

2. METADATI

Il permesso di fare copie digitali o fisiche di tutto o parte di questo lavoro per uso di ricerca o didattico è acconsentito senza corrispettivo in danaro, mentre per altri usi o per inviare a server, ridistribuire a liste di discussione o diffondere ulteriormente è necessario il permesso da parte dell'autore.

L'utilizzo per scopi di profitto non è consentito senza il permesso dell'autore.

Gli eventuali lavori derivanti dallo stesso dovranno contenere opportuna citazione.

2.1 INTRODUZIONE

Questo capitolo, ha per oggetto i metadati, ovvero i “dati sui dati”, che tanta importanza rivestono nell'ambito degli open archive e dell'iniziativa OAI, soprattutto per quel che concerne l'interoperabilità tra i vari archivi.

In particolare, oltre a precisare cosa sono i metadati e i loro possibili utilizzi, il capitolo definisce i formati di metadati e le caratteristiche che devono avere per essere conformi all'iniziativa OAI, la cui scelta è ricaduta sul Dublin Core per varie ragioni.

Infine si accenna brevemente ai due formati standard di metadati, Dublin Core e MARC 21, che verranno trattati più approfonditamente nei due capitoli successivi.

Per la realizzazione di tale capitolo, importante è stato il contributo di un lavoro della dott.ssa De Robbio riguardante i metadati all'interno del circuito della comunicazione scientifica [B3].

2.2 METADATI

I *metadati* venivano utilizzati già in tempi remoti quando i primi bibliotecari catalogavano i manoscritti redigendo degli elenchi di *item* appositi, ovvero di identificatori delle risorse cartacee le quali venivano poste sugli scaffali della biblioteca.

Il termine “meta” deriva dal greco ed ha vari significati: accanto, con, dopo, successivo. In inglese e nelle lingue neolatine tale termine viene impiegato comunemente per riferire qualcosa di trascendentale o metafisico.

Il significato attuale del termine metadati è quello di “dati a proposito di altri dati”. I bibliotecari tradizionalmente hanno inserito tali informazioni nei loro cataloghi e con l’avvento di Internet il termine si è riferito comunemente ad informazioni che descrivono le risorse del Web.

I metadati vengono utilizzati per descrivere le risorse, digitali o fisiche, in maniera strutturata e standardizzata per consentire la scoperta, il recupero e l’uso di tali risorse.

Si definisce formato di metadati un insieme di elementi atti a descrivere una determinata risorsa.

Esistono molti formati di metadati che differiscono per le seguenti caratteristiche:

- *Tipologia* – le diverse funzionalità dei metadati:
 - *Amministrativa* – per la gestione ed amministrazione delle risorse informative.
 - *Descrittiva* – per la descrizione o identificazione delle risorse informative.
 - *sulla Conservazione* – per la gestione della conservazione delle risorse informative.
 - *Tecnica* – per il funzionamento di un sistema e il comportamento dei metadati.
 - *di Utilizzo* – per il livello e il tipo di utilizzo delle risorse informative da parte dell’utente.

- *Ricchezza del formato.*
- *Specificità* per un particolare dominio di applicazione.

Non vi è un unico formato di metadati che vada bene per tutte le necessità, quindi nella scelta bisognerà analizzare, oltre i fattori precedentemente elencati, le seguenti problematiche:

- *Gestione locale* – la necessità di gestire il repository tenendo conto della crescita delle risorse nel tempo.
- *Riutilizzo dei metadati* – le informazioni associate ad una risorsa, o metadati, possono essere riutilizzate per scopi supplementari a quelli individuati inizialmente. Scegliere quindi un formato di metadati che garantisca l'*estendibilità*, permette di riutilizzare gli elementi precedenti adattando il grado di descrivibilità della risorsa alle nuove necessità emergenti.

I metadati permettono la descrizione precisa delle risorse in piccoli pacchetti informativi chiamati *record di metadati*. Per esempio, un catalogo bibliografico contiene un insieme di tali record ognuno dei quali composto da elementi: autore, titolo, data di creazione o pubblicazione, soggetto, codice identificativo per la posizione sullo scaffale, ecc., che descrivono libri o altri item bibliografici.

Il collegamento tra il metadato e la risorsa da esso descritta può essere così realizzato:

- Il metadato può essere separato dalla risorsa, come nel caso del catalogo bibliografico.
- Il metadato può essere incorporato nella risorsa stessa, come ad esempio i dati CIP (Cataloging In Publication) stampati sulla facciata della pagina titolo di un libro o l'header TEI (Text Encoding Initiative) in un testo

elettronico.

La maggior parte degli standard di metadati attualmente in uso non prescrivono l'utilizzo di un tipo di collegamento rispetto all'altro, lasciando facoltà di scelta in funzione dell'implementazione.

Il collegamento tra metadati descrittivi e oggetti di rete può migliorare notevolmente le capacità di scoperta delle risorse, permettendo:

- Ricerche basate su un determinato campo informativo, quale autore, titolo, ...
- Possibilità di capire, senza dover accedere alla risorsa, se essa soddisfa le necessità dell'utente basando la ricerca su elementi maggiormente descrittivi: formato, durata, tipo,...
- Indicizzazione di oggetti non testuali.
- Possibilità di accesso ad un contenuto sostitutivo distinto dall'accesso al contenuto della risorsa stessa.

Lo sviluppo di Internet e del Web ha portato al sovraccarico di informazioni dovuto alla grande quantità di dati digitali disponibili online, constatabile ad esempio effettuando la ricerca di una risorsa attraverso i servizi disponibili in rete, laddove si ottengono centinaia, o addirittura migliaia, di risultati, non avendo modo di poter restringere adeguatamente il campo della ricerca. In quest'ottica una possibile funzione dei metadati può essere proprio quella di affinare la ricerca utilizzando le informazioni descrittive in essi contenute.

I metadati non vengono solamente utilizzati per scopi di descrizione e/o scoperta di risorse, ma possono essere anche usati per:

- amministrare e gestire risorse, ad esempio per registrare informazioni

riguardanti la loro locazione e acquisizione

- registrare diritti di proprietà intellettuale delle risorse
- gestire l'accesso al livello utente alle risorse
- descrivere la correlazione esistente tra le risorse e gli ambienti software e hardware coinvolti
- preservare a lungo termine le risorse digitali e la loro autenticità

L'OAI prevede che i metadati disseminati dai repository abbiano le proprietà di *estendibilità* e *interoperabilità*, per soddisfare tali requisiti la scelta è ricaduta sul Dublin Core Semplice (DC).

Per quanto concerne la prima proprietà il DC permette alle comunità, i cui bisogni non vengano del tutto soddisfatti dai suoi 15 elementi base, di estendere tale insieme creando e aggiungendo nuovi elementi descrittivi.

La seconda proprietà, invece, viene soddisfatta in quanto il DC deve essere utilizzato come “lingua franca” per garantire la soglia minima di interoperabilità, definendo un insieme di attributi stabiliti a priori.

Un altro standard di metadati più ricco, utilizzato principalmente nell'ambito della catalogazione bibliografica, è il MARC 21; adottato come formato di rappresentazione interna dei metadati e come formato aggiuntivo di scambio da CDSware, applicazione trattata nella parte applicativa della tesi.

Entrambi i formati verranno trattati più dettagliatamente nei seguenti capitoli.

BIBLIOGRAFIA

- [B3] De Robbio Antonella, “Metadati per la comunicazione scientifica”,
Biblioteche Oggi, dicembre 2001.