

## Progettazione e Produzione di Contenuti Digitali

Prof.ssa E. Gentile  
a.a. 2011-2012

---

---

---

---

---

---

---

---

## I Metadati

- I metadati possono essere letteralmente definiti come “dati sui dati”, ma si interpreta generalmente il termine nel senso di dati strutturati che possono supportare un’ampia gamma di operazioni sulle risorse digitali.

Prof.ssa E. Gentile

2

---

---

---

---

---

---

---

---

## Operazioni sui Metadati

- Le operazioni supportate dai metadati possono comprendere:
  - individuazione
  - recupero
  - gestione e controllo
  - conservazione a lungo termine delle risorse

Prof.ssa E. Gentile

3

---

---

---

---

---

---

---

---

## Categorie funzionali dei Metadati

- **Descrittivi**
  - per l'identificazione e il recupero degli oggetti digitali; sono costituiti da descrizioni normalizzate dei documenti fonte (o dei documenti digitali nativi), risiedono generalmente nelle basi dati dei sistemi di *information retrieval* all'esterno dei depositi degli oggetti digitali e sono collegati a questi ultimi tramite appositi *link*.
- **Amministrativi e gestionali**
  - per le svariate operazioni di gestione degli oggetti digitali all'interno dell'archivio.
- **Strutturali**
  - per descrivere la struttura interna dei documenti (ad esempio, introduzione, capitoli, indice di un libro) e gestire le relazioni fra le varie parti componenti gli oggetti digitali.

---

---

---

---

---

---

---

---

## I Metadati

- I Metadati descrivono:
  - l'oggetto fisico digitalizzato
  - gli oggetti digitali creati durante il processo di digitalizzazione e memorizzati come master digitali
  - gli oggetti digitali derivati da tali master per la diffusione in rete
  - le nuove risorse create impiegando tali oggetti digitali
  - le collezioni di ogni sorta di oggetti sopra citati

---

---

---

---

---

---

---

---

## Metadati Descrittivi

- I metadati *descrittivi* servono per il reperimento e l'interpretazione dell'oggetto digitale.
- Il DCMES (*Dublin Core Metadata Element Set*) è uno schema di metadati descrittivi molto semplice, prodotto di un'iniziativa interdisciplinare e progettato per supportare il reperimento di risorse attraverso una vasta gamma di domini.

---

---

---

---

---

---

---

---

## DCMES

### *Dublin Core Metadata Element Set*

- |                                     |                          |
|-------------------------------------|--------------------------|
| 1. Titolo                           | 8. Tipo                  |
| 2. Creatore                         | 9. Formato               |
| 3. Soggetto                         | 10. Identificatore       |
| 4. Descrizione                      | 11. Fonte                |
| 5. Editore                          | 12. Lingua               |
| 6. Autore di contributo subordinato | 13. Relazione            |
| 7. Data                             | 14. Copertura            |
|                                     | 15. Gestione dei diritti |

Prof.ssa E. Gentile

7

---

---

---

---

---

---

---

---

---

---

## Titolo

- **Nome:** Titolo
- **Identificatore:** Title
- **Definizione:** Un nome dato alla risorsa
- **Commento:** In particolare, un Title sarà un nome con il quale la risorsa è formalmente conosciuta.

Prof.ssa E. Gentile

8

---

---

---

---

---

---

---

---

---

---

## Creatore

- **Nome:** Creatore
- **Identificatore:** Creator
- **Definizione:** Un'entità che ha la responsabilità principale della produzione del contenuto della risorsa
- **Commento:** Esempi di Creator includono una persona, un'organizzazione o un servizio. In particolare, il nome di un Creator dovrebbe essere usato per indicare l'entità

Prof.ssa E. Gentile

9

---

---

---

---

---

---

---

---

---

---

## Soggetto

- **Nome:** Soggetto e Parole chiave
- **Identificatore:** Subject
- **Definizione:** L'argomento della risorsa.
- **Commento:** In particolare un Subject può essere espresso da parole chiave o frasi chiave o codici di classificazione che descrivono l'argomento della risorsa. Si consiglia di scegliere un valore da un vocabolario controllato o da uno schema di classificazione formale.

Prof.ssa E. Gentile

10

---

---

---

---

---

---

---

---

## Descrizione

- **Nome:** Descrizione
- **Identificatore:** Description
- **Definizione:** Una spiegazione del contenuto della risorsa
- **Commento:** Una descrizione può includere ma non è limitata solo a:
  - un riassunto analitico
  - un indice
  - un riferimento al contenuto di una rappresentazione grafica
  - un testo libero del contenuto.

Prof.ssa E. Gentile

11

---

---

---

---

---

---

---

---

## Editore

- **Nome:** Editore
- **Identificatore:** Publisher
- **Definizione:** Un'entità responsabile della produzione della risorsa, disponibile nella sua forma presente
- **Commento:** Esempi di un Publisher includono una persona, un'organizzazione o un servizio. Normalmente il nome di un Publisher è usato per indicare l'entità stessa

Prof.ssa E. Gentile

12

---

---

---

---

---

---

---

---

## Autore di contributo subordinato

- **Nome:** Autore di contributo subordinato
- **Identificatore:** Contributor
- **Definizione:** Un'entità responsabile della produzione di un contributo al contenuto della risorsa
- **Commento:** Esempi di un Contributor includono una persona, un'organizzazione o un servizio. Normalmente il nome di un Contributor è usato per indicare l'entità stessa

---

---

---

---

---

---

---

---

## Data

- **Nome:** Data
- **Identificatore:** Date
- **Definizione:** Una data associata a un evento del ciclo di vita della risorsa
- **Commento:** Normalmente Date è associata alla creazione o alla disponibilità della risorsa. Si consiglia di usare un numero di 8 caratteri nella forma YYYY-MM-DD come definita in W3C, un profilo dello standard ISO 8601.
  - In questo schema l'elemento data 1994-11-05 corrisponde al 5 novembre 1994. Sono possibili molti altri schemi, ma se vengono utilizzati, devono poter essere identificati in modo univoco

---

---

---

---

---

---

---

---

## Tipo

- **Nome:** Tipo di risorsa
- **Identificatore:** Type
- **Definizione:** La natura o il genere del contenuto della risorsa
- **Commento:** Tipo include termini che descrivono categorie generali, funzioni, generi o livelli di aggregazione per contenuto. Si raccomanda di selezionare un valore da un vocabolario controllato. Per descrivere la manifestazione fisica o digitale della risorsa, si usa l'elemento Format

---

---

---

---

---

---

---

---

## Formato

- **Nome:** Formato
- **Identificatore:** Format
- **Definizione:** La manifestazione fisica o digitale della risorsa
- **Commento:** Normalmente Formato può includere il tipo di supporto o le dimensioni della risorsa. Formato può essere usato per determinare il software, l'hardware o altro apparato necessario alla visualizzazione o all'elaborazione della risorsa. Esempi di dimensioni includono grandezza e durata. Si raccomanda di selezionare un valore da un vocabolario controllato (ad esempio la lista di Internet Media Types (MIME) che definisce i formati dei supporti elettronici)

---

---

---

---

---

---

---

---

## Identificatore

- **Nome:** Identificatore della risorsa
- **Identificatore:** Identifier
- **Definizione:** Un riferimento univoco alla risorsa nell'ambito di un dato contesto
- **Commento:** Si raccomanda di identificare la risorsa per mezzo di una sequenza di caratteri alfabetici o numerici secondo un sistema di identificazione formalmente definito.
  - Esempi di tali sistemi di identificazione includono l'Uniform Resource Identifier (URL) (incluso l'Uniform Resource Locator (URN)), il Digital Object Identifier (DOI) e l'International Standard Book Number (ISBN)

---

---

---

---

---

---

---

---

## Fonte

- **Nome:** Fonte
- **Identificatore:** Source
- **Definizione:** Un riferimento a una risorsa dalla quale è derivata la risorsa in oggetto
- **Commento:** La risorsa può essere derivata dalla risorsa Source in tutto o in parte. Si raccomanda di identificare la risorsa per mezzo di una sequenza di caratteri alfabetici o numerici secondo un sistema di identificazione formalmente definito

---

---

---

---

---

---

---

---

## Lingua

- **Nome:** Lingua
- **Identificatore:** Language
- **Definizione:** La lingua del contenuto intellettuale della risorsa
- **Commento:** Per i valori dell'elemento Language si raccomanda di utilizzare quanto definito dal RFC 1766 che include un codice di lingua su due caratteri (derivato dallo standard ISO 639, seguito opzionalmente da un codice di paese su due caratteri (derivato dallo standard ISO 3166).
  - Ad esempio, "en" per l'inglese, "fr" per il francese, "it" per l'italiano o "en-uk" per l'inglese usato nel Regno Unito

---

---

---

---

---

---

---

---

## Relazione

- **Nome:** Relazione
- **Identificatore:** Relation
- **Definizione:** Un riferimento alla risorsa correlata
- **Commento:** Si raccomanda di far riferimento alla risorsa per mezzo di una sequenza di caratteri alfabetici o numerici in conformità ad un sistema di identificazione formalmente definito

---

---

---

---

---

---

---

---

## Copertura

- **Nome:** Copertura
- **Identificatore:** Coverage
- **Definizione:** L'estensione o scopo del contenuto della risorsa
- **Commento:** Normalmente Coverage include
  - la localizzazione spaziale (il nome di un luogo o le coordinate geografiche),
  - il periodo temporale (l'indicazione di un periodo, una data o un range di date)
  - una giurisdizione (ad esempio il nome di un'entità amministrativa).Si raccomanda di selezionare un valore da un vocabolario controllato (es il Thesaurus of Geographic Names) e, se possibile, di utilizzare i nomi di luogo o i periodi di tempo piuttosto che identificatori numerici come serie di coordinate o range di date

---

---

---

---

---

---

---

---

## Gestione dei diritti

- **Nome:** Gestione dei diritti
- **Identificatore:** Rights
- **Definizione:** Informazione sui diritti esercitati sulla risorsa
- **Commento:** Normalmente un elemento Rights contiene un'indicazione di gestione dei diritti sulla risorsa, o un riferimento ad un servizio che fornisce questa informazione. L'informazione sui diritti spesso comprende i diritti di proprietà intellettuale Intellectual Property Rights (IPR), Copyright e vari diritti di proprietà. Se l'elemento Rights è assente, non si può fare alcuna ipotesi sullo stato di questi o altri diritti in riferimento alla risorsa

---

---

---

---

---

---

---

---

## Metadati Amministrativi e Gestionali

- I metadati amministrativi vengono usati per la gestione dell'oggetto digitale e per offrire maggiori informazioni sulla sua creazione e su ogni restrizione d'uso.

---

---

---

---

---

---

---

---

## Metadati Amministrativi e Gestionali

- Metadati tecnici
- Metadati per la conservazione
  - Metadati relativi alla fonte
  - Metadati relativi alla provenienza
- Metadati per la gestione dei diritti
- Metadati per la conservazione a lungo termine

---

---

---

---

---

---

---

---



## File MAG

- I Metadati Amministrativi Gestionali (MAG) rappresentano lo standard ICCU per l'amministrazione dei dati relativi alle risorse digitali.
- I file MAG sono generati automaticamente a partire dalle informazioni catalografiche e di digitalizzazione.

---

---

---

---

---

---

---

---

## Sezioni del MAG

- <gen>** informazioni generali sul progetto e sul tipo di digitalizzazione
- <bib>** metadati descrittivi sull'oggetto digitalizzato
- <stru>** dati sulla struttura logica dell'oggetto digitalizzato
- <img>** metadati specifici delle immagini digitalizzate
- <ocr>** metadati relativi al riconoscimento ottico del testo
- <doc>** metadati relativi ai file di testo correlati
- <audio>** metadati relativi a un file contenente una traccia audio
- <video>** metadati relativi a uno stream video
- <dis>** dati sulla disseminazione degli oggetti digitali e la loro fruibilità

---

---

---

---

---

---

---

---

## Esempio di sezione <bib>

```
<bib level="m">
<dc:identifier>info:sbn/CFI0404365</dc:identifier>
<dc:identifier>info:bnl/99-2276</dc:identifier>
<dc:title>
Canzoni / Giacomo Leopardi; versione in prosa, note e
postfazione di Marco Santagata
</dc:title>
<dc:creator>Leopardi, Giacomo</dc:creator>
<dc:publisher>Milano: A. Mondadori, 1998.</dc:publisher>
<holdings ID="H1">
<library>Biblioteca Nazionale Centrale</library>
<inventory_number>CF0053604765</inventory_number>
<shelfmark>GEN.A05.5826</shelfmark>
</holdings>
</bib>
```

---

---

---

---

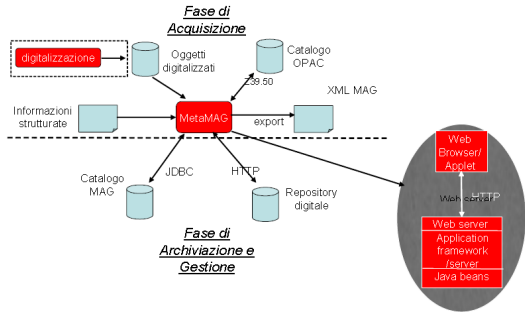
---

---

---

---

## Fasi della creazione del MAG



Prof.ssa E. Gentile

28

---

---

---

---

---

---

---

---

---

---

## Tipologia di Metadati

- **metadati tecnici**
  - descrivono le caratteristiche tecniche della risorsa digitale
- **metadati per la conservazione** comprendono:
  - metadati relativi *alla fonte*
    - descrivono l'oggetto dal quale è derivata la risorsa digitale
  - metadati relativi *alla provenienza*
    - descrivono la storia delle operazioni effettuate su un oggetto digitale fin dalla sua creazione
- **metadati per la gestione dei diritti**
  - descrivono i diritti d'autore e di riproduzione, le restrizioni e le licenze che vincolano l'uso della risorsa
- **metadati per la conservazione a lungo termine**
  - necessari a conservare adeguatamente per un periodo di tempo indefinito il contenuto digitale

Prof.ssa E. Gentile

29

---

---

---

---

---

---

---

---

---

---

## Metadati Strutturali

- I metadati strutturali descrivono le relazioni logiche o fisiche che collegano le parti di un oggetto composto.
- Un libro "fisico", ad esempio, consiste in una sequenza di pagine. Un documento archivistico può consistere in una sequenza di carte di un fascicolo inserito in una busta, a sua volta parte di una serie inserita in un fondo particolare.
- Il processo di digitalizzazione può generare un certo numero di risorse digitali distinte, ad esempio una immagine per ogni pagina, ma il fatto che queste risorse formino una sequenza e che quella sequenza costituisca un oggetto composto o faccia parte di una struttura complessa multilivellare è evidentemente essenziale per il loro uso e la loro interpretazione.

Prof.ssa E. Gentile

30

---

---

---

---

---

---

---

---

---

---

## Collezione Digitale

- Una risorsa digitale non è creata isolatamente, ma come parte di una collezione digitale, e dovrebbe essere presa in considerazione nel contesto di quella collezione e del suo sviluppo.

Prof.ssa E. Gentile

31

---

---

---

---

---

---

---

---

## Standard Terminologici

- Perché la trasmissione delle informazioni veicolate dai metadati sia efficace non è sufficiente la semplice comprensione e condivisione dello schema di metadati adottato e degli elementi che lo compongono.
- Occorre anche un accordo sui descrittori da usare come valori di quegli elementi di metadati, o attraverso l'adozione di terminologie comuni o adottando terminologie differenti in cui le equivalenze e le relazioni fra i termini siano chiaramente definite.

Prof.ssa E. Gentile

32

---

---

---

---

---

---

---

---

## Pubblicazione

- Il master digitale deve prevedere:
- Meccanismi di scalabilità
  - come la riduzione della qualità e la generazione di file di dimensione adeguate per il trasferimento su rete
- Metodi di trasmissione progressiva
  - consente una fruizione graduale e controllabile a seconda di parametri collegati al profilo dell'utente
- I file possono essere resi disponibili sia per il *download* che per lo *streaming*.

Prof.ssa E. Gentile

33

---

---

---

---

---

---

---

---

## Pubblicazione di Immagini Fisse

- La stessa immagine deve poter essere offerta in diversi formati
- Francobollo
  - Icone (Dim max 100-200 pixel)
- Cartolina
  - Sono riduzioni di immagini più grandi (Dim max 600 pixel)
- Full-screen
  - (Dim max 1024-1280 pixel)

Prof.ssa E. Gentile

34

---

---

---

---

---

---

---

---

## Download di un file di ~150 Kbyte

Alla velocità di (bps)	Tempo impiegato (s)
56 K	~28
128 K	~12
384 K	~4
512 K	~3
1 M	~1,4

Prof.ssa E. Gentile

35

---

---

---

---

---

---

---

---

## Ontologie Web

- Un'*ontologia* è una definizione esplicita, condivisa e socialmente accettata di una porzione di realtà (dominio di conoscenza) per mezzo di un modello concettuale.
- Mentre un *thesaurus* definisce e mette in relazioni termini, un'*ontologia* ha l'obiettivo di definire ed esplicitare relazioni tra concetti

Prof.ssa E. Gentile

36

---

---

---

---

---

---

---

---

## Componenti di un'ontologia

- Un insieme di concetti (entità, attributi, processi, ...) riguardanti un dato dominio
- Le definizioni di tali concetti
- Le relazioni che interconnettono le entità o classi all'interno di un dato dominio

Prof.ssa E. Gentile

37

---

---

---

---

---

---

---

---

## Costruzione di un'ontologia

- Esaminare il vocabolario utilizzato per descrivere oggetti e processi caratteristici del dominio
- Sviluppare definizioni rigorose dei termini principali di tale vocabolario
- Caratterizzare le connessioni logiche esistenti fra tali termini

Prof.ssa E. Gentile

38

---

---

---

---

---

---

---

---

## Linguaggio RDF

- RDF – Resource Description Framework
- Raccomandato da W3C è usato per rappresentare l'informazione e scambiare la conoscenza nel Web
- È orientato alla rappresentazione dei metadati (es. titolo, autore, data di modifica...)
- La struttura è basata su "risorsa", "attributo" e "valore" e permette di definire la semantica dei tag XML

Prof.ssa E. Gentile

39

---

---

---

---

---

---

---

---

## Diritti di proprietà intellettuale

- Proprietà letteraria e scientifica
  - Diritto d'autore, copyright
- Proprietà industriale
  - Brevetti, marchi, disegni e modelli industriali
- Origini differenti degli oggetti
  - Materiali pubblicati
  - Materiali di produzione interna
  - Opere commissionate da istituzioni
  - Doni, lasciti e prestiti

Prof.ssa E. Gentile

40

---

---

---

---

---

---

---

---

## Dati sui diritti

- Identificazione della risorsa digitale
- Nome della persona o organizzazione che concede la licenza
- Lo specifico diritto (o diritti) che viene concesso (es. se è permessa la modifica) e ogni specifica esclusione
- Il periodo di tempo per cui la licenza è concessa
- La comunità di utenti o gruppi che può usare la risorsa
- Ogni obbligo (inclusi i pagamenti) in cui l'utente della risorsa può incorrere

Prof.ssa E. Gentile

41

---

---

---

---

---

---

---

---

## Dublin Core

- Ogni elemento del Dublin Core è definito utilizzando un insieme di dieci attributi dello standard ISO/IEC 11179 per la descrizione dei dati.

1. **Nome** – Etichetta assegnata al dato
2. **Identificatore** – Identificativo univoco assegnato al dato
3. **Versione** – Versione del dato
4. **Registrazione di autorità** – Entità autorizzata a registrare il dato
5. **Lingua** – Lingua nella quale il dato è indicato
6. **Definizione** – Indicazione che rappresenta chiaramente il concetto e la natura essenziale del dato
7. **Obbligatorietà** – Indica se il dato è richiesto sempre o solo in alcuni casi (contiene un valore)
8. **Tipo di dato** – Indica la tipologia del dato che può essere rappresentato nel valore del dato stesso
9. **Occorrenza massima** – Indica un limite alla ripetibilità del dato
10. **Commento** – Un'osservazione che concerne l'applicazione del dato

Prof.ssa E. Gentile

42

---

---

---

---

---

---

---

---