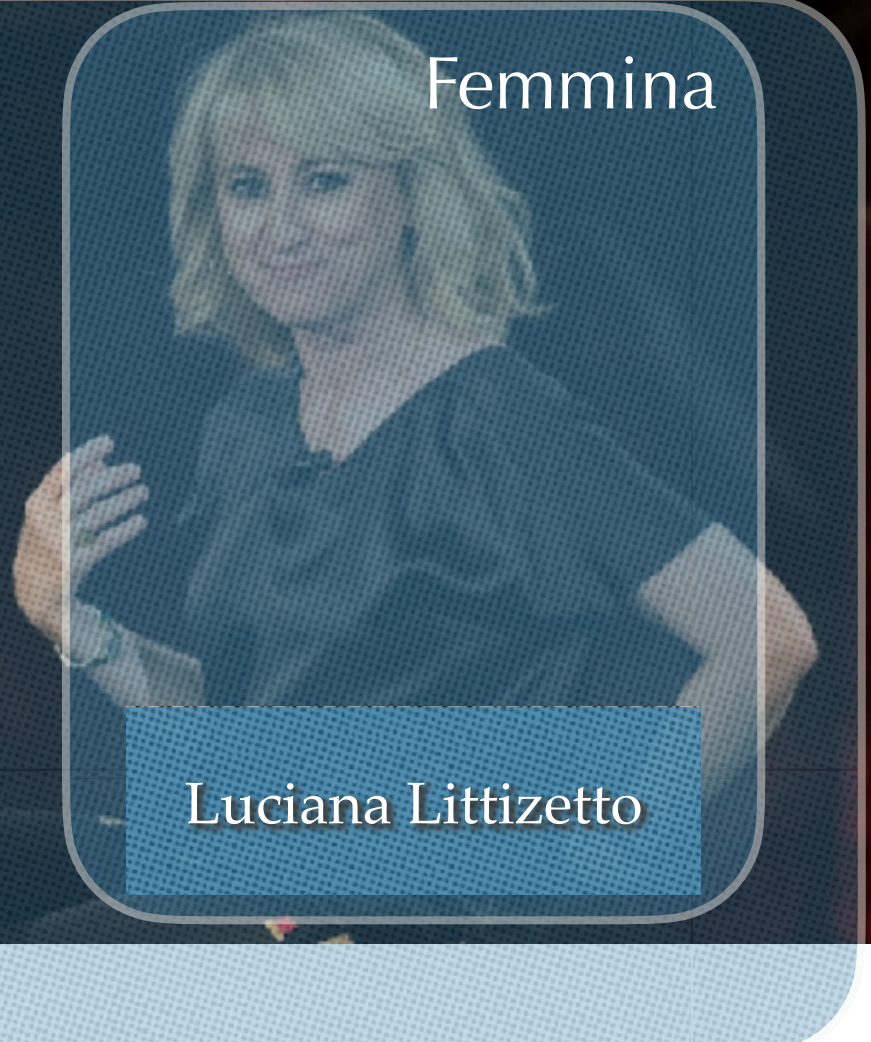
Persona

Maschio



Fabio Fazio

Femmina



Luciana Littizetto

ESEMPIO DI CONCETTUALIZZAZIONE



UNIVERSITÀ
DEGLI STUDI
DI MILANO

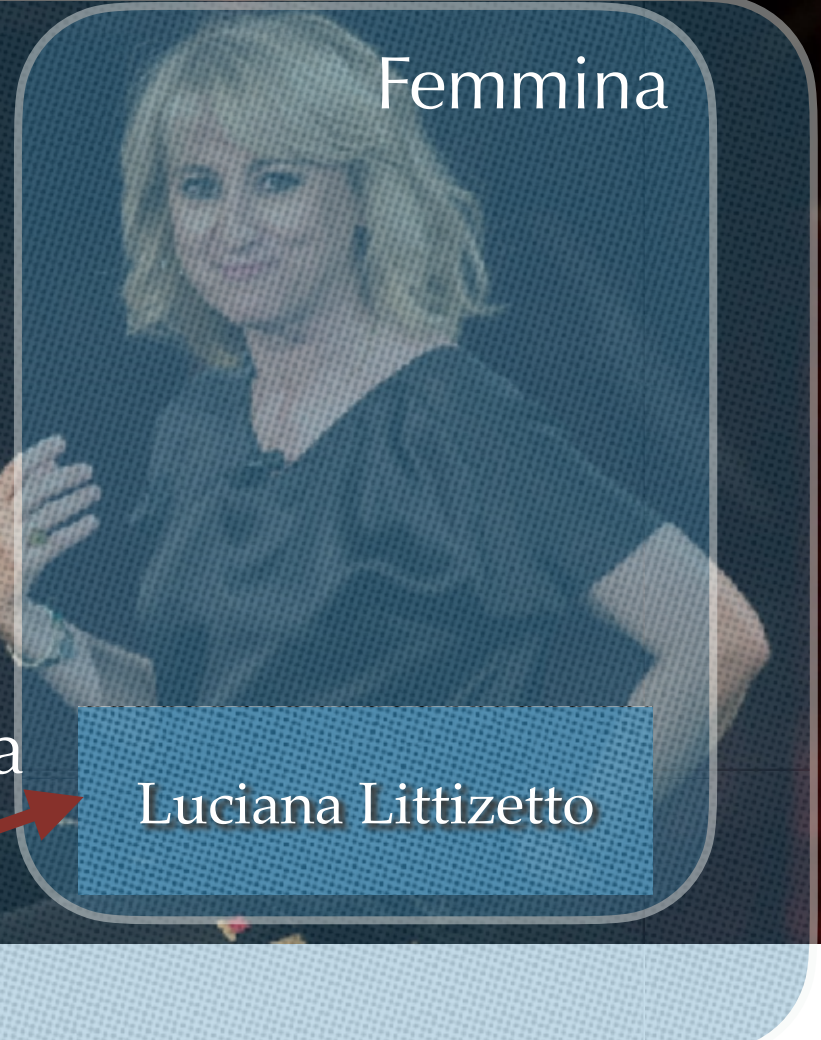
Persona

Maschio



Fabio Fazio

Femmina



Luciana Littizetto

collabora



ESEMPIO DI CONCETTUALIZZAZIONE



UNIVERSITÀ
DEGLI STUDI
DI MILANO



- Identificare un oggetto:
 - https://it.wikipedia.org/wiki/Fabio_Fazio
 - https://it.wikipedia.org/wiki/Luciana_Littizzetto
- Si tratta di associare un oggetto o risorsa ad un identificatore che potrà essere utilizzato in modo univoco all'interno del vocabolario

- Si tratta di associare un oggetto o una risorsa ad un identificatore che potrà essere utilizzato in modo univoco all'interno del vocabolario
 - Molti sono i criteri che si possono usare per decidere cosa considerare individuo:
 - un oggetto indivisibile
 - un'unità che non ammette riduzioni nel sistema di riferimento
 - un oggetto che ha una condizione di identità
 - un oggetto che è associato ad un identificatore
- Solitamente si considera individuo qualche cosa che permane (che identifichiamo all'interno di uno sfondo mutevole, o irrilevante, o composto da altri oggetti identificati)

- Classificare un oggetto:
 - https://it.wikipedia.org/wiki/Fabio_Fazio **type** <https://it.wikipedia.org/wiki/Maschio>
 - https://it.wikipedia.org/wiki/Luciana_Littizzetto **type** <https://it.wikipedia.org/wiki/Femmina>
- Si tratta di affermare che un oggetto appartiene ad una classe di oggetti: ovvero **condivide con altri oggetti uno stesso insieme di proprietà**

- Si tratta di affermare che un oggetto appartiene ad una classe di oggetti: ovvero condivide con altri oggetti uno stesso insieme di proprietà
 - In una **buona classificazione** l'insieme di proprietà definito dalle classi dovrebbe essere:
 - reciprocamente esclusivo
 - congiuntamente esaustivo
 - pertinente
 - In realtà quasi tutti i vocabolari includono diverse criteri di classificazione, diverse dimensioni o punti di osservazioni, quindi ammettono intersezioni tra le classi che appartengono a dimensioni diverse

- Nel flussi di gestione documentale che progetteremo è utile considerare tre categorie di metadati
- **Gestione documentale**
 - Informazioni che possiamo includere nei documenti di destinazione
 - Informazioni per parametrizzare le trasformazioni documentali
- **Archiviazione**
 - Informazioni descrittive del contenuto, es. Dublin Core
 - Informazioni sui diritti d'autore e licenze d'uso
- **Distribuzione**
 - Informazioni richieste dai canali di distribuzione, es. ONIX e Schema.org

- Ad esempio con **YAML** potremmo descrivere così un nostro documento

gestione_documentale:

title: "Guida Completa alla Gestione Documentale"

author:

- "Paolo Ceravolo"
- "Collaboratore A"

date: "2024-11-17"

abstract: "Le migliori pratiche di gestione documentale."

lang: "it"

bibliography: "references.bib"

csl: "apa.csl"

document_class: "report"

output_formats:

- pdf
- html
- docx

METADATI NELLA GESTIONE DOCUMENTALE



UNIVERSITÀ
DEGLI STUDI
DI MILANO

- Ad esempio con **YAML** potremmo descrivere così un nostro documento

archiviazione:

dublin_core:

title: "Guida Completa alla Gestione Documentale"

creator:

- "Paolo Ceravolo"
- "Collaboratore A"

subject:

- "Gestione documentale"
- "Archivistica"

description: "Le migliori pratiche e tecniche per gestire e archiviare i documenti."

publisher: "Università degli Studi di Milano"

date: "2024-11-17"

type: "Manuale"

format: "application/pdf"

identifier: "ISBN-978-88-555-1234-5"

language: "it"

coverage: "Italia"

rights: "Creative Commons Attribution 4.0 International License"



UNIVERSITÀ
DEGLI STUDI
DI MILANO



METADATI NELLA GESTIONE DOCUMENTALE



UNIVERSITÀ
DEGLI STUDI
DI MILANO

- Ad esempio con **YAML** potremmo descrivere così un nostro documento

distribuzione:

onix:

title: "Guida Completa alla Gestione Documentale"

contributor:

primary_author: "Paolo Ceravolo"

other_authors:

- "Collaboratore A"

publisher: "Università degli Studi di Milano"

publication_date: "2024-11-17"

isbn: "978-88-555-1234-5"

audience: "Professionisti, Accademici"

format:

type: "Digital"

file_format: "PDF"

price:

amount: "19.99"

currency: "EUR"



UNIVERSITÀ
DEGLI STUDI
DI MILANO



METADATI NELLA GESTIONE DOCUMENTALE



UNIVERSITÀ
DEGLI STUDI
DI MILANO

- Ad esempio con **YAML** potremmo descrivere così un nostro documento

distribuzione:

schema_org:

```
"@context": "https://schema.org"
```

```
"@type": "Book"
```

```
name: "Guida Completa alla Gestione Documentale"
```

author:

```
- "@type": "Person"
```

```
  name: "Paolo Ceravolo"
```

```
- "@type": "Person"
```

```
  name: "Collaboratore A"
```

```
datePublished: "2024-11-17"
```

publisher:

```
"@type": "Organization"
```

```
name: "Università degli Studi di Milano"
```

```
isbn: "978-88-555-1234-5"
```

genre:

```
- "Gestione documentale"
```

```
- "Archivistica"
```

```
language: "it"
```

```
inLanguage: "Italian"
```

```
format: "PDF"
```



UNIVERSITÀ
DEGLI STUDI
DI MILANO

 **SESAR LAB**

