Metadati per la comunicazione scientifica

Strumenti efficaci per lo sviluppo dei sistemi di informazione digitale in rete

di Antonella De Robbio

Metadata means different things to different communities [...] Although these communities have different aims and different motivations the tools and standards that are used are often the same.
(Matthew J. Dovey)[1]

I metadati sono ad oggi un tema cruciale in molti contesti, non solo bibliotecari. Anne Gilliland-Swetland, per rappresentare quello che significa "metadato", utilizza la seguente metafora: "the sum total of what one can say about any information object at any level of aggregation".
I metadati non debbono necessariamente essere percepiti come informazioni digitali: il retaggio culturale e i professionisti dell'informazione da sempre hanno creato metadati, i quali fino ad oggi sono stati gestiti all'interno delle collezioni tradizionali. Ora, con l'avvento dei sistemi di informazione digitale, i metadati si esprimono in differenti modi e differenti forme, incorporati all'interno di biblioteche digitali, incarnati nei documenti contenuti entro periodici elettronici, o collegandosi sugli archivi aperti (*open archives*) fornendo accesso ai contenuti scientifici.
I metadati, come parola chiave implicita o esplicita, incardinati in protocolli, espressi in standard entro piattaforme condivise, configurati in formati differenti, interni o esterni ai documenti che essi rappresentano, nelle loro varie manifestazioni, aprono le vie ad una comunicazione più estesa, mettendo in connessione differenti mondi con regole diverse.
Va quindi posta grande attenzione al loro trasporto tra server comunicanti.

L'articolo propone i temi affrontati dall'autrice nel seminario nazionale "Verso un'interoperabilità tra sistemi, biblioteche, musei e archivi", tenutosi a Roma il 3 aprile 2001 e organizzato dal Gruppo di studio sugli standard e le applicazioni di metadati nei beni culturali, che nasce e opera in contesto ICCU (Istituto centrale per il catalogo unico e le informazioni bibliografiche).

Comunicazione scientifica, dati e metadati

L'accesso all'universo delle risorse disponibili online è divenuto recentemente uno degli obiettivi primari di molte istituzioni scientifiche che gestiscono risorse informative, ma che soprattutto producono informazioni di contenuto.
In parallelo, la descrizione degli oggetti che compongono le collezioni di molte istituzioni, scientifiche e non (biblioteche, musei, archivi...) è stata gestita negli anni recenti attraverso sistemi di automazione che hanno fornito strumenti per un controllo più efficiente delle collezioni.
Tali descrizioni, che prima erano effettuate con sistemi non automatizzati, nel processo di transizione dalla fase tradizionale a quella automatizzata hanno subito un analogo processo di ricognizione strutturale e concettuale. Molti dei sistemi gestionali ad oggi esistenti sono stati costruiti senza nessun riferimento a standard di comunità e quindi non sono prodotti condivisibili all'interno della rete.
Molto spesso, soprattutto in ambito scientifico-accademico, l'informazione prodotta in settori disciplinari specifici è descritta con "linguaggi" compresi solo da quelle specifiche comunità che fanno riferimento a quel determinato linguaggio e quindi tali insiemi informativi restano isolati.
Esistono perciò differenti sistemi che mantengono un proprio disegno strutturale sia per la composizione dei campi (descrittivi o di accesso), sia per la presentazione complessiva dei dati descrittivi all'interno del sistema. Tali differenze si fanno ancora più sensibili quando ci si riferisce a stru-

[1] Matthew J. Dovey, *"Stuff" about "Stuff",* "Vine", 1999, 116.

menti di supporto per il recupero dell'informazione, quali schemi di classificazione e vocabolari controllati.
Una più puntuale concettualizzazione di metadato è necessaria dal momento che i professionisti dell'informazione considerano che le loro attività saranno trasferite e ricomprese all'interno della sfera dei sistemi di informazione digitale.
L'avvento del web e la crescita esponenziale delle risorse elettroniche ha incrementato anche la domanda dell'utenza relativamente alle effettive capacità di avere strumenti che consentano di ricercare attraverso differenti strutture di metadati in modo simultaneo.
Le necessità sempre crescenti delle fasce di utenza specializzata di poter recuperare informazioni da contenitori integrati, ricreando ambienti interdisciplinari, ha motivato molte istituzioni scientifiche a convertire i loro metadati "non standardizzati" in formati più facilmente accessibili.
In molti casi le istituzioni scientifiche hanno messo a disposizione i contenuti intellettuali a testo pieno prodotti all'interno dei propri circuiti disciplinari, esponendone le descrizioni (metadati) di modo che queste potessero essere "raccolte" da sistemi adeguatamente messi a punto. Tali raccolte di metadati generano solitamente indici cumulativi centralizzati, che sono il prodotto delle raccolte dai vari server distribuiti lungo la rete e che espongono i metadati descrittivi.
In tale modello distribuito, noto come Harvest, vi sono quindi due livelli: il livello dei singoli server (*repositories*), dove risiede l'informazione a testo pieno prodotta dalle istituzioni scientifiche, e il livello del server che contiene i metadati.

Le diverse categorie di metadati

I metadati non sono riferibili soltanto alla semplice descrizione di un oggetto.
Nel circuito della comunicazione scientifica, in particolare, i metadati possono assumere, all'interno di piattaforme comuni, differenti connotazioni a seconda del ruolo al quale essi sono preposti o in relazione alle funzioni che essi svolgono.
I metadati sono qualcosa di mobile nel tempo e nello spazio, essi infatti continuano ad accrescersi entro il sistema in cui dimorano, durante il ciclo di vita dell'oggetto informativo con il quale si relazionano. I metadati, intesi in senso moderno, non possono essere considerati "definitivi", in quanto una volta creati non rimangono statici, ma vengono modificati nel corso del tempo e qualche volta possono anche essere disposti in punti differenti durante l'arco della vita di una risorsa, soprattutto se digitale.
Fino ad oggi i professionisti dell'informazione, e in particolare gli archivisti e i bibliotecari, hanno focalizzato la loro attenzione su metadati in associazione ad attività di descrizione o catalogazione.
Il Dublin Core infatti, formato da metadati di tipo descrittivo, fino ad oggi ha supplito anche ad altre funzioni, ora riconducibili sotto differenti tipologie di metadati, più specifici e meglio adeguati a scopi non puramente descrittivi.
Con la parcellizzazione delle conoscenze anche le specificità degli stessi metadati è andata via via differenziandosi, a seconda delle condizioni in cui i metadati si muovono insieme agli oggetti a cui essi si riferiscono. I metadati, infatti, possono non solo descrivere o catalogare un oggetto, ma anche indicare il contesto, la gestione, i processi, la conservazione e l'uso della risorsa che è stata descritta.
Varie sono le fonti da cui i metadati possono essere originati. Per molto tempo queste "informazioni sulle informazioni" sono state create da professionisti dell'informazione, attraverso attività manuale umana. Ora anche gli utenti meno esperti sono potenzialmente messi in grado di creare metadati che descrivono i documenti da loro stessi prodotti, attraverso interfacce web che consentono inserimenti manuali in stringhe preconfezionate. Da sempre gli autori di contenuti intellettuali accademico-scientifici hanno utilizzato metadati per descrivere e indicizzare le loro creazioni nei circuiti di comunicazione all'interno del proprio ambito disciplinare.
Ad oggi vi sono metadati che possono essere creati automaticamente dalla macchina e possono essere interconnessi attraverso complesse relazioni che creano legami tra risorse differenti.
È importante comprendere che un metadato di un oggetto informativo può essere simultaneamente un dato di un altro oggetto informativo.
I metadati assumono importanza strategica nello sviluppo dei sistemi di informazione digitale in rete, e in tale contesto ogni previsione conduce a un concetto di metadato ampio e allargato. Per capire meglio il concetto di metadato risulta utile frazionarlo in categorie distinte che riflettono gli aspetti chiave delle funzionalità del metadato.
Differenti funzionalità di metadati conducono a differenti tipologie:
– *amministrativi gestionali MAG* (usati per la gestione e amministrazione delle risorse informative);
– *descrittivi* (dal MARC al Dublin Core);
– *conservazione* (compresa migrazione);
– *tecnici* (comportamento dei metadati e funzionamento dei sistemi);
– *utilizzo* (relativi al livello e al tipo di utilizzo dell'utente)
Nella tabella 1 di p. 56 sono riportati i diversi tipi di metadati e le loro funzioni.
In questo contesto, un oggetto informativo è qualcosa che può essere indirizzato e manipolato sia attraverso l'attività umana sia attraverso un sistema automatico come un'entità discreta. Tale oggetto può essere costituito da un singolo termine, o può essere un aggregato di più argomenti.
In generale tutti gli oggetti informativi, indipendentemente dalla forma fisica o intellettuale che essi assumono, hanno tre caratteristiche: contenuto, contesto e struttura, le quali possono essere ricomposte attraverso il metadato.
Il *contenuto* stabilisce un nesso tra ciò che l'oggetto contiene in sé, o nei suoi dintorni, ed è intrinseco all'oggetto informativo stesso.
Il *contesto* indica il chi, cosa, perché, dove, come; tutti aspetti associati con la creazione dell'oggetto e comunque estrinseci all'oggetto informativo stesso.
La *struttura* si riferisce all'insieme formale di associazioni situate all'interno o tra gli oggetti informativi. Tali insiemi possono essere intrinseci o estrinseci agli oggetti.

I metadati nell'infrastruttura Open Archive Initiative (OAI)

Nel sistema Open Archive Initiative (OAI), per esempio, ➤

Tab. 1 - Diversi tipi di metadati e loro funzioni[2]

Tipologia	Definizione	Esempi
Amministrativi	Metadati utilizzati nella gestione e nell'amministrazione delle risorse informative	- Informazioni sull'acquisizione - Tracciato storico dei diritti di proprietà intellettuale, cessione e passaggi ai fini della riproduzione - Documentazione dei requisiti di accesso legale - Informazioni sulla reperibilità - Criteri di selezione per la digitalizzazione (= formato, set di caratteri) - Controllo della versione e distinguibilità fra oggetti informativi simili - Tracce di controllo create da sistemi di gestione di metadati (*recordkeeping*)
Descrittivi	Metadati utilizzati per descrivere o identificare risorse informative	- Registrazioni catalografiche - Indicazioni di aiuto per il reperimento - Indicizzazione su database specialistici - Connessioni fra risorse tramite link web - Annotazioni di utenti - Metadati per sistemi di gestione (*recordkeeping*) generati dai programmi di creazione delle registrazioni
Sulla conservazione	Metadati riferiti alla gestione della conservazione delle risorse informative	- Documentazione della condizione fisica delle risorse - Documentazione delle azioni intraprese per conservare le versioni fisiche e digitali delle risorse, per esempio ripristino (*refreshing*) e migrazione dei dati
Tecnici	Metadati riferiti al funzionamento di un sistema e al comportamento dei metadati	- Documentazione sull'hardware e il software - Informazioni sulla digitalizzazione, per esempio formati, rapporti di compressione, procedure di graduazione (*scaling routines*) - Tracciato storico dei tempi di risposta di sistema - Dati di autenticazione e sicurezza, per esempio chiavi crittografiche, password
Di utilizzo	Metadati riferiti al livello e al tipo di utilizzo delle risorse informative	- Registrazioni di visualizzazione (*exhibit records*) - Tracciato storico dell'uso e gestione dei profili utenti - Informazioni sulla riutilizzazione del contenuto e sull'esistenza di una pluralità di versioni (*multi-versioning information*)

il protocollo Open Archives Metadata Harvesting (OAMH) definisce il meccanismo per la raccolta dei dati contenenti i metadati dai vari *repositories*.
L'attuale infrastruttura tecnica di OAI è strutturata in modo che a monte, i data provider, debbano esporre, sempre attraverso quanto specificato nel protocollo OAMH, i propri metadati attraverso un protocollo HTTP. Il protocollo OAMH definisce quindi i meccanismi di "esposizione" per i provider e di raccolta, attraverso un programma che colleziona le informazioni indicizzate dalle varie collezioni.
Filosofia portante di OAI, che adotta un set minimo del formato Dublin Core per lo scambio dei metadati, è la netta distinzione tra le due componenti, e precisamente i data provider e i service provider.
L'approccio noto come *metadata harvesting,* consente di mettere in relazione questi due settori distinti attraverso un colloquio di scambio di informazioni. Da una parte ci sono i server contenenti le "informazioni sulle informazioni" ovvero i "dati sui dati" o metadati, dall'altra i servizi, o service provider, raccolgono i dati "esposti" dai data provider e li organizzano in servizi a valore aggiunto da offrire alle comunità scientifiche a seconda delle necessità espresse da quella particolare comunità di utenti.
La raccolta dati in OAI avviene all'interno di un modello di tipo distribuito che si fonda sull'architettura Harvest, in un meccanismo di scambio tra *gatherer* e *brokers*.
Ogni *gatherer* estrae informazioni indicizzate dalle collezioni e le trasmette in un formato standard, attraverso un protocollo standard, a programmi chiamati *brokers*. Il *broker* costruisce un indice combinato di informazioni sulle varie collezioni.
Il protocollo OAI, per esempio, non prescrive i mezzi di associazione tra il metadato e il relativo contenuto, ma visto che le necessità primarie sono quelle di aver accesso al contenuto associato a quello specifico metadato, i data provider con i loro archivi di contenuti possono ritenere di definire link specifici all'interno del metadato verso il contenuto. In tale ottica il formato obbligatorio Dublin Core fornisce gli elementi identificatori che possono essere usati a questo scopo.

Metadati a servizio dell'utenza scientifica

Nel mondo della comunicazione scientifica, alcune esperienze di e-print server, all'interno dell'architettura OAI, si sono

[2] Tratta e da me liberamente tradotta da *Introducion to metadata: pathways to digital information,* a cura di Murtha Baca, Getty Standards Program, <http://www.getty.edu/>.

orientate verso una fase ancora più evoluta: dalla conversione dei loro metadati in un formato accessibile per una raccolta cumulata si è passati alla creazione di una singola interfaccia per la ricerca simultanea attraverso differenti ed eterogenei database.
Se l'obiettivo è quello di disegnare interfacce o convertire dati in un nuovo standard, la prima tappa fondamentale è analizzare gli elementi informativi in ogni database e correlare i campi informativi discreti nei differenti database che hanno lo stesso, o similare, significato. Ciò significa costruire mappature tra metadati o mappature semantiche.
I *crosswalks* all'interno delle mappature sono importanti non solo a supporto della necessità di offrire un unico punto di riferimento *one-stop shopping,* o quali strumenti di ricerca incrociata, ma in quanto sono strumenti utili alla conversione dei dati da un formato a un altro formato accessibile in modo più esteso.
I *crosswalks,* in sostanza, sono rappresentazioni visuali o "mappe" che mostrano le relazioni tra diversi contesti. In altre parole essi supportano l'interoperabilità semantica.
Il processo di *mapping* supporta la capacità di un motore di ricerca di interrogare campi con contenuto uguale o quanto meno similare in differenti database.
Nel prototipo UPS (Universal Preprint Service), per esempio, messo a punto da Herbert Van de Sompel,[3] Thomas Krichel,[4] Michael L. Nelson[5] e altri, i dati originati dai differenti archivi distribuiti furono raccolti e ricondotti sotto un unico formato e posti entro un sistema centralizzato munito di motore di ricerca specifico e sistema di *linking,* al fine di fornire un servizio ritagliato sull'utenza scientifica.
Il concetto di fondo era quello di offrire una biblioteca digitale multidisciplinare contenente materiale scientifico e pubblicamente accessibile. A tal fine il prototipo raccolse quasi 200.000 record dai differenti archivi creando un ambiente adatto all'utenza finale. Gli archivi interessati in tale processo furono sei: arXiv.org (noto come Los Alamos E-Print Archives), Cognitive Sciences Eprint Archive (CogPrints), Digital Library for the National Advisory Committee for Aeronautics (NACA), Networked Computer Science Technical Reference Library (NCSTRL), Networked Digital Library of Theses and Dissertations (NDLTD) e Research Papers in Economics (RePEc).
I metadati collezionati durante la fase di *data gathering* erano espressi in una varietà di formati tutti differenti l'uno dall'altro, a seconda dell'archivio dal quali essi provenivano. All'interno di UPS, come formato di metadato comune fu scelto il formato ReDIF version 1, messo a punto da Thomas Krichel[6] e usato nell'iniziativa RePEc. Tutti i dati furono quindi convertiti nel formato ReDIF e di conseguenza fu prodotta una mappatura tra i formati non ReDIF e il formato ReDIF.
Il formato ReDIF inspirato a IAFA[7] si basa sul protocollo Guildford[8] e viene utilizzato, oltre che dalla rete NetEC per gli economisti, anche dal servizio DoIS,[9] di ambito biblioteconomico, per i documenti LIS. Il formato ReDIF che nasce in contesto scientifico si riferisce a metadati che colloquiano entro il sistema e che possono avere funzioni diverse.
Riporto due esempi di metadato ReDIF, il primo (fig. 1 di p. 58) si riferisce al dato contenuto nella serie di un determinato archivio, il secondo (fig. 2 di p. 58) si riferisce al metadato per la struttura e definizione di un archivio (nel caso specifico dell'archivio correlato alla serie che contiene il metadato di cui all'esempio nella fig. 1).

Formati di metadati accademici

Il metadato Dublin Core in OAI convive con altri metadati, in quanto l'architettura OAI supporta insiemi paralleli di metadati, disseminati dagli archivi distribuiti, nei seguenti formati:
– OAI DC Dublin Core codificato in XML
– OAI RFC1807 RFC1807 codificato in XML
– ArXiv (Old e Test) codificati in XLM
– AMF Test-bed for Academic Metadata Format.
Di estremo interesse, all'interno della comunicazione scientifica, è il formato messo a punto da Thomas Krichel e Simeon Warner, denominato Academic Metadata Format (AMF).[10]
AMF codifica descrizioni di documenti creati e usati da accademici, delle organizzazioni e degli enti importanti nel mondo accademico, comprese università, centri di ricerca, editori accademici, società scientifiche, associazioni e agenzie appartenenti al settore della ricerca.
Le caratteristiche primarie di AMF sono che: è codificato in XML ed è costruito su una struttura composta da elementi e attributi:
– NOMI (*nouns* di quattro tipi: persone, organizzazioni, documenti, gruppi);
– VERBI (*verbs,* le relazioni);
– AGGETTIVI (*adjectives,* aggiungono informazioni ai *nouns;* strutture annidate).
Per quanto riguarda la relazione AMF/RDF (Resource Description Framework) va detto che il draft ancora non ➤

[3] All'epoca del prototipo UPS (1999) a Los Alamos National Laboratory-Research Library, New Mexico, US, and Automation Department of the Central Library of the University of Ghent, Belgium.

[4] All'epoca del prototipo UPS (1999) all'University of Surrey, UK.

[5] NASA Langley Research Center, Hampton VA, USA.

[6] THOMAS KRICHEL, *ReDIF version 1.* <http://openlib.org/acmes/root/docu/redif_1.html>.

[7] "Internet Anonymous FTP Archive templates. Metadata format designed for Anonymous FTP archives, now adapted for use in ROADS project", definizione tratta da *Metadata glossary:* <http://www.ukoln.ac.uk/metadata/glossary/>.

[8] Il protocollo Guildford prende il nome della città dove è situata la Surrey University dove il protocollo è stato messo a punto nell'ambito della rete NetEC per l'economia: <http://www.workingpapers.de/RePEc/all/root/docu/guilp.html>, <http://netec.wustl.edu/WoPEc/data/Papers/rpcrdfdocGuildP.html>.

[9] Documents in Information Science, <http://dois.mimas.ac.uk>.

[10] THOMAS KRICHEL – SIMEON WARNER, *Vocabulary for the Academic Metadata Format,* draft, 26.03.2001.

Fig. 1 – Metadato ReDIF che descrive un dato (documento)

Template-Type: ReDIF-Article 1.0
Title: Online Resources for Mathematics in the Scientific Virtual Reference Desk
Author-Name: Antonella De Robbio
Author-Email: derobbio@math.unipd.it
Author-Workplace-Name: Biblioteca del Seminario Matematico Università degli Studi di Padova
Author-Workplace-Homepage: http://www.math.unipd.it/~derobbio/home/antohp.htm
Abstract: The present work briefly describes the Virtual Reference Desk for mathematics elaborated during the time I worked at the CERN Library [1] (European Laboratory for Particle Physics or Laboratoire européen pour la physique des particules) in Geneva. This instrument is dedicated to the CERN librarians, with whom I have shared important moments of my professional career. In particular, I would like to gratefully acknowledge their valuable co-operation and assistance during our time spent working together. The Web metasource is comprised of three directories, annotated and interrelated with dual application: The first is intended as a work tool for librarians working in mathematics libraries, but above all for librarians of high energy physics, who more often than not must turn to mathematics and the use of mathematical applications and models for the physical sciences and in particular particle physics. The second is an on-line resource for mathematics; that is, a Virtual Reference Desk for the community of mathematicians, with whom I have been collaborating for some twenty years at the University of Padova. The bibliographical instrument is born from the need to have at our disposal a scientific Virtual Reference Desk created according to the needs of those working in physics and mathematics libraries – a tool which is comprised of materials collected during years of work as much as material available on-line through the use of new technologies.
Journal: High Energy Physics Libraries Webzine
Issue: 3
Year: 2001
Publication-Status: Published
File-URL: http://library.cern.ch/HEPLW/3/papers/4/
File-Format: Application/HTML
File-Function: Full text
Handle: ReLIS:hep:heplwz:i:3:y:2001:p:4

adotta RDF, tuttavia se le risorse descritte sono identificate, c'è una mappatura diretta tra AMF e uno schema che usa RDF.
Tale formato sarà formalmente definito più avanti nel prossimo documento denominato Academic Reality Description (ARD).
Ad oggi non vi è ancora un modo diretto per veicolare i MAG all'interno di AMF. Tale metadato amministrativo gestionale dovrebbe essere associato al trasporto stesso dei metadati.
Per esempio, se si utilizzano protocolli OAI (Dienst, Guildford), allora i MAG potrebbero essere trovati nell'HEADER OAI che avvolge il metadato AMF.
Un discorso a parte merita il trattamento degli schemi di classificazione all'interno dell'Academic Metadata Format.
In mancanza di standard sui metadati per la descrizione di schemi di classificazione, il draft AMF ha scelto di utilizzare i nomi di gruppo che permettono di specificare strutture gerarchiche (gruppi di gruppi, ecc.) conformi agli schemi.
Questa soluzione presenta ad oggi qualche problema, ed è difficilmente trasportabile all'integrazione di vocabolari controllati dentro AMF (vedi fig. 3).
L'integrazione dell'informazione sulle classificazioni dentro AMF dovrebbe necessariamente mantenere un insieme di identificatori per ogni differente schema classificatorio, o meglio per ogni versione di ciascuno schema. Inoltre c'è il problema di collegare i codici alle descrizioni e questo richiede metadati specifici sugli schemi di classificazione.
I prossimi sviluppi di AMF coinvolgeranno quindi la codifica degli schemi di classificazione entro la struttura AMF e sarà estremamente interessante dirigere l'attenzione verso tale direzione.

Fig. 2 – Metadato ReDIF di archivio

Template-type: ReDIF-Series 1.0
Name: High Energy Physics Libraries Webzine
Type: ReDIF-Article
Description: Collection of documents on high energy physics libraries from the point of view of both information workers and library clients
Provider-Name: CERN, Geneva
Maintainer-Name: Antonella De Robbio
Maintainer-Email: derobbio@mail.cern.ch
Handle: ReLIS:hep:heplwz

Metadati per la validazione della qualità dei contenuti

All'interno dei gruppi di lavoro del workshop che si è tenuto a Ginevra, presso il CERN, lo scorso mese di marzo,[11] su cui si basano molte delle osservazioni riportate nel presente intervento, si è tentato di identificare e discutere le proprietà e le caratteristiche principali richieste agli OA di materiale non soggetto a *peer-review*. Tale materiale dovrebbe essere configurato entro reali blocchi da costruire all'interno di un nuovo meccanismo di comunicazione scientifica tale da indirizzare i bisogni della scienza, della comunità scientifica e della società pubblica in generale verso gli archivi aperti messi a disposizione dalle istituzioni scientifiche, piuttosto che verso l'attuale sistema dei costosissimi periodici tradizionali.
Ci sono i server coi documenti, i data provider; ora è necessario passare a uno step successivo, ovvero alla validazione del

[11] Cfr. Antonella De Robbio, *Open Archive Initiative (OAI) in Europa,* "Biblioteche oggi", 19 (2001), 4, p. 66-69.

materiale contenuto nei server. Come attuare questo processo in modo economico ed efficace? Dato per certo il vasto consenso attorno alla questione della necessità di trovare forme di certificazione dei lavori scientifici che ne garantiscano la qualità, la questione principale che ci si è posti a Ginevra era se e come può essere possibile implementare, in una infrastruttura di nuova concezione, funzioni che provvedano, attraverso sistemi accettati dalla comunità scientifica, alla gestione di processi di certificazione.
I metadati devono quindi essere visti come utensili per gli autori e per la comunità scientifica, come strumenti in grado di conformarsi all'interno di una piattaforma adatta alla regolazione di processi di validazione. Tali processi di validazione possono essere attuabili attraverso il supporto che le biblioteche, nel ruolo di service provider, possono offrire ai creatori di contenuti.
A Ginevra si sono esposte alcune ipotesi. In particolare è risultato di estremo interesse il lavoro effettuato dal CERN stesso di Ginevra e presentato da Thomas Baron e Tibor Simko.[12] Attraverso il processo di *submission* da parte degli stessi autori, processo tutto automatizzato e altamente personalizzato (la nuova versione è stata messa a punto nel settembre scorso), è possibile attuare alcune strategie orientate alla validazione dei documenti sottomessi. Ciò è possibile utilizzando differenti percorsi, avvalendosi anche di tag entro i metadati. Sono stati illustrati cinque differenti processi di validazione e varie ipotesi correlate.
Dalla scorsa estate lo staff CDS (CERN Document Server) che lavora in stretto contatto coi bibliotecari, ha messo a punto un motore di ricerca *ad hoc* che consente l'interrogazione a testo pieno sugli oltre 170.000 documenti contenuti in CDS. La ricerca full-text è resa possibile attraverso un'interfaccia dinamica costruita con WebLib. Vediamo in sintesi in quale contesto si muovono i processi di *submission* dei lavori dei fisici che afferiscono al CERN Document Server.
Cosa contiene CDS:
– HEP documents: preprints, books, journals, photos, notes, presentations, meeting agendas, ecc. (25 tipi diversi);
– caricamento di circa 4.000 e-print/mese;
– 430.000 record bibliografici;
– 170.000 documenti full-text ricercabili da interfaccia WebLib;
– sistema gestionale: Aleph 300 (ExLibris);
– interfaccia WebLib personalizzata (creazione di interfacce dinamiche);
– database MySQL separato per documenti non della biblioteca (archivi, press cutting, directories, fotografie museo …).
Modalità di acquisizione dei metadati @ CERN:
– manuale (8 per cento):
collezione di documenti scansiti (scanner);
– elettronica:
web & email *submission mechanism,*
applicazione del sistema Uploader per il caricamento automatico e trasformazione dei metadati;
– sistema per l'archiviazione a lungo termine;
– cinque differenti *approval approaches* per la validazione dei documenti:
da niente alla recensione completa.
Una delle proposte del CERN ai fini di una certificazione di qualità dei documenti all'interno dei server è quella di una validazione attraverso i metadati. Il CDS del CERN è pronto per la compatibilità OAI nel ruolo di data provider e attualmente nella filosofia di OAI la qualità del documento non è registrata.
Come ottenere valore aggiunto attraverso la validazione?
La semplice soluzione del CERN è quella di aggiungere un'etichetta <qualità> all'interno dei metadati.
Tale etichetta può essere di due tipi: Set-wide e Record-specific.
Anche Billy Arms,[13] nel suo magistrale intervento, ha ➤

Fig. 3

```
<amf>
   <group id="csfhrd">
   <title>Classification Scheme for Human Rights Documentation</title>
   <homepage>http://www.huridocs.org/clasengl.htm<homepage>
   <haseditor>
      <person>
      <name> Ivana Caccia </name>
      <email> icaccia@web.apc.org </email> br()i()i()</person>
   </haseditor>
   <haspart>
      <group id="csfhrd:GEN II.10"><title> natural justice </title></group>
      <group id="csfhrd:GEN II.20"><title> universality / relativism </title></group>
      <group id="csfhrd:GEN II.30"><title> philosophy & human rights </title></group>
      <group id="csfhrd:GEN II.40"><title> political theories & human rights </title>
      <haspart>
         <group id="csfhrd:GEN II.41"><title> democracy </title></group>
         <group id="csfhrd:GEN II.42"><title> liberalism </title></group>
         <group id="csfhrd:GEN II.45"><title> marxism </title></group>
      </haspart></group>
   </haspart></group>
</amf>
```

Fig. 4

```
Quality Metadata: Example by W. Arms
<oai-quality>
    <category> internal </category>
    <process>
            peer review
    </process>
    <organization>
            CERN
    </organization>
    <policies>
            http://www.cern.ch/policies/review.html
     </policies>
</oai-quality>
```

[12] Thomas Baron – Tibor Simko, *CERN document server: validation and OAI,* <http://documents.cern.ch/archive/electronic/other/agenda/a01193/ a01193s3t2/transparencies/Baron.ppt>.

[13] Billy Arms, *WY, quality control in scholarly publishing. What are the alternatives to peer-review?,* <http://documents.cern.ch/archive/electronic/ other/agenda/a01193/a01193s4t3/transparencies/Arms.ppt>.

parlato di livelli diversi di certificazione attraverso l'utilizzo di metadati. I documenti a vario livello di certificazione potrebbero configurarsi entro i server nelle differenti condizioni di:
– validati o non validati;
– validazione interna o esterna;
– tramite *peer-review* o altro processo;
– istituzione che ha proceduto alla validazione
... e così via.

Il metadato di qualità proposto da Arms avrebbe una struttura come quella riportata nella figura 4.

What about intellectual property rights management inside metadata?

Alcuni metadati definiscono o prescrivono schemi per la gestione dei diritti di proprietà intellettuale e di accesso ai contenuti. Non è ancora del tutto chiaro a quali tipologie di metadati saranno demandate tali funzioni di controllo. Alcuni elementi in metadati descrittivi, per esempio nel Dublin Core, includono indicazioni sommarie relative ai diritti di proprietà, ma ciò non è sufficiente. Controlli sulla gestione dei diritti potrebbero essere configurati all'interno di metadati più specifici, per esempio metadati amministrativi gestionali, come il controllo sui profili utenti per la definizione degli accessi o delle restrizioni di accesso, in certi sistemi demandato ai metadati di utilizzo.
È opinione diffusa negli ambienti accademici che dovrebbe essere responsabilità dei data provider, adottando un determinato protocollo piuttosto che un altro, esporre i metadati in modo che tali informazioni possano essere trasparenti. È importante che i *repositories* contenenti documenti accademici definiscano i permessi di accesso o le restrizioni di accesso o di utilizzo al materiale in essi raccolto, dando indicazioni sui modi con cui si consente l'accesso ai contenuti intellettuali (accesso per dominio IP, password...). Queste informazioni dovranno essere incardinate in metadati specifici connessi ad altri metadati, affinché il sistema sia reso trasparente.
Per concludere, ciò che dobbiamo sapere è che l'esistenza di una molteplicità tipologica di metadati sarà il nodo cruciale per una continuità di accesso e di utilizzo, fisico e intellettuale, delle risorse digitali informative e degli oggetti informativi che esse contengono. Riprendendo Anne Gilliland-Swetland, è proprio in questo senso che

> i metadati ci forniscono la Stele di Rosetta che renderà possibile decodificare gli oggetti informativi e le loro trasformazioni all'interno della conoscenza nei sistemi informativi di retaggio culturale del ventunesimo secolo.