

Università di Pisa

Interfacoltà tra la Facoltà di Lettere e Filosofia e la
Facoltà di Scienze matematiche, fisiche e naturali

Laurea Specialistica in Informatica Umanistica
Curriculum Management della Conoscenza

TESI DI LAUREA

**Proposta di trans-codifica per gli oggetti
di conoscenza. Un punto d'incontro tra
l'editoria libraria e le biblioteche digitali**

Relatore

Prof. Francesco Varanini

Candidato:

Alice Guidi

Anno Accademico 2008/2009

Un libro, uno solo, può sconvolgere la vita di un singolo individuo. Una biblioteca può arricchire culturalmente una piccola cittadina. Un gigantesco archivio, invece, non cambierà in nulla la vita di una metropoli." Arthur Schopenhauer

Introduzione

L'idea della tesi nasce da una ricerca sugli standard utilizzati nella rappresentazione dei libri per le biblioteche digitali. Mentre infatti il contesto tecnologico può considerarsi piuttosto consolidato, gli oggetti digitali prodotti presentano una quantità e un'eterogeneità ancora in atto.

L'interesse si è concentrato sulla definizione di metadati utili per un catalogo editoriale. Questa necessità è nata dall'esperienza di tirocinio, ovvero il recupero del catalogo storico editoriale della casa editrice *Guerini e Associati* (22 anni di attività). Costruito un unico catalogo che contenesse tutti i titoli (sono state integrate diverse fonti, digitali e cartacee) è stato successivamente codificato con XML. Poiché non esistono precedenti di utilizzo di metadati per archivi editoriali, lo sguardo si è soffermato sull'utilizzo di quest'ultimo nel paradigma delle biblioteche digitali, dove esistono sicuramente delle diversità dovute a necessità non del tutto affini, ma si tratta sempre di catalogazione di risorse librarie.

Tra decine di standard funzionali alla catalogazione dei record bibliografici si è scelta la TEI. Come punto di partenza è stato fissato quello di riuscire a rispondere a due funzioni a cui il catalogo deve assolvere: il reperimento e la localizzazione delle informazioni. Il catalogo codificato pone l'utente in grado di trovare, attraverso una ricerca in cui si utilizzano attributi o relazioni, risorse bibliografiche presenti nel catalogo, di localizzare una singola risorsa e/o insieme di risorse che corrispondono alle specifiche richieste.

TEI nasce come standard per la codifica di testi umanistici letterari, dove si guarda molto al contenuto, ma non solo. Mi sono lasciata convincere da questa scelta, pensando ad una prospettiva. Data la grande versatilità e la vasta offerta di tag che offre TEI si presta bene anche in previsione di un nuovo progetto: l'inserimento del full text on line, che viene associato ad una risorsa, con possibilità di vendita, integrale o parziale dei capitoli. Sarebbe un servizio con un valore aggiunto che permetterebbe di distinguersi dalla concorrenza, specialmente se si riesce a presentare i testi correlati in modo intelligente. Se

inserisco i tag TEI anche al testo (non solo elementi strutturali, quali capitolo, fine pagina ecc.), scavando la sua semantica, l'utente può reperire un materiale selezionato e ragionato, che va al di là del testo trovato con una semplice ricerca testuale. Si possono scrivere *query* più raffinate, perchè se ho aggiunto un elemento, specificando ad esempio tutti i <persName> con un attributo che ne descrive la professione o il ruolo i risultati saranno molto dettagliati. A tale proposito citerò, per capire quale può essere un uso appropriato, il mio lavoro di tesi della laurea triennale.

Il programma di ricerca ha evidenziato le caratteristiche peculiari dell'informazione di carattere scientifico avvalendosi degli studi finora effettuati ed evidenziando l'evoluzione formale e applicativa dei formati descritti per la catalogazione delle risorse nel corso degli anni più recenti e valutandone il grado di affidabilità.

Questo nuovo modello prospetta una diversa interpretazione dei formati descrittivi che consentono l'organizzazione di un sistema dinamico che nella rete si configura, più che come una base di dati di registrazioni statiche, come un insieme di processi informativi dinamici.

L'informazione in rete è un flusso dinamico che esalta la natura flessibile ed il continuo mutamento della comunicazione: il profilo dell'ambiente digitale che emerge, con le interconnessioni, si distacca nettamente dalla visione precedente guidata dal principio di accesso diretto ai contenitori, sia pure magazzini virtuali.

L'attività di ricerca si è quindi focalizzata sullo studio e sulla definizione di un modello di riferimento della meta-informazione per le risorse librarie: tale modello permette di costruire un catalogo elettronico in un formato di interscambio indispensabile al funzionamento avanzato di applicazioni tecnologiche orientate alla rete, in grado di rispondere alle necessità gestionali e di fruizione.

La ricerca si è articolata in una fase compilativa e una applicativa.

Fase compilativa

La fase compilativa, quella vera e propria di ricerca, riguarda l'indagine web, la raccolta e l'analisi delle esperienze nel settore degli standard utilizzati per la

catalogazione di libri. Trattati sulla gestione dei metadati con particolare riferimento alle esperienze americane e europee.

Lo studio ha tenuto conto delle pubblicazioni e delle direttive pubblicate dalle varie organizzazioni di riferimento, ed in particolare, delle raccomandazioni prodotte dal W₃C in materia di semantic web.

Fase applicativa

La fase applicativa si è concretizzata nella partecipazione a un corso estivo, TE@SUMMER CAMP organizzato dalla Oxford University che introduceva tale standard a livello avanzato, e dall'esperienza di tirocinio presso la casa editrice Guerini e Associati.

Tali esperienze hanno permesso di sviluppare una capacità specifica di analisi del significato, ruolo e valore informativo di categorie di metadati e di mettere a punto criteri di scelta personalizzati per la gestione del catalogo. L'esperienza di ricerca ha quindi confermato:

- la correttezza dell'approccio di studio intrapreso in quanto la necessità di fornire sempre nuovi format di metadati assume la caratteristica di una costante della progettazione di applicazioni di rete;
- la necessità di definire un modello di riferimento e una metodologia per effettuare personalizzazioni che, da una parte permettono di concordarsi con le scelte della Dublin Core Metadata Initiative, dall'altra corrispondono alle esigenze delle comunità di riferimento non bibliotecarie (casa editrice nel nostro caso) che generano e ricercano libri in rete;
- il valore del formato dei metadati come componente strutturale multilivello, una componente portante che fa da ponte tra le necessità di organizzare in modo semi-strutturato le informazioni e la necessità di costruire schemi di funzionamento tecnologici;
- l'originalità del risultato complessivo della ricerca nella costruzione di modelli per la conoscenza in un contesto di produzione documentale digitale emergente e fortemente innovativo.

L'investimento in un lavoro concettuale sui metadati diventa quindi ora strettamente legato alla costruzione di una nuova intelaiatura delle informazioni, valutando peso e ruolo degli elementi nelle varie fasi di vita dell'oggetto in analisi e delle numerose interconnessioni che lo riguardano.

La tesi è stata strutturata nel modo seguente:

★ *Capitolo 1*: sguardo al promotore del pensiero delle biblioteche digitali, J.C.R. Licklider

★ *Capitolo 2*: panoramica dei metadati utilizzati sul web per raccolte bibliografiche, archivi, progetti sviluppati a livello internazionale.

★ *Capitolo 3*: accenno alla standardizzazione dei linguaggi ed ai principi di catalogazione. Breve descrizione dei due linguaggi più affermati ed utilizzati oggi nelle principali istituzioni, in materia di biblioteche digitali: MARC21 e Dublin Core. Breve presentazione di WorldCat e GoogleBook, una cooperazione che opera su scala mondiale.

★ *Capitolo 4*: excursus sull'utilizzo di internet da parte delle case editrici e le tipologie di siti web che utilizzano.

★ *Capitolo 5*: descrizione del progetto di stage e tesi, la costruzione del catalogo storico editoriale, la codifica XML/TEI associata e l'utilizzo del motore di ricerca Tauro-core per la formulazione delle query. Creazione del sito web, screenshot delle pagine realizzate.

★ Conclusioni

★ Bibliografia e webliografia.

In appendice sono presenti delle tabelle di conversione di metadati, i cosiddetti crosswalk.

Sono riportate le riproduzioni delle schede di catalogazione utilizzate dalla casa editrice per archiviare il catalogo, oltre ad una scansione del catalogo cartaceo che mostra come si presenta al suo interno e quali informazioni riporta.

E' inserita anche un'immagine che riproduce MetaMap: la mappa vettoriale che rappresenta i metadati con una metafora particolare.

Indice

1. Libraries of the Future	10
1.1 Precognitive system	12
1.1.1 Man's interaction with recorded knowledge	13
1.1.2 Stimare la size	13
1.1.3 Scopi, requisiti progetti e criteri del precognitive system	14
2. Metadati e archivi	21
2.1 Scenario di riferimento	21
2.2 Definizione di biblioteca digitale	25
2.3 Definizione di metadati	32
2.3.1 Caratteristiche ed uso	32
2.3.2 Tipologie di metadati	33
2.3.3 Categorizzare i metadati	35
2.3.4 Etichette (Tags)	41
2.3.5 Gli SCHEMAS	42
2.3.6 Modelli di metadati	48
2.3.7 Metadati per domini	49
2.4 Analisi delle esperienze innovative schematizzazione delle applicazioni	56
2.4.1 Alcuni esempi di digital repository	58
2.4.2 OAIS: Open Archive Information System	61
3. Standard: risorse bibliografiche	64
3.1 Principi di catalogazione	67
3.1.1. Un po' di storia	67
3.2 Automazione delle biblioteche, MARC e DC	83
3.3 WorldCat e Google Book	88

4. Libro e la rete	92
4.1 Le case editrici e l'uso della rete	94
4.2 Il libro dal cartaceo al digitale	97
4.2.1 L'editoria on-demand e il self-publishing	102
5. Tirocinio presso Guerini e Associati	106
5.1 Catalogo anagrafica	106
5.2 Catalogo on line del sito	107
5.3 Catalogo editori Alice	109
5.4 Altre fonti	110
5.5 Modifiche effettuate sui cataloghi	110
6. Codifica XML/TEI	118
6.1 XML	118
6.2 La TEI	120
6.3 Scelta dello standard TEI	122
6.3.1 Elementi utilizzati per la codifica del catalogo storico editoriale	123
6.3.2 Il teiHeader e il tag TEI	128
6.4 Scelta del motore di ricerca per XML: Tauro-core	131
6.5 TRQL, TauRoCore Query Language	135
6.6 Sito web realizzato	136
Conclusioni	141
Tabelle Crosswalk	145
Riproduzione scheda produzione	155
Catalogo cartaceo	159
MetaMap	160
Bibliografia	163
Webliografia	167

Part I

Excursus storico:

Libraries of the future

[I]f information is the currency of the knowledge economy, digital libraries will be the banks where it is invested."

How to build a digital library, written by
Ian H. Witten and David Bainbridge,
Series Editor, 2002.

Libraries of the future

Vorrei iniziare la trattazione della tesi partendo dal promotore, colui che più di cinqt'anni fa aveva intravisto la possibilità di creare un sistema per l'archiviazione di risorse che permettesse un più rapido accesso e reperimento del materiale. Stiamo parlando di Joseph Carl Robnett Licklider, psicologo americano (1915-1990). Durante la seconda guerra mondiale inizia la sua ricerca sulla comunicazione presso il laboratorio di psico-acustica dell'Università di Harvard. Le sue ricerche risultano subito interessanti per l'Office of Strategic Research and Development e per il National Defense Research Council. Licklider nel 1950 entra a far parte del Lincoln Laboratory del MIT. Nel 1957 diventa vicepresidente del BBN dove compra il suo primo calcolatore PDP-1 in produzione ed esegue la prima dimostrazione di file-sharing. Le sue teorie si concentrano sullo sviluppo dei computer, ed ha aperto la strada allo sviluppo di internet. Fu uno dei primi ad intuire l'importanza del computer come strumento di comunicazione di massa e della libera circolazione dell'informazione per la democrazia. Ha finanziato e incoraggiato Douglas Engellbart. Nel 1960 ha pubblicato "Man Computer Symbiosis" in cui propone un'interazione semplificata tra l'uomo e la macchina. Non ha pensato di sostituire l'uomo con la macchina, il suo pensiero era ben preciso: i computer hanno il compito di eseguire grandi calcoli, ma il comando resta all'uomo che sa valutare e ragiona sui risultati ottenuti. Già nel 1962, Licklider formula l'idea di una rete globale di computer che scambiano dati, anticipando di molti anni ciò che sarebbe stato

Internet. Tra il 1962 e 1963 progetta diversi prototipi: internet, mouse, windows e l'ipertesto. Molte delle sue visioni sono ancora oggi attuali. Nel 1959 ha scritto *Libraries of the future* dal quale prenderemo alcuni spunti, delineando il suo pensiero.

Il testo *Libraries of the future* è un resoconto di un lavoro di ricerca durato due anni che studia come la tecnologia possa essere pensata per rendere fruibile l'informazione e facilitarne l'organizzazione. L'uomo lamenta una quantità enorme di materiale di lettura e il poco tempo a disposizione per leggere una mole così cospicua. Nel 1956 la Ford Foundation stabilì *The Council on Library Resources* che aveva l'obiettivo di trovare una soluzione alla proliferazione delle pubblicazioni e al mal funzionamento della ricerca dovuto agli scaffali colmi e al disordine generale. Il concilio finanziò un progetto che cercasse una soluzione per immagazzinare e reperire i dati impiegando personale specializzato. Dunque, agli inizi del 1961, il Concilio avviò una ricerca per determinare le caratteristiche necessarie che dovevano possedere le *library of the future*.

La domanda più ricorrente era senz'altro "Come dovrà essere consultata la biblioteca del futuro?" La questione fu demandata al Dr. Bush, che nel suo articolo del 1945, *As we may think*, aveva idealizzato il Memex, uno strumento per la gestione documentale, dove ogni documento può essere memorizzato e collegato con un indice ad esso associato che ne permette il reperimento. Licklider il quale aveva una formazione orientata alla comunicazione e alle problematiche inerenti alla rappresentazione delle informazioni, fu attratto dal problema e scrisse un trattato eloquente sulla *Library of the future*.

Quindi negli anni '60, Licklider anticipa il cambiamento delle biblioteche attraverso l'uso del computer e parla di "precognitive utility net" come servizio di supporto all'apprendimento. *Precognitive* perchè doveva essere un sistema per l'avanzamento e l'applicazione della *conoscenza*. Piuttosto che essere basato su una collezione di documenti e tags, il sistema precognitivo, si sviluppa su tre elementi chiave:

- ◆ corpus della conoscenza;
- ◆ domande;
- ◆ risposte.

Prevedeva l'elaborazione delle informazioni in conoscenza, delle domande in risposte, tutto doveva avvenire digitalmente. Come poteva essere elaborata l'informazione in conoscenza? Come doveva essere organizzato il corpus della conoscenza? Come fece Bush, anche Licklider osserva come opera il cervello umano. Egli riconobbe che il cervello è costituito da un sistema di neuroni e processi. Questa attività neurale accoglie diversi stimoli come il parlato e la scrittura e in qualche modo li processa e li memorizza affinché supportino il disegno delle inferenze e siano capaci di rispondere alle domande. Il cervello umano processa gli stimoli nel momento in cui l'input viene recepito. Licklider credeva anche che gli uomini pensano manipolando, modificando e combinando schemi o modelli che mettono in relazione cose con altre. La nuova conoscenza, secondo il pensiero di Licklider, si ottiene adattando uno o più schemi già acquisiti rimodellandoli a nuove situazioni.

1.1 Precognitive system

L'elaborazione delle informazioni secondo il modello del precognitive system è suggerita dalle seguenti attività che giocano un ruolo primario nella sua formazione:

- 📁 Accesso random alla memoria;
- 📁 Memoria indirizzata ai contenuti;
- 📁 Elaborazioni parallele;
- 📁 CRT e penna ottica;
- 📁 Procedimenti, subroutine, componenti affini ai programmi del computer;
- 📁 Strutture gerarchiche e programmi ricorsivi;
- 📁 Strutture a liste;
- 📁 Linguaggi *problem-oriented* e *procedure-oriented*;
- 📁 Unità di uscita xerografiche;
- 📁 Sistema *time-sharing* con postazione remota dell'utente.

Ciò che è importante da valutare è ad esempio il modello in cui: un uomo seduto alla scrivania, scrive o disegna su una superficie con una penna, e così un processore dovrebbe comunicare l'informazione.

1.1.1 Man's interaction with recorded knowledge

L'analisi dei concetti e dei problemi nel dominio del *precognitive system*, tratta di quattro argomenti:

- ✿ misurare le informazioni rispetto al deposito mondiale di conoscenza;
- ✿ scopi, requisiti, criteri e piani per il sistema precognitivo;
- ✿ *Schemata* di memorizzazione, organizzazione, recupero e divulgazione delle informazioni;
- ✿ Interazione uomo-macchina nel sistema precognitivo.

1.1.2 Stimare la size

Come base per pensare un sistema precognitivo, è necessario stimare quanta informazione dobbiamo trattare. I concetti - visti come unità di misura delle informazioni e ridondanza - sono delicati da stimare. La stima più semplice da valutare è il numero di caratteri alfanumerici che devono essere richiesti per spiegare chiaramente i contenuti di tutti i documenti appartenenti alle biblioteche mondiali, ogni documento "type" (che si oppone a "token", o copia individuale) deve essere considerato soltanto una volta. Ad esempio se di una stessa risorsa ho più traduzioni devo creare un nuovo documento "type"? Le varie suddivisioni, di parti del totale, sono interessanti. Va stabilito un requisito affinché il numero totale di caratteri alfanumerici sia equo semplice e diretto. Se un numero definito di "bits" viene assegnato a ogni carattere alfanumerico, è possibile moltiplicarlo per il numero di bits per carattere, in modo da poter affermare che: "c'è un numero n di bits nelle informazioni annotate nelle librerie del mondo". Durante i primi mesi, fu fatta una stima approssimativa

(Licklider, 1962), basata principalmente sul lavoro di Bourne (1961) e sulla grandezza della Library of Congress, insieme ad alcune impressioni assortite. La prima stima dava $2 \cdot 10^{14}$ caratteri oppure 10^{15} bits.

La grandezza del corpo vs la capacità e la velocità del processore di un calcolatore: uno dei più importanti prerequisiti per un'organizzazione efficace della conoscenza è quella di raggrupparla in un grande contenitore un po' alla volta.

L'accesso all'informazione richiede tempo, di solito devono essere considerati due o più metodi differenti di "tempo di accesso". Anche se, si conosce precisamente il percorso dell'informazione che vogliamo recuperare, normalmente si impiega una certa quantità di tempo per iniziare la ricerca. Dopo di che, ci si può spostare da una parola all'altra con un rapido passaggio. Da questi pensieri già emerge l'idea di *ipertesto*, il saltare di parola in parola in modo veloce e ragionato. Più avanti descrive il concetto di "accesso random": si può saltare da registro a registro in modo casuale. La velocità dell'accesso random era già conosciuta prima della Seconda Guerra Mondiale. Segue una trattazione sulla velocità del processore per accedere alla memoria.

1.1.3 Scopi, requisiti, progetti e criteri per il Precognitive System

Generalmente parlando, gli scopi del precognitive system sono di promuovere e facilitare la conoscenza tramite:

- ☞ l'acquisizione;
- ☞ l'organizzazione;
- ☞ l'utilizzo.

Acquisizione di conoscenza

L'acquisizione della conoscenza - l'iniziale apprensione per l'aumento del deposito di conoscenza - coinvolge la registrazione e la rappresentazione di eventi. Implica anche un'attività di selezione, un'analisi e organizzazione di attività relative all'inserimento di conoscenza che si aggiunge a quella già esistente.

Sia l'aspetto di acquisizione che quello di interpretazione sono riconosciuti come importanti passaggi, e giocano un ruolo attivo nell'"esperienza" e nella

“sperimentazione”. Tuttavia, anche se il concetto di interpretazione è ampiamente discusso, quello di acquisizione è escluso dal concetto attuale di biblioteca.

Pensando ad un nuovo sistema che superi talune lacune, dovremmo tenere ben presente la possibilità di sviluppare una più forte interazione tra il processo di acquisizione e quello che si occupa di trattare la conoscenza già esistente. L'idea è illustrata sinteticamente nella *Figura 1*.

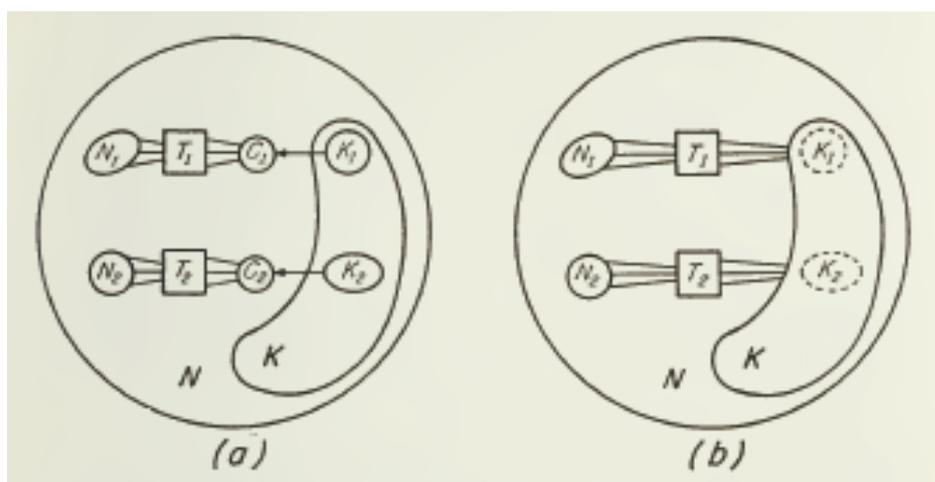


Figura 1: (a) Schematizzazione delle relazioni esistenti tra l'acquisizione della conoscenza, tramite la sperimentazione e il sistema librario. La *N* rappresenta la “Natura”, *K* identifica il corpo della conoscenza memorizzato nella biblioteca. Una piccola parte K_1 di *K* è compresa nella classe di una struttura cognitiva *C* mediante un sperimentatore che conduce un esperimento T_1 su una piccola parte N_1 di *N*. Le tre linee che connettono una figura con l'altra rappresentano un'interazione costretta soltanto dalla natura di T_1 . Quando l'esperimentatore ha raccolto e interpretato i dati, può aggiungere i risultati a K_1 . (b) La *figura 2* illustra l'eliminazione dei vincoli e delle limitazioni imposte dall'interposizione di C_1 tra T_1 e *K* del diagramma A. Gli esperimenti possono quindi interagire con l'intero *K*, ed in particolare con tutti i K_1 , usando altri canali di interazione in aggiunta a quelli forniti dal diagramma a.

Il vantaggio del diagramma b dipende dall'efficienza degli arrangiamenti per l'interazione.

Le difficoltà di integrare i risultati di molti progetti di ricerca simultanei che operano con connessioni molto sciolte l'uno dall'altro, e dal corpo di conoscenza, è attualmente oggetto di molta preoccupazione, particolarmente nel campo della farmaceutica.

Organizzazione della conoscenza

Abbiamo accennato più volte al “pozzo di conoscenza”, al “corpo della conoscenza”, e al “corpus”. Gli schemi più concreti che sono utili alla definizione dei concetti connessi con tali termini sono quelli che rappresentano stringhe di caratteri alfanumerici, i diagrammi associati, grafi, immagini, e così via, che costituiscono i documenti che vengono conservati in archivi universalmente riconosciuti. Tuttavia, tali semplici schemi, non sono di per sé sufficienti.

La neuroanatomia e la neurofisiologia, insieme al comportamento umano, forniscono schemi complementari meno definiti, ma tuttavia necessari, che arricchiscono il concetto. Tali arrangiamenti complessi di elementi e processi neurali ricevono diversi stimoli, incluse frasi parlate e scritte, e in qualche modo li processano e li memorizzano per favorire il disegno delle inferenze e le risposte alle domande. Se le risposte sono spesso imprecise, di solito sono più adeguate alle esigenze reali rispetto al mero ripristino di input passati. Quando noi parliamo di organizzare l'informazione in conoscenza, noi assumiamo un set di concetti che coinvolgono alcuni schemi.

Le materie prime, o gli input, per l' “organizzatore della conoscenza” sono i dati alfanumerici, i disegni geometrici, le immagini, le funzioni di tempo e così via. Gli outputs del sistema di organizzazione sono espressi in una o in più di una delle forme di input, ma essi non sono mere riproduzioni di traduzioni di input particolari; gli outputs sono suggerimenti, risposte a quesiti, sintesi fatta su misura di ciò che potrebbe essere un uomo se avesse una memoria più grande e precisa così da poter processare le informazioni più velocemente. I concetti del processo di organizzazione, e dell'organizzazione stessa, sono oggetto delle tematiche che si svilupperanno di seguito.

Nell'organizzazione della conoscenza, così come nell'acquisizione della conoscenza, sembrerebbe opportuno portare l'attenzione sull'intero corpus, o in ogni caso su grandi parti di esso, comunque grandi abbastanza da poter essere compresi dall'uomo.

Una parte del concetto di organizzazione, chiamata “organizzazione della memoria” si occupa della progettazione di strutture di memorie e di sistemi. Il suo scopo è quello di raggiungere due risonanze o congruenze:

1. tra la memoria e i modelli dell'informazione che possono essere memorizzati in essa;

2. tra la memoria e le richieste (per es. le domande) che possono essere rivolte ad essa.

Utilizzo della conoscenza

La conoscenza viene utilizzata nel promuovere il progresso, l'organizzazione della conoscenza, guida lo sviluppo tecnologico, e nella realizzazione della maggior parte delle attività che concernono professioni delle arti, degli affari, dell'industria e del governo. Vale a dire che il pozzo della conoscenza trova una continua e universale applicazione.

La *figura 2* suggerisce che potrebbe essere possibile trasferire la conoscenza da un pozzo generale alle applicazioni più specifiche. Tale approccio si riferisce al sistema precognitivo.

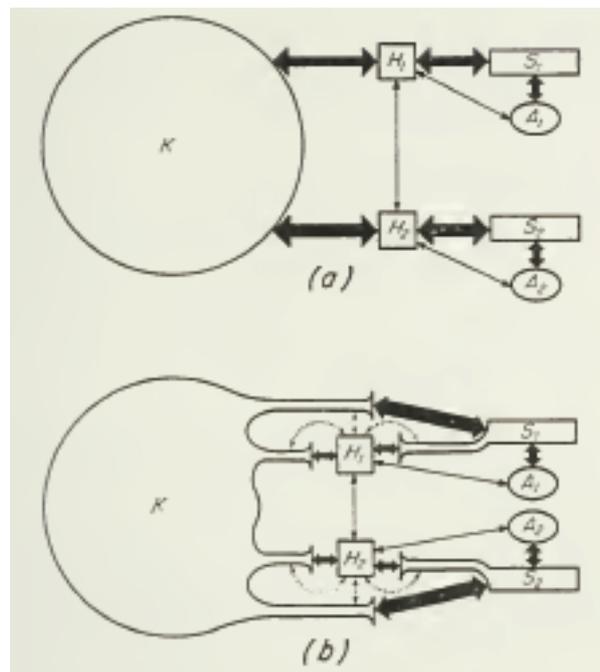


Figura 2: (a) rappresentazione schematica che illustra il flusso di informazioni ad un pozzo di conoscenza così com'è attualmente (al tempo del progetto).

(b) la situazione che potrebbe prevalere se, attraverso lo sviluppo del sistema precognitivo, il fondo di conoscenza venisse esteso fino all'interazione più profonda (rappresentata dalle proiezioni svasate e le loro interfacce) con esseri umani e i loro sistemi di applicazione. Le linee tratteggiate sono i percorsi di controllo. Piccole quantità di informazioni di controllo sono in grado di dirigere la selezione, la trasformazione, e la trasmissione di grandi quantità di informazioni sostanziali.

La struttura complessiva del precognitive system (ipotizza la Reti delle Reti):

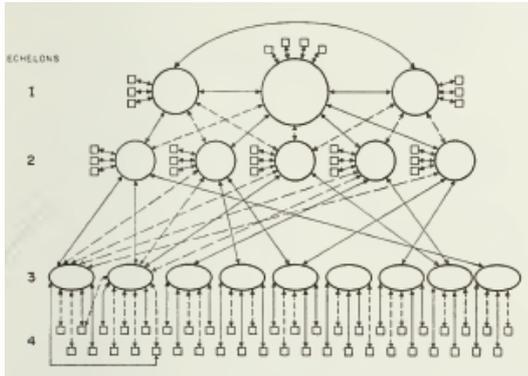


Figura 3: i cerchi e le ellissi rappresentano sistemi informatici avanzati e specialistici. I quadrati rappresentano le interfacce uomo-computer, quelli del gruppo 4 (in basso), che diventano stazioni o console per gli utenti sostanziali del sistema. La maggior parte delle connessioni sono collegamenti di telecomunicazione commutabili. Quelli indicati come linee continue rappresentano le connessioni che si potrebbero stabilire in un

dato momento durante il funzionamento. Le linee tratteggiate sono introdotte a suggerire altri collegamenti che potrebbero essere stabiliti. I cerchi del gruppo 1 si occupano in primo luogo del mantenimento del fondo di conoscenza, quelli del gruppo 2 hanno la funzione di organizzare i corpora di campi o sottocampi di conoscenza, e quelli del gruppo 3 hanno a che fare con il processamento delle richieste da parte degli utenti. L'utente posizionato nel gruppo 4, fornisce input e output (controlla e visualizza) di strutture ed il loro trattamento. Per tutti, tranne che per il gruppo I, il numero di sottosistemi previsti per il sistema di proiezioni è molto superiore al numero rappresentato in figura.

Lo schema ricorda vagamente gli utenti che si collegano alla rete, tramite personal computer, e si può intravedere uno schema progenitore del web, delle interconnessioni, dello scambio e sviluppo di conoscenza che da un'unica fonte si arricchisce e compie il passaggio di "fondo di conoscenza".

Il network ancor più simile a come lo intendiamo noi è mostrato in figura 4:

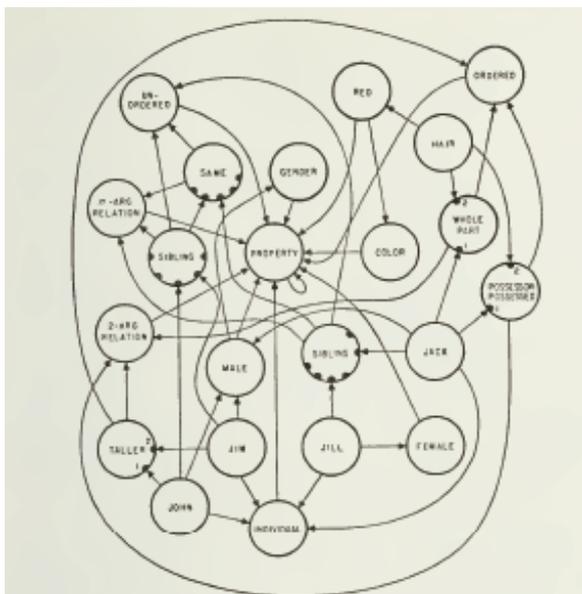


Figura 4: il diagramma rappresenta una rete relazionale. I cerchi rappresentano le entità, le relazioni e le proprietà. Entità, relazioni, e proprietà possono partecipare nelle relazioni. Gli "input finali" delle relazioni sono caratterizzati da macchie nere. Alcune interpretazioni del diagramma, relazioni che si ricavano: "John e Jim sono fratelli, e Giovanni è più alto di Jim" oppure "Jack ha i capelli rossi".

Finché in *The Computer as a Communication Device*, articolo apparso nel 1968, Lick ci offre una lucida anticipazione di ciò che abbiamo imparato a chiamare *social network*, *reti sociali*, *virtual community*, *comunità virtuali* . Nelle parole di Lick, già molto precise, “On-line interactive communities”. Persone intente a collaborare, “face to face through a computer”.

Part II

Archiviazione dei dati, i metadati e la loro standardizzazione

[C]lassification, broadly defined, is the act of organizing the universe of knowledge into some systematic order. It has been considered the most fundamental activity of the human mind.

"Cataloguing and Classification: An Introduction" Lois Mai Chan

Metadati e archivi

2.1 Scenario di riferimento

A distanza di oltre dieci anni dalla nascita del web il suo ruolo nella professione bibliotecaria e documentaria continua ad evolversi: la tecnologia è utilizzata quotidianamente e le professioni legate alla scienza dell'informazione sono strettamente legate alla necessità di creare e mantenere contenuti web. Il personale coinvolto nello sviluppo di un sito web *library-oriented* sa che l'implementazione di servizi on line richiede una conoscenza più spinta della funzionalità applicative di forme interattive, aggiungendo grafica e materiale multimediale che presuppone la conoscenza dei diversi formati tecnologici. Molte tecnologie presentano tendenze di sviluppo simili e collaborano nella realizzazione di questa strategia: i costi della memorizzazione on line, sebbene siano minori che in passato, incidono ancora sensibilmente, i *repositories* e la tecnologia su cui sono basati sono ora acquistabili e gli standard come l'Open Archives Metadata Harvesting Protocol sono operativi, mentre continua incessante l'impegno per la realizzazione di metadati sottostanti. La riflessione sulla *digital preservation* degli ultimi cinque anni ha inoltre chiarito la necessità di un approccio tecnico adeguato.

Lo sviluppo di collezioni di articoli e di riviste scientifiche liberamente accessibili on line ha mostrato come la comunicazione in ambito accademico può cambiare, alterando la disseminazione e le modalità di accesso. D'altra parte lo sviluppo di una serie di straordinari lavori digitali ha almeno suggerito il po-

tenziale specifico della creatività autoriale e del mezzo digitale nel trasformare la presentazione e la trasmissione della conoscenza.

Gli studi effettuati dall'Università di Berkley (California) sulla crescita del web mostrano che il 93% dell'informazione prodotta nel 1999 è in formato digitale¹, mentre le ricerche dell'OCLC Office Research Web Characterization Project² evidenziano le tre tendenze principali nello sviluppo del web pubblico: grandezza e crescita, internazionalizzazione e uso dei metadati. In particolare in riferimento all'utilizzo dei metadati, l'analisi del web relativa agli anni 1998-2002 (vedi *tabella 1*) ha rilevato diverse caratteristiche interessanti. Innanzitutto l'uso dei metadati è in crescita costante, soprattutto cresce la loro presenza nelle homepage, e crescite analoghe sono registrate nella percentuale di pagine web raccolte da siti pubblici che contengono una qualche forma di metadati. Va comunque tenuto presente che la comparsa di editor HTML più sofisticati permettono la creazione e l'applicazione automatica di alcuni *tags* META direttamente come parte integrante del modello di documento. Questo aspetto influisce sicuramente nella percezione della crescita della diffusione dei metadati nel web, ma è improprio pensare che vengano fornite descrizioni dettagliate delle "risorse" web.

La riluttanza da parte dei creatori di contenuti ad adottare schemi formali di metadati con cui descrivere i propri documenti è evidente: il Dublin Core appare applicato solo nello 0,5% delle home page dei siti pubblici nel 1998 e registra una piccolissima crescita (lo 0,7% nel 2002). In realtà la grande maggioranza dei metadati forniti nel web è creazione specifica e non segue alcuno schema formalizzato.

¹<<http://www2.sims.berkeley.edu/research/projects/how-much-info/charts/charts.html>>.

² <<http://www.oclc.org/research/projects/archive/wcp/>>.

Tabella 1 - Uso dei metadati nei siti web pubblici 1998-2002³.

	1998	1999	2000	2001	2002
Numero di siti web pubblici	1.457.000	2.229.000	2,94E+06	3.119.000	3,08E+06
Numero di tags % di pagine con meta-dati	70%	79%	82%	85%	87%
Media numero di tags	2,75	2,32	2,72	2,97	3,14
% di pagine con metadati	2,75%	2,32%	2,73%	2,98%	3,15%
Numero di tags	2,27	2,16	2,46	2,63	2,75

Emerge dunque il bisogno primario di organizzare e rendere *ricercabile* il web: tale obiettivo si può ottenere soltanto tramite la conoscenza della struttura dei dati, tenendo conto delle caratteristiche di varietà, non consistenza, costante evoluzione nel tempo, e perdita di univocità nel significato dei termini e dei concetti della struttura della rete globale. La soluzione è apparsa in alcune descrizioni standardizzate o in un linguaggio capace di accrescere le funzionalità (un meccanismo per una più precisa descrizione degli oggetti nel web) e nell'integrazione dei dati leggibili sia dalle macchine che dall'uomo: i *metadati*.

³ Fonte: Edward T.O'Neill - Brain F. Lavoie - Rick Bennett, Trends in the evolution of the public web, "D-Lib Magazine", April 2003, <<http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>>.

Nell'ambiente digitale, volatile per natura, i metadati descrivono risorse elettroniche, testi e documenti multimediali: essi esistono e hanno significato solo in relazione al documento (o oggetto) cui si riferiscono e di cui valorizzano le informazioni.

Attualmente esistono molti standard di metadati, emessi da gruppi formali, ossia organizzazioni nazionali ed internazionali (ISO, ANSI, NISO) e gruppi informali (W₃C Consortium, Dublin Core, Library of Congress). Questa proliferazione di schemi ha determinato confusione e, necessariamente, incompatibilità: molteplici iniziative che a fronte di questo problema avviano accordi a livello nazionale ed internazionale coinvolgendo esperti del settore. Non mancano aspetti negativi e punti deboli quali la limitatezza della prospettiva e la re-invenzione di cose esistenti. Solo di recente è emersa nel web la tendenza, a non investire ancora nella realizzazione di nuovi schemi di metadati confezionati ad hoc per esigenze delle singole comunità di utilizzatori, e ri-utilizzare il frutto di sforzi (anche finanziari) già compiuti, soprattutto grazie alle possibilità offerte dal W₃C.

I servizi bibliotecari, in particolare, hanno una lunga esperienza e tradizione nell'uso dei metadati, grazie agli OPAC (OnLine Public Access Catalogue) che hanno seguito regole di catalogazione standard. Il sistema di metadati considerato come l'archetipo è il sistema di catalogazione collegato al formato MARC (Machine Readable Cataloguing) che consentì la prima esperienza di scambio di record catalografici a livello mondiale. Molti formati tecnologici di interscambio si sono succeduti negli anni per garantire l'interoperabilità tra gli OPAC che rispondevano all'uso di software ed hardware proprietari. La necessità di mediazione tra le realtà esistenti ha portato allo sviluppo da una parte di protocolli di comunicazione come lo Z_{39.50}, che consentono l'interazione di client con molti server su cui risiedono i cataloghi, dall'altra alla creazione di schemi di metadati descrittivi per risolvere le difficoltà.

L'assegnazione di metadati si rivela una componente chiave nei processi di digitalizzazione e di editoria elettronica che implicano una scelta piuttosto ampia di metadati da selezionare e applicare. L'attribuzione dei metadati comporta del resto costi più o meno elevati, data la necessità di un'opera manuale, a cui solo una forte attività cooperativa degli utilizzatori può porre rimedio.

2.2 Definizione di biblioteca digitale

“Digital libraries are meant to provide intellectual access to stores of information, digital libraries can serve as intellectual infrastructure” (Digital Library Initiative-phase 2).⁴

La definizione di *digital library* ormai comunemente accettata vede in essa una o più collezioni di informazioni in formato elettronico gestite tramite l'uso di tecnologie informatiche per l'acquisizione, la memorizzazione, la conservazione e l'accesso, con relativi servizi associati per comunità di utenti che la utilizzano tramite supporto tecnologico. La ricerca applicata in questo campo comprende l'impatto complessivo delle tecnologie digitali e di rete sulle biblioteche e in generale sull'informazione, includendo anche l'influenza di questi sviluppi sui fornitori delle informazioni e sugli utenti.⁵

Associato al concetto di digital library è il concetto di *information retrieval* che, nel suo senso più ampio, consiste nei processi attraverso i quali l'informazione memorizzata (ed i documenti che la contengono) è resa disponibile agli utenti. In tutti i sistemi automatizzati ed informatizzati di organizzazione dei documenti e delle loro descrizioni, questo processo consente il recupero mirato e selettivo dei materiali rilevanti per le necessità informative di un utente.

Nel processo convergono da una parte dunque le operazioni biblioteconomiche catalografiche, dall'altra tecniche più sofisticate per il recupero di descrizioni, immagini e dati multimediali.

Le definizioni di biblioteca digitale che nel corso degli ultimi anni sono state elaborate ne hanno messo in primo piano aspetti diversi. In uno dei workshop

⁴ <<http://www.nsf.gov/pubs/1998/nsf9863/nsf9863.htm>>.

⁵ I programmi di ricerca su larga scala sono formalmente iniziati negli Stati Uniti più di una decina di anni fa con il programma NSF/DARPA Digital Libraries che ha prodotto un certo numero di biblioteche digitali: in questa fase la maggior parte dell'attività è consistita nella digitalizzazione di risorse esistenti creando diverse collezioni, soluzioni tecnologiche e fornendo semplici forme di accesso.

DELOS,⁶ Nicholas Belkin l'ha definita “un'istituzione che esegue e supporta le funzioni di una biblioteca nel contesto di collezioni di oggetti informativi in forma digitale distribuiti in rete”.

I significati del termine digital library spaziano quindi da quello più intuitivo di collezione digitalizzata di materiali tradizionali a quello più avanzato di collezione di informazioni esclusivamente digitali associate ai servizi che le rendono fruibili per utenti reali e potenziali.

L'espressione viene talvolta usata come sinonimo di *electronic library* e di *virtual library*: in effetti esistono elementi che accomunano tutte queste denominazioni, individuabili nei seguenti punti:

- ✿ non sono singole entità;
- ✿ richiedono tecnologie per realizzare relazioni tra molte risorse;
- ✿ le interconnessioni ed i servizi di informazione sono trasparenti per l'utente finale;
- ✿ uno degli scopi è l'accesso universale;
- ✿ le collezioni non sono limitate a surrogati di documenti ma si estendono ad oggetti digitali che possono essere rappresentati o distribuiti in formato stampa.

La Digital Library Federation statunitense fornisce una definizione che tiene conto principalmente del contesto in cui la collocazione è inserita, puntando semplicemente alla nozione di *collezione*, senza quindi fare riferimento all'organizzazione, all'accessibilità intellettuale o ai servizi correlati⁷ e ponendo in evidenza un concetto più ampio di servizio.

⁶ DELOS Network of Excellence on Digital Libraries <<http://delos-noe.iei.pi.cnr.it/>> è un'iniziativa finanziata dal Quinto Programma Quadro (IST-FP5) della Commissione Europea (Information Society Technologies) con gli obiettivi di contribuire a migliorare l'efficacia della ricerca europea in questo settore, creare un forum in cui i ricercatori e i rappresentanti di comunità di applicazione interessate e le industrie possano scambiare idee ed esperienze e i risultati di progetti finanziati dalla Comunità Europea possano essere presentati e discussi in un'ottica di cooperazione, formare giovani ricercatori in questo campo, contribuire a definire una politica di ricerca comunitaria nell'ambito delle digital libraries, cooperare con società di standardizzazione in corso e facilitare la cooperazione tra iniziative su tale argomento a livello internazionale.

⁷ “Digital Libraries are organization that provide the resources, including the specialized staff, to select, structure, offer intellectual access to interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that are readily and economically available for use by defined community”.

Una serie di attributi raccolti nel tempo danno coerenza e concretezza al concetto di digital library: le funzioni di organizzazione, mantenimento, accesso ed economia. Sulla base di questi elementi vanno definiti e misurati i programmi di sviluppo realistico delle *digital libraries*.

Nell'ambito dello statunitense Working Group on Digital Libraries Metrics, un gruppo di lavoro finalizzato allo sviluppo del consenso su un set di parametri di valutazione e comparazione dell'efficacia delle digital libraries e delle relative componenti tecnologiche in un ambiente distribuito⁸, è stato delineato che la digital library è molto di più della collezione dei materiali nei suoi *repositories*: l'accento sui servizi forniti a tutti i suoi utenti (uomini e macchine, produttori, manager e consumatori dell'informazione), dal supporto alla gestione della collezione, alla realizzazione di memorizzazione attendibile, all'ausilio nella formulazione ed esecuzione della *query* costituiscono i punti di forza di questo sviluppo.

Spingendosi indietro, fin dagli inizi dell'automazione bibliotecaria, è possibile distinguere quattro generazioni di biblioteche digitali⁹:

 la **prima generazione** (a partire dalla fine degli anni Sessanta), ossia l'era del *record bibliografico* e dei cataloghi on line, con la nascita del formato MARC e delle varie iniziative di digitalizzazione nazionali e internazionali;¹⁰

 la **seconda generazione** (dai primi anni Novanta), ossia l'era della *sperimentazione*, fortemente influenzata dalla massiccia diffusione di internet in ambiente statunitense, specie presso la Cornell University e la Library of Congress.¹¹ In questa fase la biblioteca si serve della rete come un mezzo aggiuntivo per espletare i servizi on line tramite l'uti-

⁸ <<http://www.dlib.org/metrics/public/>>.

⁹ Daniel Greenstein, *Next Generation digital libraries*, Victorian Association for Library Automation, February 2002, <www.vala.org.au/vala2002/2002pdf/01Grnstn.pdf>.

¹⁰ Smith, K. Wayne (editor), OCLC, 1967-1997: *thirty years of futhering access to the world's information*., New York, Haworth Press, 1998.

¹¹ The Library of Congree National Digital Library Program, "*American Memory*" project: *digital library users*, <<http://memory.loc.gov/ammem/dli2/html/lcndlp.html>>, Cornell University and the University of Michigan, *Making of America*, <<http://quod.lib.umich.edu/m/moagrp/>> e <<http://digital.library.cornell.edu/m/moa/>>.

lizzo di modalità di interconnessione (Telnet, ad esempio).¹² I cambiamenti non sono limitati al catalogo: è in questo periodo infatti che inizia la digitalizzazione delle raccolte e la conservazione digitale.

 la **terza generazione** (dalla metà degli anni Novanta) ossia la fase del *consolidamento* e dell'*istituzionalizzazione* in internet tramite l'ampia disponibilità di web opac. Dalla metà alla fine degli anni Novanta si moltiplicano la crescita e l'estensione dei web opac che offrono accesso all'intera gamma di collezioni della biblioteca, oltre a rendere fruibili le collezioni di altre istituzioni, a distanza in modalità remota, come nel caso del progetto europeo System for Electronic Support of Academic Material (SESAM).¹³ La penetrazione capillare di internet ha portato a valutare l'effettiva utilità ed efficacia di questo nuovo modus operandi, soprattutto rispetto al ruolo e alla struttura organizzativa dei servizi bibliotecari. Questi servizi sono supportati da sistemi, locali e globali, ognuno dei quali fornisce specifiche funzioni e interagisce con altri in modo rappresentabile tramite una struttura modulare. In questa fase storica vengono inoltre considerati gli utenti, con particolare interesse alle loro aspettative e ai comportamenti nell'uso dei sistemi informativi.¹⁴

 la **quarta generazione**, quella attuale, infine pone un'enfaticizzazione della prospettiva funzionale della biblioteca digitale vista come una componente di un sistema esteso di servizi, in cui è chiamata a mediare tra risorse informatiche diverse e distribuite e la comunità di utenti in continua evoluzione, stabilendo un ambiente di servizio consistente in uno spazio informativo on line, in cui gli utenti possono cercare, localizzare, accedere, e soprattutto utilizzare l'informazione.

¹² Commission on Preservation and Access and Research Libraries Group, *Report of the task force on archiving and digital information*, Washington, DC, 1996, <<http://www.oclc.org/community/rlg/transitions/>>.

¹³ Augusta Maria Paci, *Orientare i sistemi di informazione all'utente: gli obiettivi del progetto di ricerca europeo SESAM*, Nuovi annali della Scuola speciale per archivisti e bibliotecari, a . XII, 1998. Sito web del progetto <<http://cordis.europa.eu/libraries/en/projects/sesam.html>>.

¹⁴ Denise Covey, *Usage and usability assessment*, Washington, DC, CLIR, 2002, <<http://www.clir.org/pubs/reports/pub105/contents.html>>.

Tale ambiente supporta inoltre tutti gli aspetti amministrativi e gestionali correlati.¹⁵

Anche Howard Besser introduce una ripartizione storica nello sviluppo delle biblioteche digitali mettendo in luce tre epoche. Nelle prime realizzazioni in rete di sistemi bibliotecari attraverso il web gli utenti accedevano esclusivamente al web opac che permetteva il servizio di ricerca e recupero dell'informazione; i moduli software che implementano gli altri servizi per motivi di sicurezza non comunicavano via web. Ciò ha ben presto determinato una situazione per cui ogni web opac era accessibile tramite la propria interfaccia, ognuna caratterizzata da punti di accesso (e relativi nomi), lingua e grafica diversi.

Tabella 2- Stadi di sviluppo delle digital library

Stadio	Data	Sponsor	Scopo
I: fase sperimentale	1994	NSF/ARPA/ NASA	Esperimenti su collezioni di materiali digitalizzati
II: fase di sviluppo	1998/1999	NSF/ARPA/ NASA DLF/ CLIR	Si inizia a considerare la conservazione, la sostenibilità, le comunità di utenti
III: fase di applicazione	In corso	Finanziata attraverso canali pubblici e privati	Pieno interoperabilità tra digital library reali

¹⁵ Daniel Greenstein, *Next-generation digital libraries*, Joint Conference on Digital Libraries, July 2002, Portland, Oregon.
<<http://www.dlib.org/dlib/september02/rasmussen/ograsmussen.html>>.

Gli utenti dovevano familiarizzare con interfacce diverse, ed era impossibile effettuare ricerche incrociate in più web opac.

Una prima risposta a questo problema è stata la ricerca sperimentale e l'applicazione dello standard Z39.50, protocollo di comunicazione che consente la ricerca multidatabase su piattaforme e sistemi diversi in modo trasparente per l'utente tramite un'unica interfaccia grafica.

Questo standard ha permesso alle grandi istituzioni bibliotecarie l'implementazione web dei servizi informativi.

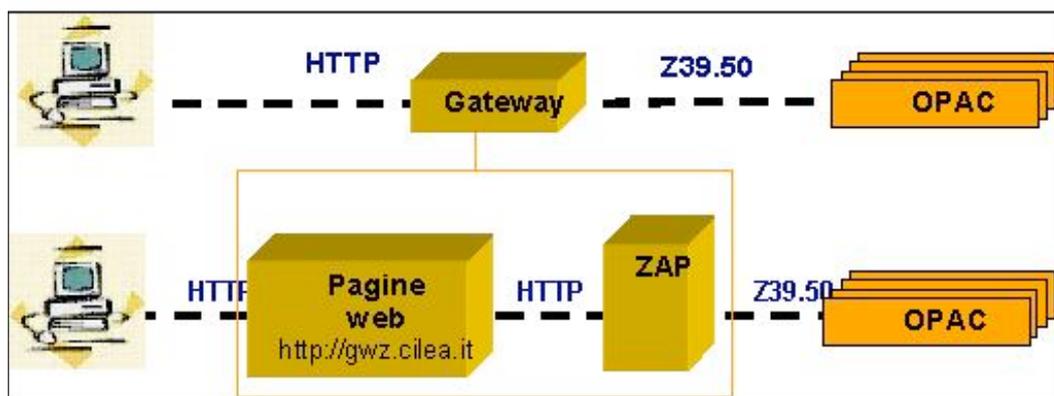


Figura 5: mostra il flusso delle operazioni eseguite dal Gateway HTTP-Z39.50 durante l'interrogazione di un OPAC Z39.50

Z39.50 è il nome più conosciuto per lo standard ISO 23950: "Information Retrieval (Z39.50): Application Service Definition and Protocol Specification", e per l'ANSI/NISO Z39.50.

Il Gateway HTTP-Z39.50:

- ✿ riceve richieste in HTTP
- ✿ traduce la richiesta in Z39.50
- ✿ manda la richiesta all'OPAC Z39.50 selezionato
- ✿ riceve la risposta in Z39.50
- ✿ traduce la risposta in HTTP
- ✿ manda la risposta al client.

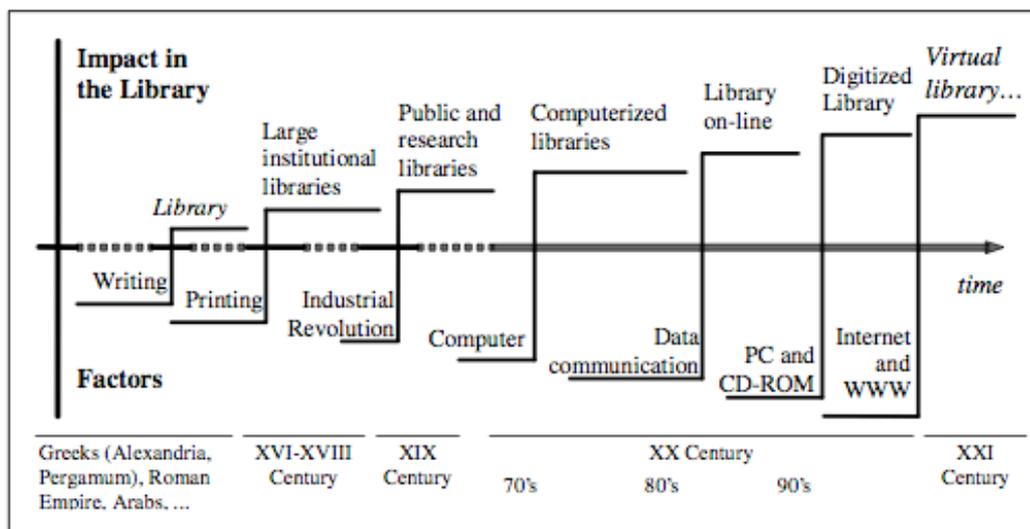


Figura 6: Biblioteche e tecnologia nel tempo, come cambiano gli strumenti e l'organizzazione in base alle scoperte tecnologiche. I cambiamenti sono molto concentrati negli ultimi cinquant'anni.

2.3 Definizione di metadati

I metadati sono definiti *informazioni sulle informazioni* o *dati strutturati sulle risorse* utilizzabili a supporto di un'ampia gamma di operazioni come descrizione e presentazione delle risorse, gestione delle risorse informative e conservazione.

Tim Beerners-Lee, l'inventore del world wide web, definisce il metadato "machine understandable information about web resources or other things"¹⁶ in cui la frase *machine understandable* è considerata un punto chiave, e si riferisce alle informazioni che gli agenti intelligenti sono in grado di utilizzare per ottimizzare le procedure di lavoro.

Si può anche pensare ai metadati come ad una sorta di riassunto formalizzato dell'informazione, riguardo alla forma e al contenuto, o, meglio ancora, come alla somma totale di ciò che si può dire riguardo ad un oggetto informativo, ad ogni livello di aggregazione.¹⁷

2.3.1 Caratteristiche ed uso

I metadati, organizzati in sistemi, rappresentano la componente logica di una raccolta.

Le risorse che i metadati descrivono possono essere elettroniche e non si limitano a risorse testuali, ma possono essere anche manufatti, materiali visivi, iconografie, file audio e altro materiale di un'eterogeneità estrema. I metadati possono essere applicati a diversi livelli, dalla descrizione globale di un'intera collezione, ad una particolare risorsa nell'ambito di una collezione o ancora ad un singolo elemento della risorsa, come un oggetto specifico o un singolo frame in un prodotto multimediale. I metadati aiutano a descrivere e strutturare in modo stabile l'informazione memorizzata su vari supporti e forniscono

¹⁶ Tim Beerners Lee, *Metadata architecture*, January 1997, updated September 2000, <<http://www.w3.org/DesignIssues/Metadata>>.

¹⁷ A. J. Gilliland-Swetland, *Setting the stage. Introduction to metadata*, Pathways to Digital Information, May 2000, <<http://www.slis.kent.edu/~mzeng/metadata/Gilland.pdf>>.

il contesto e l'accesso facilitato¹⁸, recupero e comprensione dell'informazione nel tempo superando il problema dei cambiamenti tecnologici. Come hanno notato Weiber e Lagoze, due dei principali studiosi del settore, i metadati permettono di accedere ad un contesto surrogato distinto dall'accesso alla risorsa. Qualsiasi ambiente che produca o gestisca considerevoli quantità di documentazione ha buoni motivi per utilizzare i metadati: nell'ambiente Internet i set di metadati sono stati sviluppati per selezionare la grande massa di informazioni disponibili e per migliorarne l'accesso. L'adozione su larga scala di standard descrittivi per le risorse elettroniche migliora il recupero di informazione strutturata nei contesti della ricerca; i metadati in effetti aiutano ad organizzare le risorse elettroniche, promuovono la loro interoperabilità, verificano l'identificazione e assicurano la conservazione a lungo termine.

2.3.2 Tipologie di metadati

Per rispondere alle varie necessità di gestione sono stati sviluppati diversi tipi di metadati.

I metadati descrittivi sono usati per l'identificazione, la scoperta, l'accesso e la valutazione delle risorse.

Quelli amministrativi, di mantenimento dei record (*recordkeeping*), hanno la funzione di ordinare, valicare, archiviare le risorse per assicurare il significato, la gestibilità, e la longevità dei record e dell'informazione in essi contenuta¹⁹.

I metadati per la conservazione contribuiscono alla memorizzazione a lungo termine delle risorse digitali.

Queste tipologie tendono ad aumentare perché emergono nuove necessità descrittive e gestionali.

Fino alla metà degli anni Novanta il termine *metadata* era usato prevalentemente dagli informatici coinvolti nella gestione e nell'interoperabilità di basi di dati geo-spaziali, per riferirsi ad un insieme di standard industriali o disci-

¹⁸ La National Library of Australia sostiene che la principale ragione d'essere dei metadati è proprio quella di migliorare la gestione e il recupero dell'informazione: <<http://www.nla.gov.au/librariesaustralia/metadata.html>>.

¹⁹ New South Wales *recordkeeping metadata standard*, September 2000, <<http://www.records.nsw.gov.au/recordkeeping/dirks/doing-a-dirks-project/specify-and-apply-recordkeeping-metadata/#what-is-recordkeeping-metadata>>.

plinari, a documentazione interna o esterna e agli elementi necessari per l'identificazione, la rappresentazione, la performance e l'uso di informazioni di un sistema informativo in cui un oggetto, semplice o composto, può essere manipolato da un essere umano o da un sistema come un'entità discreta.

Risale al 1995 il **Metadata workshop** organizzato da Stuart Weibel di OCLC in cui si prese la storica decisione di definire un modello di metadati Dublin Core per la comunità bibliotecaria.²⁰

Da quel nucleo primario di significato, i confini si sono estesi per rispondere alle necessità di informazione descrittiva standardizzata. Ha anche aggiunto un livello di collegamento logico con le catalogazioni dei documenti pregressi: per tale motivo i cataloghi cartacei della biblioteca, gli archivi e la documentazione catalografica dei musei vengono associati a strutture che contengono i metadati.

In generale gli ambiti di applicazione dei metadati agli oggetti informativi, indipendentemente dalla loro forma fisica o intellettuale, presentano tre caratteristiche:

-  contenuto
-  contesto
-  forma

In particolare ognuna di queste può essere gestita attraverso i metadati che esprimono la qualità intrinseca o estrinseca della caratteristica nel singolo oggetto:

-  il *contenuto* si riferisce a ciò che l'oggetto contiene ed è intrinseco;
-  il *contesto* indica l'ambiente associato alla creazione ed è estrinseco;
-  la *struttura* è relativa al set formale di associazioni all'interno dell'oggetto o tra oggetti informativi, e può essere intrinseca o estrinseca.

Le istituzioni archivistiche e museali hanno tradizionalmente posto molta attenzione al contesto in quanto permette di identificare e preservare il valore assoluto nel tempo, facilitando l'autenticazione degli oggetti e assistendo i ricercatori nelle analisi e nelle interpretazioni.

²⁰ DC-I: CLC/NCSA Metadata workshop. *The essential elements of network object description*, <<http://dublincore.org/workshops/dci/>>.

Lo sviluppo dei metadati ha posto meno attenzione alla struttura degli oggetti informativi il cui ruolo è potenziato dalla contemporanea crescita delle caratteristiche tecnologiche e delle attrezzature informatiche.

Le comunità degli specialisti dell'informazione, bibliotecari e documentalisti, sono diventate invece sempre più consapevoli del fatto che maggiore è l'attenzione al livello strutturale di una risorsa, ovvero ai metadati di contenuto, migliore è il rendimento qualitativo delle procedure di ricerca, manipolazione e interrelazione con altri oggetti informativi: per catturare, documentare e rinforzare tale struttura è necessario sviluppare tipi specifici di metadati.

La tipologia di metadati relativi all'amministrazione è fondamentale per i *repositories* che garantiscono accesso, conservazione e uso delle collezioni.

Nelle *digital libraries* in cui convergono musei virtuali, raccolte bibliotecarie e archivi automatizzati, si trovano disponibili insieme copie digitali e relativi metadati.

Per l'amministrazione delle collezioni e nel mantenimento dell'integrità intellettuale presente e futura è importante incorporare altri tipi di metadati in tali sistemi, come afferma Paul Conway, che allude a questa capacità quando discute dell'impatto della digitalizzazione sulla conservazioni, sostenendo che il mondo digitale trasforma i concetti tradizionali di conservazioni dalla protezione dell'integrità fisica dell'oggetto alla specificazione della creazione e del mantenimento di esso, di cui l'integrità intellettuale è primaria.²¹

2.3.3 Categorizzare i metadati

Dalla letteratura internazionale emergono metodi diversi di approccio alle categorizzazione dei metadati.

Un'indagine condotta nell'ambito del progetto DESIRE²² nel 1996 individua oltre venti tipologie di metadati diverse per formati, segnale di una *meta-diver-*

²¹ Paul Conway, *Preservation in the digital world*, Washington, DC, Commission on Preservation and Access, 1996, <<http://www.clir.org/pubs/reports/conway2/>> "The digital world transforms traditional preservation concepts from protecting physical integrity of the object to specifying the creation and maintenance of the object whose intellectual integrity is its primary characteristic".

²² <<http://www.desire.org>>.

sità²³, termine usato anche da Clifford Lynch nelle sue numerose indagini relative alle *digital libraries*.

Di norma vengono definiti tre livelli di distinzione:

 tipologico-strutturale

 funzionale

 di comunità

Antonella De Robbio²⁴ nell'intervento al convegno *La biblioteca ibrida: verso un servizio informativo integrato* svoltosi a Milano nel marzo 2002, ha proposto una visione integrata molto interessante delle tipologie di metadati schematizzate da Martin Dillon²⁵, Locarn Dempsey e Rachel Heery²⁶ e da Micheal Day²⁷ riassumendo e comparando gli approcci più importanti su questo tema.

²³ Clyfford Lynch, *Building digital libraries for the metadiversity: federation across disciplines*, *Preprints of the II Metadiversity Conference Proceedings*, Charleston, SC, June 2002.

²⁴ Antonella De Robbio, *Metadati: parola chiave per l'accesso alla biblioteca ibrida*, in *La biblioteca ibrida: verso un servizio informativo integrato*, Milano, Editrice Bibliografica, 2003, pp. 103-128.

²⁵ Martin Dillon, *Metadata for web resources: how metadata works on the web*, Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenge of Networked Resources and the Web, Novembre 2000, <http://www.loc.gov/catdir/bibcontrol/dillon_paper.html>.

²⁶ Locarn Dempsey - Rachel Heery, *Specification for resources description methods. Part 1: A review of Metadata: a survey of current resources description formats, deliverable A review of metadata - Development of a European Service for Information on Research and Education*, March 1997, <<http://www.ukoln.ac.uk/metadata/desire/overview/>>.

²⁷ Micheal Day, *The metadata challenge for libraries: a view from Europe*, *Metadiversity - A Call to Action. Responding to the Grand Challenge for Biodiversity Information Management through Metadata*, Natural Bridge, Virginia, November 1998, <<http://www.ukoln.ac.uk/metadata/presentations/metadiversity/challenge.html>>.

Tabella 3 -Differenti tipi di metadati e loro funzioni

Tipo	Definizione	Esempi
Amministrativi	Metadati usati nella gestione e nell'amministrazione delle risorse informative	<ul style="list-style-type: none"> - Informazioni sull'acquisizione - Diritti e modalità di riproduzione - Documentazione di requisiti per l'accesso legale - Informazioni sulla localizzazione - Criteri di selezione per la digitalizzazione - Controllo delle versioni e differenziazione tra oggetti informativi simili
Descrittivi	Metadati usati per descrivere o identificare le risorse informative	<ul style="list-style-type: none"> - Record di catalogazione - Sussidi per il ritrovamento - Indici specializzati - Relazioni di interconnessione tra le risorse
Strutturali	Metadati relativi alla gestione della conservazione delle risorse informative	<ul style="list-style-type: none"> - Documentazione sulle condizioni fisiche delle risorse - Documentazione sulle azioni intraprese per conservare le versioni fisiche e digitali delle risorse (per es. refreshing e migrazione)
Tecnici	Metadati relativi a come un sistema funziona e a quali metadati comporta	<ul style="list-style-type: none"> - Documentazione hardware e software - Informazioni sulla digitalizzazione (per es. formati, entità della compressione, procedure di scaling) - Traccia dei temi di reazione del sistema - Autenticazione e sicurezza dei dati (per es. chiavi di criptazione, password)

Tipo	Definizione	Esempi
Di uso	Metadati relativi al livello e al tipo di uso delle risorse informative	<ul style="list-style-type: none"> - Esposizione - Traccia di uso e utente - Ri-uso del contenuto e informazioni sul multi-versioning

Fonte: *Introduction to metadata: pathways to digital information*, a cura di Murtha Baca, Getty Standards Program, Introduction to Metadata, <http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html>.

Tra i due estremi proponiamo la categorizzazione dei metadata in cinque tipologie relative alle funzionalità amministrative, descrittive, conservative, tecniche ed uso, fornendo esempi delle funzioni comuni che ognuna di essa può realizzare in un sistema informativo digitale.

La complessa attività associata alla varie tipologie e descritta estesamente in letteratura è stata sintetizzata allo scopo di rappresentare i vari aspetti, attributi e caratteristiche particolari che la generazione di metadata comporta, facendo riferimento al testo di Murtha Baca che funge da linea guida di riferimento per applicazioni nei vari contesti. Questa risorsa consente un'efficace trasferimento delle conoscenze, costituisce una base di ispirazione e richiede tuttavia di essere calata ed interpretata sotto il profilo culturale nonché tecnologico ai fini di nuove concezioni catalografiche da applicare alle varie realtà nazionali.

Tabella 4 - Attributi e caratteristiche dei metadata

Attributo	Caratteristiche	Esempi
Fonti di metadata	Metadati interni generati da un agente creativo per un oggetto informativo nel momento in cui questo viene creato o digitalizzato	<ul style="list-style-type: none"> - Nomi di file e informazioni di intestazione - Strutture di directory - Formato dei file e schemi di compressione
	Metadati esterni relativi ad un oggetto informativo creato successivamente da qualcuno che non è il creatore originario	<ul style="list-style-type: none"> - Record di registrazione e catalogazione - Informazioni sul copyright ed altri aspetti legali

Attributo	Caratteristiche	Esempi
Metodo di creazione dei metadati	Metadati generati automaticamente da un computer	<ul style="list-style-type: none"> - Indici di parole chiave - Log di transazione utenti
	Metadati creati manualmente da una persona	<ul style="list-style-type: none"> - Surrogati descrittivi come record di catalogo e metadati Dublin core - XML, RDF ecc.
Natura dei metadati	Di base: metadati creati da persone che non sono specialisti dell'informazione, spesso i creatori dell'oggetto informativo	<ul style="list-style-type: none"> - Metatag create per una pagina web personale - Filing system personali - Composizione tramite maschere di inserimento predefinite
	Esperti: metadati creati da specialisti dell'informazione, spesso non coincidenti con i creatori dell'oggetto informativo	<ul style="list-style-type: none"> - Intestazioni per soggetto specializzate - Record MARC - Sussidi di ricerca archivistici
Stato	Metadati statici che non cambiano una volta che sono stati creati	<ul style="list-style-type: none"> - Titolo, provenienza e data di creazione di una risorsa informativa
	Metadati dinamici che possono cambiare in base all'uso e alla manipolazione dell'oggetto informativo	<ul style="list-style-type: none"> - Struttura della directory - Log di transazione utenti - Risoluzione delle immagini
	Metadati a lungo termine necessari per assicurare continuità di accesso ed uso dell'oggetto informativo	<ul style="list-style-type: none"> - Formato tecnico e informazioni sull'elaborazione - Informazioni sul copyright - Documentazione di gestione della conservazione
	Metadati a breve termine	<ul style="list-style-type: none"> - Relativi a transazioni
Struttura	Metadati strutturali che sono conformi ad una determinata struttura, standardizzata o meno	<ul style="list-style-type: none"> - Dublin Core - MARC - TEI ed EAD - Formati di database locali

Attributo	Caratteristiche	Esempi
	Metadati non strutturati che non sono conformi a una struttura prefissata	- Campi note non strutturati e annotazioni
Semantica	Metadati controllati conformi ad un vocabolario o a una forma di authority	- AACR2
	Metadati non controllati e non conformi ad un vocabolario standardizzato o una forma di authority	- Note a testo libero - Metatag HTML - Tag XML senza DTD riconosciuta
Livello	Metadati di collezione relativi a collezioni di oggetti informativi	- Record a livello di collezione, per es. record MARC o sussidi per la ricerca - Indici specializzati
	Metadati di item relativi a singoli oggetti informativi spesso contenuti in collezioni	- Caption trascritte di immagini - Informazioni sul formato

Fonte: *Introduction to metadata: pathways to digital information*, a cura di Murtha Baca, Getty Standards Program, Introduction to Metadata, <http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html>.

La generazione e la gestione dei metadati sono divenute un complesso mix di processi e livelli manuali e automatici. Diverse funzioni e peculiarità di disseminazione e fruizione degli oggetti informativi digitali presentano diverse fasi di trattamento informativo.

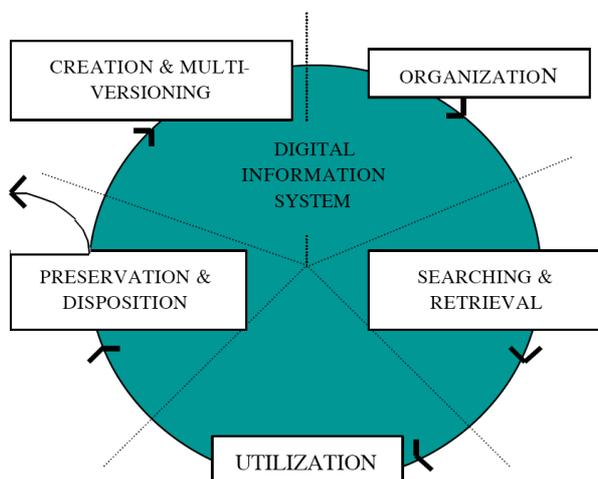


Figura 7: ciclo di vita di un information object in un sistema informativo digitale.

La *figura 7* mostra quali sono le trasformazioni che un oggetto informativo attraversa in un ambiente digitale. Via via acquisisce livelli di metadati associati in modi diversi: incapsulati nell'oggetto stesso (come intestazione per un file di immagine, per esempio) oppure allegati tramite puntatori bidirezionali. La relazione tra i metadati e l'oggetto informativo e tra i diversi aspetti dei metadati è frutto di un'operazione di registrazione esterna a questo processo in un ambiente separato di gestione dei metadati.

Gli oggetti entrano in un sistema informativo in forma digitale nativa o attraverso la conversione da altro formato e versioni diverse dello stesso oggetto possono essere utilizzate per la conservazione, la ricerca, la disseminazione o per altri sviluppi (*creation and multi-versioning*). L'autore può inserire in questo caso alcuni metadati amministrativi. Gli oggetti vengono organizzati manualmente o automaticamente nella struttura del sistema informativo digitale e vengono creati metadati aggiuntivi tramite la registrazione, la catalogazione ed i processi di indicizzazione (*organization*) finalizzati alla ricerca e al recupero da parte degli utenti. Il sistema crea metadati relativi agli algoritmi di recupero, alle transazioni degli utenti e all'efficacia (*searching and retrieval*). Una volta recuperati, gli oggetti vengono utilizzati, riprodotti, modificati: per monitorare tali processi sono utili metadati relativi alle annotazioni degli utenti, ai diritti e al controllo delle versioni (*utilization*). Infine gli oggetti subiscono processi di *refreshing*, migrazione e controlli di integrità che ne assicurano la disponibilità continua: i metadati documentano sia la conservazione sia la disponibilità (*preservation and disposition*).

2.3.4 Etichette (Tags)

I metadati sono di norma espressi come gruppi di elementi e attributi obbligatori o opzionali cui è associata una procedura di etichettatura: *tagging*. Il contenuto degli attributi ha a che vedere con le funzioni per i cui i metadati sono usati.

Dillon²⁸ suddivide i metadati in:

* *etichette o tags* che presentano le risorse.

Le etichette definiscono il contenuto dei campi, si tratta di metadati generici e semplici, non strutturati, incapsulati nella risorsa che descrivono e sono prodotti direttamente dal linguaggio di codifica. Esempi di formati con struttura semplice sono: Dublin Core, IAFA templates, ROADS templates.

Questi metadati sono importanti perché contengono una descrizione dettagliata al punto tale che l'utente può valutare se si trova di fronte una risorsa di reale interesse, e sono inoltre riconoscibili dai motori di ricerca che leggono all'interno dei codici sorgenti. Non gestiscono le relazioni tra gli oggetti.

* *descrittori* che descrivono le risorse.

I descrittori sono quei metadati che si riferiscono ai vari modelli descrittivi di risorse elettroniche, compreso il Dublin Core. Visto come componente, lo standard Dublin Core può essere creato automaticamente in modo indipendente dal documento che rappresenta. Un esempio importante è stato applicato a **CORC (Cooperative On Line Resource Catalog)** di OCLC, un ambiente di reference che ha permesso di creare descrizioni di materiali secondo le logiche dei due formati principali di metadati esistenti (MARC e Dublin Core) condivisi dalla comunità internazionale.

La produzione di metadati avviene con una conversione automatica da un formato all'altro.

2.3.5 Gli SCHEMAS

Con il rapido sviluppo del web ed il consolidamento dei nuovi schemi di metadati nei diversi domini, gli sviluppatori di servizi di informazione sul web stanno facendo i conti con il problema che ovunque i metadati siano necessari, è fondamentale che gli schemi devono essere il più possibile compatibili con gli schemi esistenti.

I requisiti di un registro di metadati consistono:

✦ definizione controllata dello schema;

²⁸ Martin Dillon, *Metadata for web resource*,
<http://www.loc.gov/catdir/bibcontrol/dillon_paper.html>.

- ✦ controllo dei cambiamenti e dell'evoluzione del linguaggio dei metadati;
- ✦ dichiarazione delle relazioni tra diversi schemi di metadati;
- ✦ definizione di *best practice*;
- ✦ promozione degli schemi esistenti per evitare duplicati inutili e costosi.

I registri di metadati ed i loro sistemi di mantenimento stanno conoscendo uno sviluppo incrementale perchè è sempre più forte l'esigenza di documentare l'evoluzione degli schemi e di allertare gli sviluppatori sui cambiamenti di versione.

Lo sviluppo del web e dei sistemi informativi digitali in rete fornisce molte opportunità ai professionisti dell'informazione ma richiede al contempo di affrontare problemi quali:

- ✦ accrescere l'accessibilità;
- ✦ mantenere le relazioni degli oggetti digitali con i rispettivi contesti all'interno delle collezioni (non creare decontestualizzazione);
- ✦ disseminare l'uso dei sistemi informativi digitali (con scalabilità dell'accessibilità intellettuale per comunità di utenze diverse);
- ✦ gestire la coesistenza di versioni multiple di oggetti digitali;
- ✦ gestire aspetti legali relativi alla proprietà intellettuale e alle modalità di riproduzione;
- ✦ affrontare il problema della conservazione dell'informazione digitale (refreshing e migrazione);
- ✦ valutare i miglioramenti e le economie dei sistemi tramite tecniche di *benchmarking* (molte dei quali possono essere raccolti automaticamente dal computer per la valutazione dell'efficacia del sistema).

Gli implementatori si accorgono spesso di aver bisogno di alcuni elementi chiave, spesso qualificati, insieme a parti di *set di elementi* di un dominio specifico oppure si trovano di fronte alla necessità di inventare nuovi elementi per casi particolari che non sono oggetto di standardizzazione.

Idealmente essi devono definire degli schemi che rispondono alle necessità locali di un progetto o servizio con un livello di precisione adeguato, tenendo conto degli standard e dei formati esistenti in modo da rendere le risorse di cui si occupano integrabili con il web , quindi *interoperabili*. Questo aspetto ha

attirato l'attenzione della ricerca che ha dedicato una parte del lavoro all'analisi del significato di questa necessità per elaborare un proprio punto di vista capace di cogliere le ricadute culturali e informative di questa problematica.

A questo scopo culturale il lavoro di ricerca si è interessato al progetto SCHEMAS,²⁹ un forum per implementatori di schemi di metadati che, attraverso il monitoraggio delle iniziative in questo campo in cui sono coinvolte organizzazioni che partecipano a progetti finanziati dalla Commissione Europea (Programma IST) e di analoghe iniziative a livello nazionale, funge da orientamento nel panorama confuso di nuovi modelli descrittivi emergenti nel web.³⁰ Analogamente è interessante il lavoro svolto da UKOLN, UK Office for Library and Information Networking, il centro inglese per eccellenza per le biblioteche digitali e le relative tecnologie presso l'Università di Bath, il quale ospita il database di schemi di metadati, il "Registro di metadati", costituito da collegamenti agli elementi e alle loro definizioni in qualsiasi forma essi siano disponibili.³¹

Un altro esempio britannico è DESIRE Metadata Registry, progetto per lo sviluppo di un servizio europeo per l'informazione sulla ricerca e l'istruzione, all'interno del quale UKOLN sta costruendo una demo di registro di metadati per indagare sulla funzionalità di un uso controllato di metadati a vari livelli: definizione degli elementi e loro uso, schemi permessi e mappature ad altri

²⁹ SCHEMAS, iniziato nel febbraio 2000 e terminato nel dicembre 2001 aveva anche un'attività di osservatorio SCHEMAS Metadata Watch (MD Watch) che forniva, tramite report quadriennali un panorama a livello mondiale dei progressi nel campo dei metadati <<http://www.schemas-forum.org/>>.

³⁰ Molti di questi progetti in realtà necessitano di schemi personalizzati sebbene interoperabili: il supporto di SCHEMAS consiste proprio nel migliorare l'accesso ai risultati di progetti europei contribuendo alla standardizzazione degli schemi di metadati adottati, attività che ha una ricaduta anche nell'ambito del CEN/ISSS (CENN/ISSS - Information Society Standardization System), <<http://www.cen.eu/cenorm/homepage.htm>>.

³¹ <<http://www.ukoln.ac.uk/projects/iemsr/>>. Tali schemi comprendono sia standard riconosciuti sia schemi di progetti specifici pubblicati in modo meno formale, come ad es. tramite pagine web, schemi RDF ecc. Lo **standard ISO/IEC 11179 MetaModel** è un modello per gli elementi di un'architettura informativa progettato per gestire sia gli elementi individuali dell'informazione, sia i gruppi di elementi tramite gli schemi di classificazione che possono essere usati per raggruppare gli elementi e i valori associati ad uno standard di dati, un sistema informativo o una schema XML. Inoltre, attraverso un'organizzazione gerarchica di schemi di classificazione, gli elementi e i valori numerati possono essere organizzati per rappresentare la struttura di un database o di uno schema XML.

namespaces, cioè i nomi degli elementi. Questo registro dei metadati è evoluto rispetto ai precedenti in quanto integra la funzione di mappatura.

Diversi vocabolari di metadati³², tramite tabelle di conversione codificate, dettano le relazioni tra gli elementi in un vocabolario di origine ed uno di destinazione. Tale approccio innovativo invece di fare le mappature tra coppie di vocabolari, mappa ogni vocabolario su uno strato semantico intermedio che media semanticamente il vocabolario di metadati A con il vocabolario B.

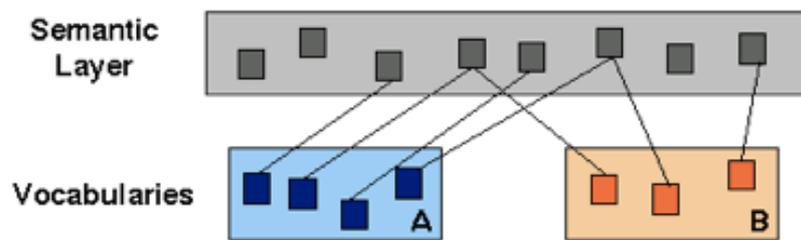


Figura 8: come opera la mappatura di un vocabolario A e B che si relazionano allo strato semantico.

Vantaggio risultante è che invece di dover creare mappature tra ogni coppia di vocabolari, è necessario soltanto fare il mapping tra ogni vocabolario e il livello semantico. Questo significa che se introduco un nuovo vocabolario C, in un registro che possiede già venti vocabolari, aggiungo un unico mapping (tra il nuovo vocabolario e lo strato semantico), piuttosto che eseguire venti mappature. Poiché le mappature non sono fatte manualmente, questo approccio in realtà si può rivelare molto limitato nella qualità delle mappature generate automaticamente con perdita di informazione. La soluzione risiede nella creazione di uno strato semantico molto dettagliato e complesso in cui gli elementi dei vocabolari possono essere spiegati con precisione. In DESIRE i *namespaces* vengono registrati e ad essi sono associati gli elementi appropriati: i quindici elementi dello schema Dublin Core, ad esempio, possono essere registrati come appartenenti al *namespaces Dublin Core* (Figura 9 è illustrato un esempio).

³² <<http://desire.ukoln.ac.uk/registry/crosswalk.php>>.

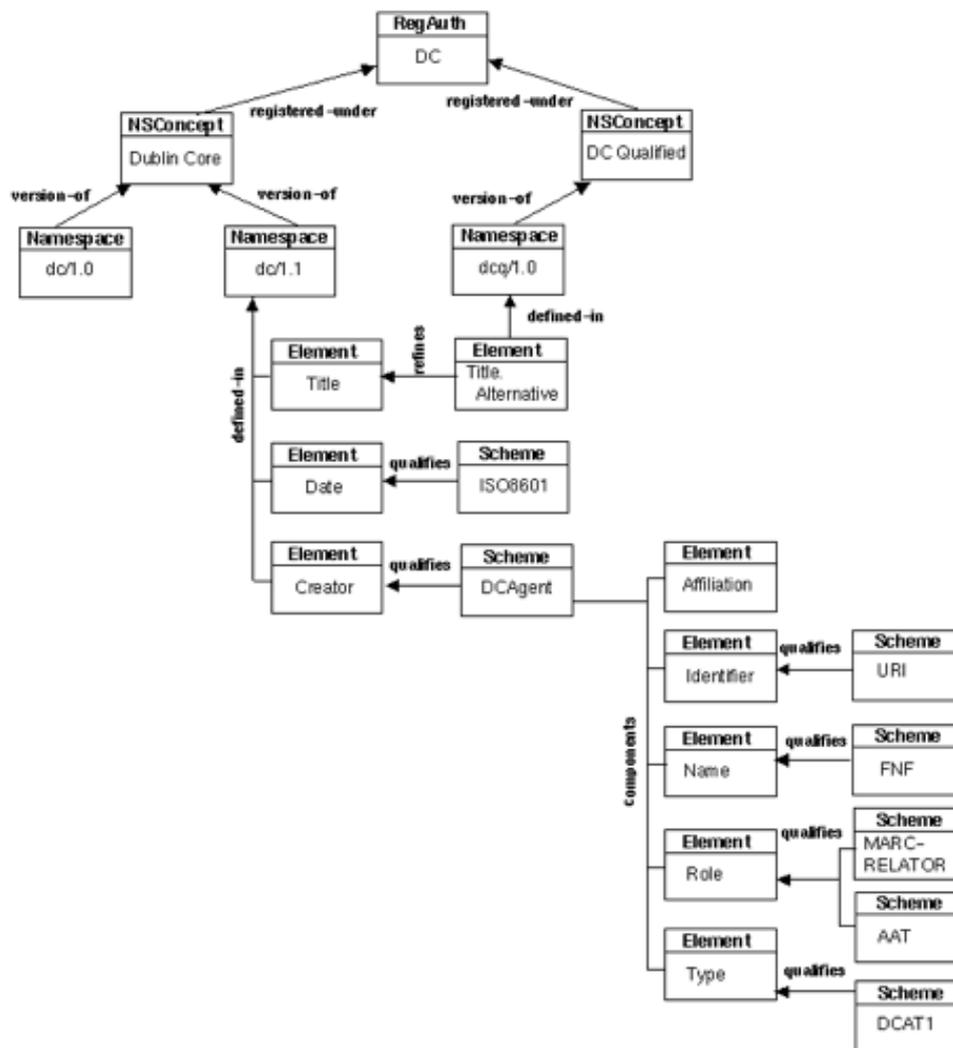


Figura 9: architettura di DESIRE, si crea lo SCHEMA di Dublin Core, in questa immagine sono rappresentati tre elementi (*Title*, *Date*, *Creator*).

Gli elementi associati ad un particolare *namespace* formano il vocabolario, con una struttura in cui è previsto anche l'aggiornamento e le nuove versioni degli schemi.

Attualmente la ricerca scientifica internazionale sta quindi investendo proprio su questo nucleo della mappatura: l'attività di SCHEMAS, conclusosi nel 2001, e di DESIRE concluso nel 2002, è integrata in CORES, un progetto mirato alla definizione di un *data model* per la dichiarazione e la condivisione della semantica dei metadati in un linguaggio comprensibile alla macchina.

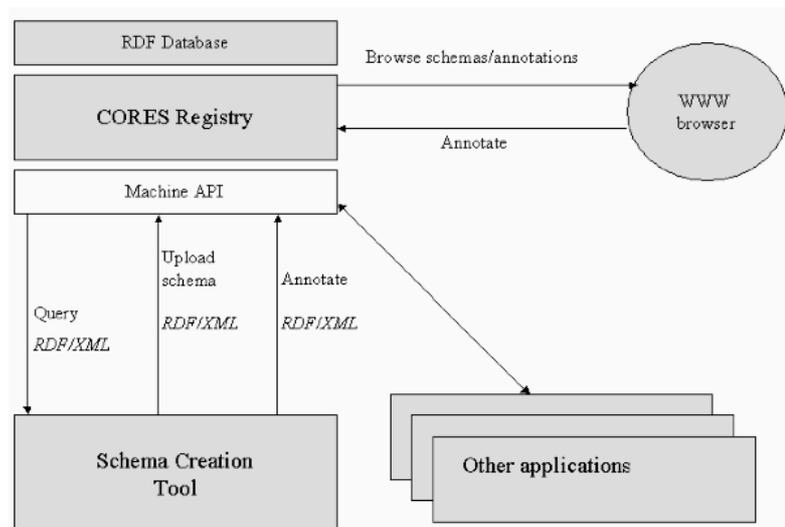


Figura 10: architettura di CORES, progetto che esplora le potenzialità degli schema di metadati e del loro ri-uso nella tecnologia del Semantic Web.

Come hanno illustrato Rachel Heery e Pete Johnston alla Dublin Core Conference 2003³³ di Seattle, i problemi sono legati alla promozione di servizi per l'individuazione e la navigazione di set di metadati in un contesto in cui si moltiplicano le applicazioni. Il progetto ha individuato la necessità di potenziare il registro per condividere le informazioni riguardo agli schemi di metadati. Il registro CORES ed il relativo strumento per la creazione di schemi è influenzato da due fattori principali:

- ❖ dai principi del Dublin Core Metadata Element Set, in particolare dai concetti di rifinitura degli elementi e codifica degli schemi;
- ❖ dall'idea di applicazioni e personalizzazioni specifiche catturabili come profilo da implementare.

Il tool per la definizione di set di elementi, profili e applicazione e schemi di codifica è scaricabile on line e fornisce, in modalità standalone o interattivamente con il registro, un'interfaccia per la creazione e l'edizione di schemi da memorizzare localmente e/o da sottoporre al registro CORES (solo però se si hanno i diritti di accesso necessari).

³³ Rachel Heery - Pete Johnston, *Metadata schema registries in the partially semantic web: the CORES experience*, DC2003, Seattle, USA, 28 September - 2 October 2003.

Gli sviluppi nel campo dei registri dei metadati sono collegati all'ampio lavoro in corso nel web semantico che include campi di ricerca quali le ontologie, i *web services*, che possono apportare miglioramenti notevoli alla gestione condivisa degli schemi di metadati.

2.3.6 Modelli di metadati

I metadati rappresentano un linguaggio di codifica, aggiungono affermazioni riguardo alle risorse: le frasi di questo linguaggio sono definite in standard, la grammatica ed i modelli dei dati stabiliscono quali informazioni possono aggiungere.

Come tutti i linguaggi, anche questo è destinato ad evolversi con l'uso a causa di diversi fattori: i linguaggi sono inevitabilmente restii ad ogni stabilità, la gente deforma le definizioni ufficiali, gli implementatori equivocano il significato inteso o l'uso degli elementi e coniano termini locali ed estensioni. Inoltre, se l'applicazione non si adatta allo standard, è questo che viene personalizzato per adattarlo all'applicazione. Da ciò scaturisce l'esigenza di gestire i vocabolari delle diverse comunità dell'informazione:

- ✿ mondo bibliotecario (vocabolari controllati, *thesauri*, classificazioni);
- ✿ mondo delle aziende (tassonomie, tipi di prodotti);
- ✿ mondo delle ricerca (ontologie di concetti, *topic maps*).

Nel linguaggio dei metadati fondamentale mancano:

- 🌹 dei dizionari completi dove trovare una panoramica dei termini usati nei vari linguaggi;
- 🌹 un contesto di pubblicazione per gli implementatori dove sia possibile vedere come i fornitori di informazioni stanno usando i metadati;
- 🌹 grammatiche standard per comprendere come i modelli di metadati condividono principi comuni.

Questa problematica, sin dal suo nascere, è stata affrontata e supportata dalla Library of Congress e dalle attività delle varie istituzioni statunitensi che hanno creato una dimensione informativa su questa tematica catalizzando gli sforzi ed i risultati raggiunti dalle esperienze sperimentali in corso, anche italiane come il *Gruppo di lavoro sui metadati dell'Istituto centrale per il catalogo unico*,

operativo dal 2000 al 2002 con lo scopo di coordinare a livello nazionale le implementazioni di metadati nei progetti di digitalizzazione nei diversi settori dei beni culturali anche tramite osservatorio sulle applicazioni di metadati nelle diverse istituzioni italiane.³⁴

2.3.7 Metadati per domini

L'Osservatorio dei metadati del progetto SCHEMAS ha individuato sei domini in cui ha condotto la sua attività di monitoraggio:³⁵

- 🌻 *audiovisual* (film industry, broadcast production and archiving, multimedia production);
- 🌻 *cultural heritage* (libraries, library service, museums, museum portals, archives);
- 🌻 *education* (interactive courseware, life-long, learning, distance learning, schools, curricula);
- 🌻 *government* (e-Government initiative, public sector information);
- 🌻 *publishing* (book trade and distribution, music/video/multimedia/distribution, music recording industry, scientific journals, news agencies, newspapers, copyright management)
- 🌻 *other* (geospatial information, industry, environmental information).

Il settore multimediale è dominato dal Moving Picture Experts Group³⁶ con l'emergere recente degli standard MPEG4 (per la codifica e l'interazione con oggetti audiovisivi), MPEG7 (interfaccia per la descrizione del contenuto multimediale degli stessi), MPEG2I, un'ambiziosa cornice concettuale per descrivere l'infrastruttura di recupero e di consumo del contenuto multimediale.

Lo standard utilizzato per collezioni di materiale archivistico è l'EAD (Encoded Archival Description) che specifica come etichettare gli inventari usando il linguaggio di marcatura SGML (il padre dell'XML). Si tratta di un esempio

³⁴ Il gruppo si è impegnato per favorire l'applicazione uniforme di standard di metadati per la ricerca e l'interoperabilità di metadati nei diversi settori,
<<http://www.iccu.sbn.it/genera.jsp?id=101#nota%201>>.

³⁵ <<http://www.schemas-forum.org/info-services/info-mats.html#Metadata>>.

³⁶ <<http://www.chiariglione.org/mpeg/index.htm>>.

di descrizione precisa e dettagliata dei vari elementi necessari all'identificazione di materiale archivistico, specificando gli elementi stessi ai vari livelli (fondi, sottofondi, serie, sotto-serie, singoli dossier). Diversi progetti stanno studiandone l'integrazione con i metadati descrittivi di risorse elettroniche gestite da biblioteche e servizi bibliografici.

Il CRM (CIMI-CIDOC Reference Model)³⁷ messo a punto dal CIMI (Consortium for the Computer Interchange of Museum Information) e dal CIDOC (Committee for Documentation of the International Council of Museums), è un modello strutturato ad oggetti per descrivere entità fisiche (gli oggetti del museo), concettuali (lavori relativi), temporali (periodici storici) e spaziali (luoghi). La sua prospettiva implica uno spazio informativo di scopo "enciclopedico" e fornisce accesso anche a altri standard correlati nel mondo web.

L'incompatibilità semantica e strutturale fra i sistemi esistenti ha spinto all'elaborazione di questo modello nell'intento di superare le difficoltà di scambio non solo con organismi come biblioteche, archivi, servizi di indicizzazione in rete, ma anche fra le stesse istituzioni preposte alla conservazione e documentazione di oggetti museali.

Le risorse di tipo accademico hanno avuto uno sviluppo particolare negli ultimi anni con la realizzazione dell'Open Archive Initiative (OAI),³⁸ la creazione di portali ad accesso per soggetto (tipo il progetto europeo Renardus, basato su Dublin Core, che integra portali di biblioteche nazionali), lo sviluppo di sistemi archiviazione, recupero tesi e dissertazioni tipo Networked Digital Library of Theses and Dissertation (NDLTD).

Riguardo al settore prettamente didattico, l'OCLC E-Learning Task Force³⁹ sostiene che le barriere culturali sono l'impedimento maggiore per lo sviluppo di strategie dei metadati per i cosiddetti *learning object*, per i quali il paradigma sta cambiando da un modello centrato sul possesso ad un modello centrato sull'utilizzo.

³⁷ Le linee guida sono disponibili all'indirizzo: <<http://cidoc.mediahost.org/>>.

³⁸ <<http://www.openarchives.org/>>.

³⁹ OCLC E-Learning Task Force, *White Paper -Libraries and the Enhancement of E-learning*, October 2003, <<http://www5.oclc.org/downloads/community/elearning.pdf>>.

Questo comporta un cambiamento di mentalità: storicamente i tre aspetti legati alla creazione, all'uso e alla descrizione degli oggetti (proprietà, gestione, controllo dell'accesso) si sono mantenuti all'interno di uno stesso ambiente (facoltà, dipartimento, istituto di ricerca). La dimensione web ha permesso di oltrepassare i confini spaziali verso una *cultura di uso allargato degli oggetti informativi*, con una nuova prospettiva sulle tre attività sopra citate, che nella cultura dove al centro è posto l'utente, sono assunte in carico da diversi gruppi o individui: dal docente che crea l'oggetto, al tecnico che ne produce un'adeguata versione on line affidata poi ad un documentalista o bibliotecario che la gestisce e ne determina il controllo di accesso in base alle preferenze del creatore. Poi vi è un'ulteriore figura che è l'amministratore che gestisce la risorsa. E' dunque necessario che le figure coinvolte nella predisposizione e nella diffusione dei *learning object* abbiano una chiara consapevolezza dei nuovi ruoli e delle interrelazioni con l'oggetto⁴⁰ allo scopo di definire nuovi strumenti per la creazione di metadati che siano possibilmente soluzioni *open-source* facilmente integrabili nelle architetture dei sistemi esistenti.

Ad un livello strategico superiore vi è un disperato bisogno di analisi dei requisiti degli schemi, dei vocabolari controllati, delle tassonomie e dell'adozione di profili di applicazione che assemblano elementi provenienti da domini diversi. La ricerca e lo sviluppo di specifiche e standard per le comunità di e-learning a livello internazionale hanno una storia molto breve che inizia circa dieci anni fa con la costituzione dell'IMS⁴¹ Global Learning Consortium, grazie al quale sono stati fatti significativi progressi con la definizione di modelli per la descrizione delle risorse didattiche e la produzione di specifiche tecniche aperte per le tecnologie interoperabili divenute standard di fatto e disponibili gratuitamente (con la relativa documentazione) in rete. L'IMS ha deciso di estendere la sua attività alle comunità bibliotecarie ed ha istituito un Digital Libraries Special Interest Group che ha pubblicato con la Coalition for Networked Information il draft pubblico del *Interoperability between information and learning*

⁴⁰ Wendy Pradt Lougee, *Diffuse Libraries: emergent roles for the research library in the digital age*, Council of Libraries and Information Resources, August 2002, <<http://www.clir.org/pubs/reports/pub108/contents.html>>.

⁴¹ <<http://www.imsglobal.org/>>, IMS Global Learning Consortium inizialmente era denominato Instructional Management System project, ma il nome generava confusione per cui oggi viene semplicemente denominato IMS.

*enviroments. Bridging the gaps*⁴², che costituisce la base per l'esplorazione delle potenziali interazioni tra gli ambienti informativi e quelli didattici, con il coinvolgimento degli standard, i modelli architetturali o le interfacce per permettere a questi due di coesistere ed evolversi con maggiore produttività.

Uno standard fondamentale per la gestione delle risorse didattiche digitali è LOM (Learning Object Metadata)⁴³ prodotto dall'IEEE, in cui ogni risorsa didattica è descritta da metadati suddivisi in 9 categorie contenenti gruppi di attributi strutturati ad albero. Gli elementi principali del *Base Schema* sono:

- ✿ *general*: gruppi di caratteristiche generali, indipendenti dal contesto del contenuto dell'oggetto didattico descritto;
- ✿ *lifecycle*: gruppi di caratteristiche relative al ciclo di vita della risorsa;
- ✿ *meta-metadata*: gruppi di caratteristiche della descrizione stessa;
- ✿ *technical*: caratteristiche tecniche della risorsa;
- ✿ *educational*: caratteristiche educative e pedagogiche della risorsa;
- ✿ *right*: indicazioni sull'utilizzo della risorsa;
- ✿ *relation*: gruppi di caratteristiche della risorsa che la collegano ad altre;
- ✿ *annotation*: caratteristiche che permettono la descrizione di note sull'utilizzo educativo della risorsa;
- ✿ *classification*: caratteristiche utili per classificare i contenuti della risorsa.

Il portale GEM (Gateway to Educational Metadata)⁴⁴ usa una versione una versione estesa di Dublin Core (21 elementi) ed offre l'editor di creazione automatica GEMCat, fornisce un accesso selettivo a materiali didattici di qualità presenti nelle rete.

⁴² A Joint White Paper on behalf of the IMS Global Learning Consortium and the Coalition for Netwroked Information, Joint White Paper - *Interoperability between information and learning enviroments. Bridging the gaps*, DRAFT by N.McLean and C.Lynch, June 2003, <http://www.imsglobal.org/dlims_white_paper_publicdraft_1.pdf>.

⁴³ Definizione dello standard:
<http://ltsc.ieee.org/wg12/files/LOM_I484_I2_I_V1_Final_Draft.pdf>.

⁴⁴ <<http://67.23.47.252/about/documentation2/schemas>>, dove sono ricercabili oltre 24000 record da oltre 350 collezioni di materiali didattici.

Lo standard **ONIX** prevede oltre duecento elementi per registrare una gamma molto ampia di informazioni valutative, promozionali, commerciali, oltre che bibliografiche di base. Sebbene la specifica sia stata concepita per i libri, con il tempo è stata usata anche per periodici, articoli di giornale, atti di convegni libri elettronici.

The ONIX for Books Product Information Message è lo standard internazionale per la rappresentazione e la comunicazione dell'informazione prodotta in forma elettronica, consentendo agli editori, distributori, biblioteche e librerie on line, come Amazon.com lo scambio di informazioni bibliografiche. ONIX for Books è sviluppato e mantenuto da EDItEUR in collaborazione con Book Industry Communication (UK) e del Book Industry Study Group (USA), e con gruppi di utenti in Australia, Canada, Francia, Germania, Italia, Paesi Bassi, Norvegia, Spagna, e della Repubblica di Corea. La ONIX for Books sito web fornisce tutti i dettagli dello standard e la versione più recente.

Per gli editori, l'esperienza ha dimostrato che ONIX for books porta due importanti benefici di business. Come formato di comunicazione, rende possibile lo scambio di informazioni sui prodotti nella catena di approvvigionamento con un modulo standard, a grossisti e distributori. Secondo, fornendo un modello per il contenuto e la struttura di un record di prodotto, ONIX ha contribuito a stimolare l'introduzione di migliori sistemi d'informazione interni, in grado di riunire tutti i metadati necessari per la descrizione e la promozione di nuovi titoli e backlist.

Lo standard ONIX definisce un elenco di campi di dati su libri e come inviare i dati sotto forma di un messaggio XML. Alcuni elementi sono necessari, ad esempio ISBN, il nome dell'autore e il titolo. Altri sono facoltativi, ad esempio le recensioni e le immagine di copertina.

Nella descrizione della documentazione, ONIX, è stato mappato (convertendo elementi e attributi) in UNIMARC e Formato MARC 21 e altri formati nazionali MARC utilizzati in vari paesi e addirittura può essere convertito in Dublin Core. La capacità di utilizzare set dei dati in formato XML e convertirlo in un altro formato XML permette il raggiungimento di una gamma più ampia di potenziali utenti e ancora una volta conferma che l'XML è un formato che facilita lo scambio di dati tra le applicazioni.

Bisogna sottolineare comunque alcuni svantaggi nell'utilizzo di ONIX: se il grande dettaglio del set elementi da un lato può esser ritenuto un vantaggio è al tempo stesso un punto a sfavore data la grande quantità di aggiornamenti sugli strumenti di sviluppo e di traduzione, dei nuovi campi di input e di conseguenza crescono i tempi necessari per l'immissione dei dati. Un altro aspetto problematico da considerare è che non sempre vi è piena compatibilità rispetto alla versione precedente. Per esempio, ONIX versione 2.1 è compatibile con la versione 2.0, ma la versione 2.0 non è compatibile con la precedente (versione 1.0).

PR.7 Title		
<i>Title composite</i> <Title> The Title Composite contains the text of a title, including a subtitle when necessary.		
<b202>	<TitleType>	(Used to set field 246 2nd indicator: If b202 = 00, 246 I2 = #; if b202 = 01, 246 I2 = 2)
<b276>	<AbbreviatedLength>	
<b203>	<TitleText>	246 \$a
<b030>	<TitlePrefix>	245 \$a
<b031>	<TitleWithoutPrefix>	245 \$a
<b029>	<Subtitle>	245 \$b; 246 \$b

Figura 11: crosswalk tra ONIX e MARC, nella prima colonna sono descritti i tag ONIX, nella seconda gli ONIX Reference Name e nella terza l'elemento corrispondente in MARC21. Fonte: <<http://www.loc.gov/marc/onix2marc.html>>.

DATA ELEMENT	SPEC	REFERENCE NAME	TAG	5	4	3	2	1	NOTES
Collection identifier composite		<CollectionIdentifier>						0-n	
Collection identifier type code, List 13	P.5.3	<CollectionIDType>	<x344>					1	Either <CollectionIdentifier> or <TitleDetail> or both must be present in each instance of the <Collection> composite.
Identifier type name	P.5.4	<IDTypeName>	<b233>					0-1 x	
Identifier value	P.5.5	<IDValue>	<b244>					1	
Title detail composite		<TitleDetail>						0-n	New composite, with new and existing elements.
Title type, List 15	P.5.6	<TitleType>	<b202>					1	Existing element, used in new location.
Title element composite		<TitleElement>						1-n	New composite, with new and existing elements ² .
Title element level, List 149	P.5.7	<TitleElementLevel>	<x409>					1	New element.
Part number	P.5.8	<PartNumber>	<x410>					0-1	New element.
Year of annual	P.5.9	<YearOfAnnual>	<b020>					0-1	Existing element, used in new location.
Title text	P.5.10	<TitleText>	<b203>					0-1 x x	Existing element, used in new location.
Title prefix	P.5.11	<TitlePrefix>	<b030>					0-1 x x	If <TitlePrefix> is present, <TitleWithoutPrefix> must also be present.
Title text without prefix	P.5.12	<TitleWithoutPrefix>	<b031>					0-1 x x	Existing element, used in new location.
Subtitle	P.5.13	<Subtitle>	<b029>					0-1 x x	Existing element, used in new location.
Contributor composite		<Contributor>						0-n	P.5.14 to P.5.61: same as P.7.1 to P.7.48
"No series" indicator (empty element)	P.5.62	<NoCollection/>	<x411/>					0-1	Formerly <NoSeries/>, renamed.
P.6 Title									
Title detail composite		<TitleDetail>						1-n	One instance of product title detail is mandatory.
Title type, List 15	P.6.1	<TitleType>	<b202>					1	Existing element, used in new location.
Title element composite		<TitleElement>						1-n	New composite, with new and existing elements ² .
Title element level, List 149	P.6.2	<TitleElementLevel>	<x409>					1	New element.
Part number	P.6.3	<PartNumber>	<x410>					0-1	New element.
Year of annual	P.6.4	<YearOfAnnual>	<b020>					0-1	Existing element, used in new location.
Title text	P.6.5	<TitleText>	<b203>					0-1 x x	Existing element, used in new location.
Title prefix	P.6.6	<TitlePrefix>	<b030>					0-1 x x	If <TitlePrefix> is present, <TitleWithoutPrefix> must also be present.
Title text without prefix	P.6.7	<TitleWithoutPrefix>	<b031>					0-1 x x	Existing element, used in new location.
Subtitle	P.6.8	<Subtitle>	<b029>					0-1 x x	Existing element, used in new location.

Figura 12: ONIX, descrizione di un titolo, (fonte: documentazione della versione 3 rilasciata ad Aprile 2009, <<http://www.editeur.org/93/Release-3.0-Downloads/>>).

2.4 Analisi delle esperienze innovative e schematizzazione delle applicazioni

In questo paragrafo vengono prese in considerazione alcune iniziative per la codifica dei metadati di oggetti digitali in depositi durevoli, vengono proposte in quanto apportano significativi contributi nella definizione del nuovo paradigma descrittivo per la gestione delle risorse digitali. Nella *tabella 5* sono riportati i dati identificativi, proponendone una vista sintetica e comparativa.

In ambito internazionale si registrano attualmente iniziative nello sviluppo teorico e nell'implementazione pratica di *trusted digital repositories*, ossia di sistemi digitali affidabili e durevoli per la raccolta, la gestione, la conservazione e l'accesso alle risorse digitali.

Nella prima parte della tabella che segue sono quindi descritte tre iniziative (OASIS, METS, VRA) di sviluppo teorico che hanno prodotto le basi di modellizzazione per la definizione di sistemi dotati di caratteristiche e funzionalità specifiche per la gestione di oggetti digitali in formati sempre più eterogenei.

Tabella 5 - Modelli di riferimento per la rappresentazione di oggetti digitali

Servizio e istituzione di riferimento	Scopi	Funzionalità / Componenti	Tipo di metadati	Tecnologia
OASIS Open Archive Information System del Consultative Committee for Space-Data System Panel	Proposta di un modello per gli oggetti digitali e i loro metadati	<ul style="list-style-type: none"> • Selezione • Acquisizione • Cattura • Registrazione • Verifica • Descrizione • Storage • Conservazione • Delivery • Accesso • Monitoraggio 	<p>Set di metadati vari legati ai progetti di implementazione:</p> <ul style="list-style-type: none"> ▶ National Library of Australia ▶ CEDARS 2000 ▶ Nedlib ▶ OCLC/RLG 	XML
METS Metadata Encoding Standard della Digital Library Federation e Network Development and MARC Standards Office Library of Congress	<p>Sintassi di trasferimento dati:</p> <ul style="list-style-type: none"> • tra le applicazioni • tra le istituzioni Driver per le applicazioni: <ul style="list-style-type: none"> • page turner • audio player • multimedia presentation 	<p>Documenti complessi riformattati o digitali nativi</p> <p>Namespaces di altri schemi usati come estensioni</p>	<ul style="list-style-type: none"> ✓ Descrittivi ✓ Amministrativi ✓ Strutturali 	XML
VRA Core Categories Type = Digital Visual Resources dell'Association Data Standards Committee	Record descrittivi di documenti di natura visiva e immagini	<ul style="list-style-type: none"> • Selezione • Acquisizione • Descrizione • Storage • Conservazione • Delivery • Accesso 	Set specifico per tipologia basato su Dublin Core Element Set	XML

Erpanet (Electronic Resource Preservation and Access Network)⁴⁵ rappresenta un ambiente di discussione internazionale a questo proposito e la definizione del RLG-OCLC *Report Trusted Digital Repositories: Attributes and Responsibilities* fornisce un buon punto di partenza: “a trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designed community, now and in the future” la definizione del RLG-OCLC continua affermando “all trusted digital repositories must accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users”, riassumendo di seguito fissa le caratteristiche:

- ✿ si caratterizza per un sistema organizzativo che non solo consente una sopravvivenza a lungo termine del *repository*, ma dichiara anche chi ha la responsabilità fisica e esegue l'aggiornamento;
- ✿ progetta il suo prototipo in conformità alle convenzioni accettate comunemente e agli standard per assicurare la gestione in corso, l'accesso e la sicurezza dei materiali depositati;
- ✿ stabilisce metodologie per la valutazione del sistema che soddisfano le aspettative di affidabilità della comunità;
- ✿ effettua la sua responsabilità a lungo termine in modo aperto ed esplicito per gli utenti;
- ✿ prevede controlli sulla pratica sui quali basa l'esecuzione.

Il report prevede inoltre che i trusted digital repositories siano compatibili con il modello di riferimento OASIS.

2.4.1 Alcuni esempi di digital repository

Il modello proposto in Open Archival Information System (OASIS reference model), sviluppato a partire dal 1995 ma proposto nella sua versione definitiva nel 2002, descrive un framework concettuale per un sistema di archiviazione con un modello informativo complementare che ha rappresentato il punto di partenza di molte ricerche e molto lavoro di implementazione. Sviluppato dal

⁴⁵ Sito ufficiale dell'Erpanet <<http://www.erpanet.org/>>, inoltre è stato significativo l'Erpaworkshop - *Trusted repositories for preserving cultural heritage*, ERPANET Workshop, Rome, November 2003, <<http://www.erpanet.org/events/2003/rome/RomeFinalReport.pdf>>.

Consultative Committee for Space Data System Panel 2, il modello è stato usato in progetti di ricerca quali InterPARES⁴⁶ e nella progettazione e realizzazione di archivi digitali⁴⁷ e di biblioteche digitali universitarie, nelle quali è stato spesso usato unitamente alle raccomandazioni RLG/OCLC (anch'esse sviluppate sul modello OAIS).⁴⁸

OAIS, un modello standard ISO⁴⁹, definisce la terminologia e i concetti rilevanti per l'attività di archiviazione digitale, identifica le componenti chiave e i processi ed essa persistenti e propone un modello informativo per gli oggetti digitali e i loro metadati. Tale modello non specifica alcuna implementazione ed è dunque neutrale riguardo al tipo di oggetti e agli aspetti tecnologici: è infatti possibile applicarlo ad ampio livello a file di immagini, oggetti *born-digital*, e persino ad oggetti fisici. Un modello di questo tipo opera in un ambiente composto da quattro entità (per maggiori dettagli su OAIS vedi 2.4.2):

-  *produttori* (fornitori dell'informazione)
-  *consumatori* (coloro che utilizzano l'informazione archiviata)
-  *management* (entità responsabile delle politiche a lungo termine dell'archivio)
-  *archivio* stesso che contiene le risorse

Un'interessante implementazione è rappresentata anche dal progetto NEDLIB⁵⁰ (maggiori dettagli sono presenti nel paragrafo successivo).

⁴⁶ The InterPARES (International Research on Permanent Authentic Records in Electronic System) 1 report: The long-term preservation of Authentic electronic records: findings of the InterPARES project, Vancouver, BC, Canada, September 2002 <<http://www.gseis.ucla.edu/us-interpares/pdf/InterPARES1FinalReport.pdf>>. Il progetto si trova nella seconda fase della ricerca (2002-2006) e si occupa del ciclo di vita dei record elettronici.

⁴⁷ Tra le iniziative di archivi digitali: United States National Archives and Records Administration's Electronic Records Archives (ERA), <<http://www.archives.gov/era/>>, e l'OCLC Digital Archive, <<http://www.oclc.org/digitalarchive/>>.

⁴⁸ Sull'argomento Nancy Y.McGovern, *Mapping organizational activities to the OAIS reference model*, Los Angeles, August 2003.

⁴⁹ ISO 14721:2003, Space Data and information transfer system - Open archival information system - Reference model, <http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683>.

⁵⁰ <<http://nedlib.kb.nl/>>.

Il **METS**, Metadata Encoding and Transmission Standard, con i relativi DTD e Schema XML⁵¹, è un'iniziativa della Digital Library Federation mantenuta presso il Network Development and MARC Standard Office della Library of Congress⁵² che ha sviluppato uno standard per la codifica di metadati descrittivi, amministrativi e strutturali basato su XML. Sono metadati necessari per la gestione degli oggetti nel repository e per lo scambio degli stessi *repositories* o con gli utenti.

Un software open source utilizzato in ambito accademico, sviluppato dal MIT per la realizzazione di digital repository è DSpace⁵³ creato per catturare, distribuire e conservare la produzione intellettuale in formato digitale: archivia testo, immagini, mpegs, e data sets. L'utente può caricare i propri documenti e poi inserire alcuni metadati come: titolo, autore, data, parole chiave ecc., su cui successivamente effettuerà le ricerche.

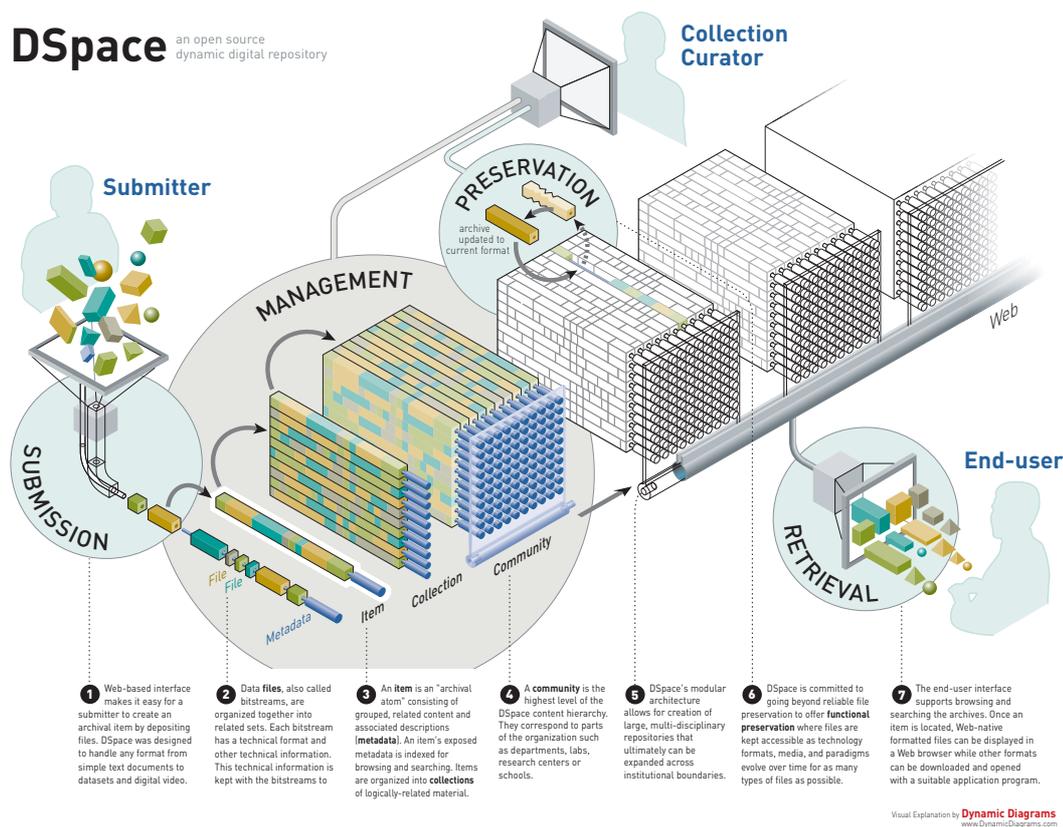


Figura 13: diagramma di DSpace che descrive visualmente la sua architettura

⁵¹ <<http://www.loc.gov/standards/mets/>>.

⁵² <<http://www.loc.gov/marc/ndmso.html>>.

⁵³ <<http://www.dspace.org>>.

Un altro esempio di software che permette la creazione di un digital repository è: **FedoraCommons**.⁵⁴

Si ricorda un altro software open source per la creazione di digital library: **Greenstone**.⁵⁵

Greenstone è un pacchetto di software adatto a gestire collezioni di biblioteche digitali e a creare nuove collezioni. Rappresenta un modo di organizzare informazioni e di pubblicarle su Internet o su CD-ROM. Greenstone è un prodotto della New Zealand Digital Library Project presso l'Università di Waikato. E' distribuito in cooperazione con l'UNESCO e lo Human Info NGO. Attualmente è disponibile la versione 3, completamente ridisegnata e implementata rispetto alla precedente versione. Mantiene le caratteristiche della versione 2: è multilingua, multiplatforma e altamente configurabile. Implementato in Java è strutturato in una rete di moduli indipendenti che comunicano tramite XML. L'architettura modulare aumenta la modularità e l'estensibilità di Greenstone. La versione 3 è ancora incompleta e quindi non del tutto stabile.

2.4.2 OAIS: Open Archival Information System

Il modello proposto da OAIS è implementato come sopra accennato nel progetto NEDLIB. Il modello fornisce la terminologia e i concetti rilevanti per l'attività di archiviazione digitale, identifica le componenti chiave e i processi persistenti ad essa e propone un modello informativo per gli oggetti digitali e i loro metadati. L'oggetto fondamentale in questo modello è rappresentato dall'oggetto informativo (*information object*), che a sua volta risulta composto da due elementi: l'oggetto digitale vero e proprio (*data object*), cioè un oggetto composto da un insieme di sequenze di bit, e le informazioni sulla sua rappresentazione (*representation informations*), cioè i metadati.

I set di metadati utilizzati nel modello OAIS sono stati impiegati nel corso degli anni in progetti quali la National Library of Australia (1999), CEDRAS (2000), Nedlib (2000), OCLC/RLG (2001).

⁵⁴ <<http://www.fedora-commons.org/>>.

⁵⁵ <<http://www.greenstone.org/>>.

Il modello di funzionamento Delivery and Capture di NEDLIB, ad esempio, descrive molto bene le varie fasi del *data flow* nell'assegnazione dei metadati: ricezione della pubblicazione, controllo della qualità, creazione di metadati (tecnici, strutturali, di controllo di accesso, bibliografici) aggiornamento del sistema esterno, creazione e sottomissione del SIP (Submission Information Package) secondo il modello OAIS. Nella fase di registrazione nel sistema di una nuova risorsa vengono considerati i seguenti metadati già presenti in quanto forniti da chi effettua l'invio della risorsa (quindi sono tutti forniti dall'esterno):

- ☀ i metadati descrittivi (per esempio *titolo*)
- ☀ metadati strutturali (per esempio la chiave di accesso, la struttura di navigazione, tipo il file index.html per una pubblicazione web)
- ☀ metadati di accesso (ad esempio specifica dei limiti di accesso)
- ☀ metadati tecnici (ad esempio specifiche tecniche per installare o visualizzare un tipo particolare di risorsa, readme.txt)
- ☀ file di definizione della risorsa per specificarne le varie componenti e le unità logiche.

I metadati esterni aggiunti nella fase di integrazione per completare la descrizione della risorsa e inserirla successivamente nel sistema comprendono i metadati bibliografici (per esempio identificatori come ISSN o ISBN) ed i record bibliografici forniti ad esempio dalla bibliografia nazionale, oltre ai metadati generati dal sistema di acquisizione.

Le tre tipologie di pacchetti di informazioni sono:

- ☀ “pacchetto di versamento”, il **SIP** Submission Information Package, per l'immissione dei dati
- ☀ “pacchetto di archiviazione”, l'**AIP** Archival Information Package, per l'archiviazione
- ☀ “pacchetto di distribuzione”, il **DIP** Dissemination Information Package, per distribuire le informazioni degli utenti

Per ognuna di queste diverse tipologie di pacchetti informativi il modello prevede poi ulteriori informazioni finalizzate a garantirne la conservazione, ovvero le PDI - Preservation Description Information -. Queste si suddividono in

quattro aree distinte, ognuna delle quali assolve a una particolare esigenza ai fini della conservazione:

- ☀️ informazioni sull'identificazione, *reference information*, informazioni rappresentate da un identificatore univoco come ad es. ISBN;
- ☀️ informazioni sul contesto, *context information*, informazioni che descrivono relazioni tra il contenuto del pacchetto informativo e i contenuti di altri pacchetti;
- ☀️ informazioni sulla provenienza, *provenance information*, informazioni che documentano l'iter del contenuto, come la creazione, le trasformazioni subite, i passaggi di possesso di proprietà ecc.;
- ☀️ informazioni sull'integrità, *fixity information*, informazioni sull'integrità del contenuto, cioè informazioni che garantiscono che il contenuto non ha subito alterazioni non autorizzate.

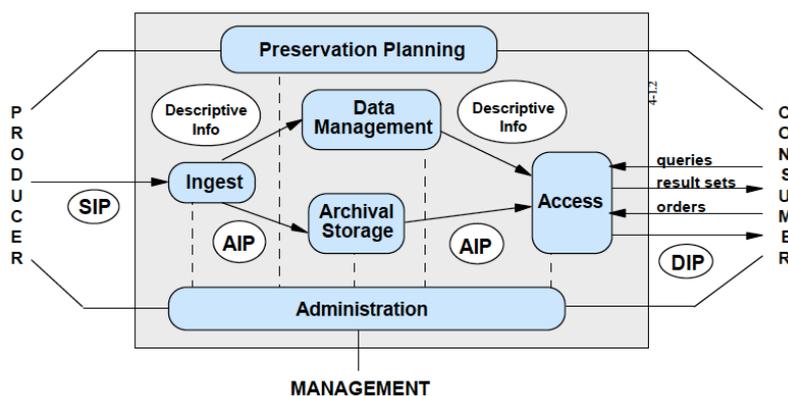


Figura 14: Reference Model di OAIS, fonte RLG-OCLC Report.

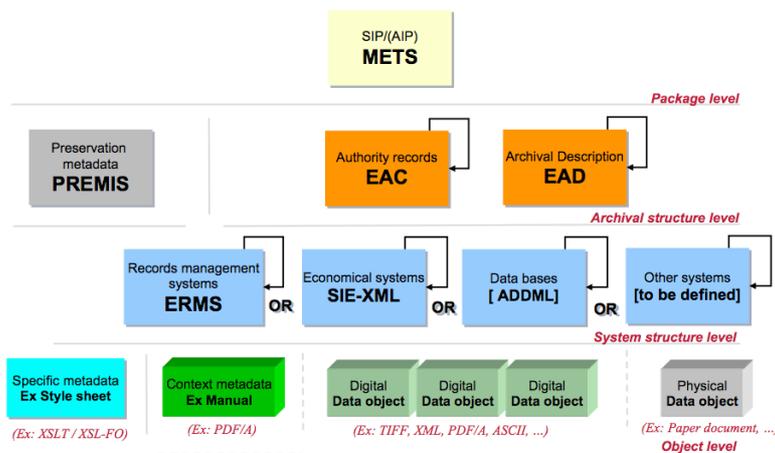


Figura 15: architettura generale di un SIP.

“[T]he dissemination of knowledge is one of the cornerstones of civilization.”

John F. Budd

Standard:

risorse bibliografiche

Nell'era della globalizzazione del digitale e del cyberspazio, l'innovazione tecnologica ha generato e sta comportando nell' "universo libro" e nell' "universo biblioteca" cambiamenti epocali. Nella delicata fase di transizione che vivono oggi i documenti, le risorse e le istituzioni preposte alla conservazione ed alla fruizione delle raccolte documentarie, al di là di posizioni "apocalittiche" sul destino prossimo venturo del libro e della biblioteca tradizionalmente intesi, è possibile ed auspicabile una prospettiva di *integrazione* tra trazione e innovazione. Quest'ultima deve guardare ai documenti, alle biblioteche ed ai cataloghi digitali (di e-commerce di una casa editrice, per esempio) come a un valore culturale aggiunto per la nostra contemporaneità ed una quasi "naturale" evoluzione di quanto ci consegna la tradizione, come attestano le sottili linee di demarcazione con teorie e pratiche bibliografiche relative, in generale, al mondo delle biblioteche e dei cataloghi, alle norme, agli standard e quant'altro.

Lo stato della ricerca relativa ai temi di catalogazione on line e delle biblioteche digitali, è in fieri, aperto, con una pubblicistica di settore copiosa, con la moltiplicazione di interventi su periodici specializzati e monografie di autorevoli studiosi di bibliologia e informatica.

Mutate la teoria e la pratica gestionale delle biblioteche in ordine allo spazio, ai servizi erogati, alle priorità, ai modelli della cooperazione e all'architettura dei sistemi di automazione, muta anche il ruolo del bibliotecario, che, in

quanto professionista e mediatore dell'informazione⁵⁶, oltre a catalogare i documenti, orienta l'utente nel reperimento di risorse diverse.

Al fine di consentire alle biblioteche lo scambio agevole delle registrazioni catalografiche e di permettere agli utenti di fruire delle registrazioni dei cataloghi elettronici, si è diffusa l'esigenza di *interoperabilità*, ossia di riutilizzo e scambio dei dati già esistenti; essa ha dato vita fin dagli anni 70 del Novecento a progetti di standardizzazione per conseguire il controllo bibliografico universale⁵⁷.

Secondo l'UK Interoperability Focus, l'**interoperabilità** è classificabile in:

- 🌸 tecnica, presuppone lo sviluppo e l'applicazione di standard quali ad esempio protocolli, i formati e i linguaggi;
- 🌸 semantica, presuppone lo sviluppo di dispositivi logici in grado di creare corrispondenze fra termini in sistemi di ricerca appartenenti ad ambiti disciplinari diversi;
- 🌸 politica/umana, presuppone la scelta delle istituzioni di rendere disponibili le proprie risorse informative;
- 🌸 multidisciplinare, presuppone l'avvio di iniziative di cooperazione tra biblioteche, archivi e musei;
- 🌸 internazionale, presuppone lo scambio internazionale dei dati organizzati in forma standardizzata.

L'interoperabilità fra sistemi eterogenei si connota in termini di capacità di un sistema o di un prodotto di operare con altri sistemi o prodotti senza richiedere un impegno particolare da parte dell'utente.

E' da sottolineare che l'organizzazione bibliografica, che svolge la funzione di archiviazione e registrazione dei documenti e di mediazione nell'accesso all'in-

⁵⁶ Si parla di "disintermediazione", ovvero della capacità dell'utente della biblioteca di soddisfare i propri bisogni informativi in modo autonomo, senza la mediazione del bibliotecario. Sul tema cfr. la miscellanea collettiva *Bibliotecario nel 2000: come cambia la professione nell'era digitale*, a c. di O. Foglieni, Editrice Bibliografica, Milano 1999.

⁵⁷ Nel 1998 è stato elaborato un nuovo modello concettuale per l'organizzazione delle unità informative (record) inserite in bibliografie e cataloghi, denominato *Functional Requirements for Bibliographic Records* (FRBR).

formazione e al documento che l'utente ricerca, si basa sul controllo bibliografico universale e sulla disponibilità delle pubblicazioni.⁵⁸

In generale, la standardizzazione è basilare per la realizzazione del **controllo bibliografico universale**⁵⁹, ossia una serie di strategie e di strumenti relativi al controllo ed allo scambio delle informazioni riguardanti le pubblicazioni.

L'espressione *controllo bibliografico universale* designa "un sistema internazionale finalizzato a rendere universalmente e immediatamente disponibili, su diversi supporti e in forma standardizzata definita e accettata in ambito internazionale, le registrazioni bibliografiche delle pubblicazioni prodotte in tutti i paesi; tale sistema si fonda sull'insieme dei sistemi nazionali di controllo bibliografico, responsabili, ciascuno per il proprio paese, della documentazione completa e autorevole della propria produzione editoriale".⁶⁰

L'UNESCO, in quanto organizzazione delle Nazioni Unite per l'educazione, la scienza e la cultura, e l'IFLA (*International Federation of Library Association and Institutions*), in quanto federazione delle associazioni ed istituzioni bibliotecarie operante a livello internazionale, si sono da anni attivate per elaborare e concretizzare programmi relativi all'organizzazione bibliografica e a quella bibliotecaria, promuovendo l'impiego di tecnologie elettroniche, lo scambio di registrazioni bibliografiche e la circolazione dei documenti (servizi di prestito interbibliotecario).

Allo scopo di rendere disponibili, in forma normalizzata accettata a livello internazionale, i dati bibliografici base di ogni pubblicazione edita in qualsiasi paese, dopo l'elaborazione dell'UBC (*Universal Bibliographic Control*), programma per il controllo bibliografico universale (Congresso IFLA, Liverpool 1971), e dell'UAP (*Universal Availability of Publications*), programma finalizzato ad assicurare la possibilità di accesso alle pubblicazioni, l'IFLA ha costituito l'*Office for UBC* (1974).

⁵⁸ Si tratta dell'insieme di strategie e strumenti che permettono la fruizione effettiva delle pubblicazioni.

⁵⁹ L'idea di dominare l'universo del sapere non è nuova nella storia: si pensi alle grandi biblioteche dell'antichità, come ad esempio la biblioteca di Alessandria di Egitto, o alle bibliografie come la *Biblioteca Universalis* di Conrad Gesner.

⁶⁰ G. Solimine, *Controllo bibliografico universale*, AIB, Roma 1995, p.5.

Anche l'ISO (*International Organization Standardization*), fondata nel 1946, che opera per promuovere la cooperazione attraverso la standardizzazione, ha elaborato standard per varie operazioni e procedure biblioteconomiche, per la traslitterazione degli alfabeti, per il set di caratteri da utilizzare nell'immissione dei dati, per il sistema di numerazione per i libri⁶¹: **ISBN**, *International Standard Book Number*, un codice che permette l'identificazione inequivocabile di un titolo o di un'edizione di un determinato editore, e per i periodici l'ISSN, *International Standard Serial Number*, un codice che identifica ogni testata.

3.1 Principi di catalogazione

3.1.1. Un po' di storia

I fondamenti di catalogazione, definita come la tecnica che ha come fine principale la mediazione, cioè il mettere in relazione un oggetto con i bisogni informativi dei suoi utenti⁶², furono posti da Antonio Panizzi⁶³ (1797-1879); le sue norme, note come “*91 regole*”, costituirono la base teorica dei successivi codici catalografici, tra i quali il Joint Code (1908) che, elaborato dall'*American Library Association* e dalla *Library Association* britannica, ha teso all'elaborazione di una scheda uniforme adottata in tutti i paesi di lingua inglese.

Un altro principio risale alla *Classificazione decimale Dewey* (anche DDC, acronimo dalla dizione inglese Dewey Decimal Classification o Dewey Decimal System) uno schema di classificazione, inizialmente rivolto all'ambiente bibliotecario, sviluppato da Melvil Dewey (1851-1931) nel 1876, da allora più volte modificato ed espanso in occasione delle ventidue revisioni principali di cui è stato oggetto fino al 2004. Nelle biblioteche pubbliche si utilizza frequente-

⁶¹ Si ricorda anche il codice a barre EAN destinato alla lettura ottica e composto da 13 cifre (per i libri le prime 3 cifre assumono il valore 978 che identifica il prodotto libro, le successive 9 cifre sono le prime 9 dell'ISBN, l'ultima cifra è il check number, numero di controllo). Vedremo una spiegazione più dettagliata nel prossimo capitolo.

⁶² B. Fadini, C. Savy, *Informatica per le scienze umane*, Franco Angeli, Milano 1999, p. 167.

⁶³ Rifugiato politico in Inghilterra, Antonio Panizzi fu bibliotecario e poi direttore del British Museum di Londra.

mente un'edizione ridotta, dal momento che non servono notazioni molto specifiche.

In Italia si segnalano le *Regole per la compilazione del catalogo* pubblicate, a cura di Luciano Bonazzi, nel 1922 (aggiornate poi nel 1956), le *Norme per il catalogo degli stampati* della Biblioteca Apostolica Vaticana pubblicate nel 1930 e le *Regole Italiane di Catalogazione per Autori* (1979).

Un progetto di regole internazionali di catalogazione fu oggetto di discussione nel corso di diversi congressi di bibliografia e documentazione, tenutisi a Bruxelles, a Roma e Venezia (1929).

Nel corso del Congresso internazionale di Madrid (1935) la *Fédération Internationale des Associations des Bibliothécaires* (FIAB) creò una commissione preposta all'unificazione delle regole di catalogazione.

In seguito, nel 1954, l'UNESCO propose alla FIAB di affrontare lo studio della possibilità di una cooperazione internazionale dei principi di catalogazione, e promosse e realizzò il Congresso a Parigi, finalizzato a conseguire l'unificazione delle diverse norme catalografiche in merito al problema della scelta e della forma di quegli elementi delle schede sui quali di basa l'ordinamento delle voci in un catalogo o lista.⁶⁴

Il codice *Rules of Descriptive cataloging in the Library of Congress*, adopted by the ALA (Washington LC 1949)⁶⁵ definiva i principi di base della catalogazione descrittiva recependo le indicazioni contenute nell'*Appendix A, Principles of descriptive cataloging* del Rapporto Henkle⁶⁶ (1946), ovvero lo scritto intitolato *Studies of descriptive cataloging: a report to the Librarian of Congress*⁶⁷, in cui si sosteneva che, oltre all'indicazione delle fonti, nelle registrazioni catalografiche i dati, completi, integrabili e non ripetuti (economia), erano da organizzare in forma ordinata e fissa.

I *Principi di Parigi* che sono stati alla base dell'elaborazione della RICA (*Regole Italiane di Catalogazione per Autori*, 1979), confermarono le due funzioni e gli

⁶⁴ La *Definizione di Principi* approvata dalla conferenza internazionale sui principi di catalogazione (Parigi, ottobre 1961).

⁶⁵ *ALA Catalog Rules: author and titles entries. Preliminary American 2nd edition.* è datato 1941.

⁶⁶ Herman Henkle era bibliotecario del Congresso e direttore del *Processing Department*.

⁶⁷ Cfr. *Il futuro della descrizione bibliografica*, Atti della giornata di studio, Firenze, 13 Novembre 1987, AIB, Roma 1988, pp. 89-157.

obiettivi del catalogo: consentire all'utente di stabilire se la biblioteca possiede o meno un documento di suo interesse, e segnalare, attraverso una sola forma del cognome dell'autore, quali opere e quali edizioni o traduzioni di una determinata opera possiede la biblioteca.

Negli anno Sessanta, il *Committee on uniform cataloging rules* avviò il progetto per la definizione di uno standard internazionale per la descrizione catalografica, affidando a Micheal Gorman l'analisi e la comparazione di alcune bibliografie nazionali.⁶⁸

Nel corso del fondamentale *International meeting of cataloging experts*, tenutosi nel 1969 a Copenaghen, si dibatté sull'applicabilità ai cataloghi di biblioteca dei criteri descrittivi delle bibliografie internazionali, sui criteri di trascrizione del frontespizio e sui temi della descrizione e dell'intestazione.

Fu creato un gruppo di lavoro - *Working group on an International bibliographic description* - allo scopo di elaborare un sistema per lo scambio internazionale delle informazioni basato su una descrizione bibliografica normalizzata, per forma e per contenuto, dei documenti su schede o su registrazioni leggibili dalla macchina.

Analizzati gli aspetti catalografici ed altre esigenze bibliografiche e conseguito un accordo internazionale sulla pratica di catalogazione, nacque l'**ISBD** (*International Standard Bibliographic Description*), che oggi si connota in termini di standard di riferimento per la catalogazione.

ISBD

L'*International Standard Bibliographic Description* nasce alla fine del Novecento tra gli anni 60 e 70, e raccoglie gli esiti del Congresso Internazionale di Copenaghen del 1969 in cui, come si è sottolineato, si decise di redigere uno standard di descrizione bibliografica di documenti valido, condiviso ed accettato a

⁶⁸ Furono esaminate e confrontate le seguenti bibliografie nazionali: *Bibliografija Jugoslavije*, *British National Bibliography*, *Bibliographie de la France*, *Bolétin bibliografico nactional*, *Buenos Aires*, *Deutsche Bibliographie*, *Magyar könyveszet*, *National Union Catalog*, *Svensk Bokförteckning*. Cfr. M. Gorman, *Bibliographical data in national bibliography entries: a report on descriptive cataloguing made for UNESCO & IFLA. Provisional abridged text*. (1969).

L'analisi dei dati e della struttura delle descrizioni costituì una base per l'elaborazione dello *Standard Bibliographic Description (SBD)*, apparso nel maggio 1971.

livello internazionale e tale da superare le eventuali difformità dei codici nazionali di catalogazione.⁶⁹

Lo standard, in grado di mediare le diversità delle prassi catalografiche, nasce dai seguenti scopi:

- 🦋 definire in tutto il mondo i termini per una descrizione compatibile;
- 🦋 consentire lo scambio di registrazioni catalografiche provenienti da agenzie bibliografiche differenti;
- 🦋 favorire l'abbattimento delle barriere linguistiche consentendo l'interpretazione dei dati da parte di utenti di diverse lingue;
- 🦋 favorire la conversione dei dati bibliografici in una forma leggibile dalla macchina.

ISBD è uno standard caratterizzato da una struttura normalizzata delle aree descrittive che registrano i dati catalografici e da un sistema di interpunzione convenzionale.⁷⁰

Gli ISBD⁷¹ sono vari e presentano alcune caratteristiche diverse a seconda del materiale trattato. Essi sono:

- 🌸 ISBD (A), *International Standard Bibliographic Description for Older Monographic Publications (Antiquarian)* (per il libro antico);
- 🌸 ISBD (CF), *International Standard Bibliographic Description for Computer Files* (per computer file);
- 🌸 ISBD (CM), *International Standard Bibliographic Description for Cartographic Materials* (per il materiale cartografico);
- 🌸 ISBD (ER), *International Standard Bibliographic Description for Electronic Resources* (per le risorse elettroniche);
- 🌸 ISBD (G), *International Standard Bibliographic Description for General* (lo standard generale di riferimento per tutti i tipi di materiale documentario);

⁶⁹ Anteriormente alla pubblicazione degli ISBD, le tradizioni catalografiche privilegiavano le intestazioni e gli aspetti organizzativi del catalogo oppure le registrazioni bibliografiche unificati.

⁷⁰ Lo standard stabilisce quali sono i dati bibliografici da includere, l'ordine di presentazione di tali dati e la punteggiatura attraverso cui questi devono essere contrassegnati.

⁷¹ Ogni manuale ISBD raccoglie le indicazioni per la propria categoria di pubblicazioni. Tutti gli ISBD si fondano sull'ISBD generale, ISBD (G).

- 🌸 ISBD (M), *International Standard Bibliographic Description for Monographic Publications* (per le monografie);
- 🌸 ISBD (NBM), *International Standard Bibliographic Description for Non-book Materials* (per il materiale non librario);⁷²
- 🌸 ISBD (PM), *International Standard Bibliographic Description for Printed Music* (per le opere musicali a stampa);
- 🌸 ISBD (S), *International Standard Bibliographic Description for Serials* (per le descrizioni delle pubblicazioni a carattere seriale).

Un primo standard provvisorio **SBD** (Standard Bibliographic Description for single and multi-volume monographic publications) fu elaborato nel 1971.

Nel 1977, tre anni dopo la pubblicazione di ISBD (M)⁷³, fu pubblicato ISBD (G).

Seguì la pubblicazione degli altri standard.

Nel 1990 sono state pubblicate le *Guidelines for applications of the ISBDs to description of component parts*, ovvero alcune raccomandazioni per la descrizione di contributi editi in veste non autonoma.⁷⁴

I diversi standard, che prevedono la registrazione di alcuni dati diversi a seconda delle caratteristiche del documento catalogato, si adeguano a ISBD (G), nato sulla base di un'idea elaborata nel 1975 dallo *Joint Steering Committee for the Revision of the Anglo-American Cataloguing Rules* (JSCAA-CR) che avanzò all'IFLA la proposta di creare uno standard bibliografico adatto a tutti i materiali presenti in biblioteca.⁷⁵

ISBD (G) *International Standard Bibliographic Description for General*

⁷² I *Non book Materials* comprendono materiale bibliografico non librario, ovvero materiali dal contenuto informativo presentati in forma diversa dal libro. Alcune "guide" per la catalogazione di questo tipo di materiale si trovano nelle edizioni del 1967 e 1978 delle *Anglo-American Cataloguing Rules*. I materiali trattati dagli ISBD (NBM) sono: ologrammi, edizioni Braille, microformati, pellicole cinematografiche, modelli tridimensionali, registrazioni sonore, videoregistrazioni ecc.

⁷³ ISBD (M), pubblicato nel 1974, fu riveduto e ripubblicato nel 1978 e nel 1987.

⁷⁴ Secondo il progetto originario questi documenti avrebbero dovuto essere descritti mediante lo standard ISBD (AN), e successivamente, ISBD (CP).

⁷⁵ In seguito furono organizzati alcuni incontri internazionali, fu redatta una bozza da Micheal Gorman e fu creato un Working Group on ISBD (G).

ISBD (G) è uno standard per la normalizzazione delle identificazioni e delle descrizioni bibliografiche, un sistema descrittivo uniforme con segni di interpunzione da collocare in aree, che consente, oltre allo scambio delle registrazioni catalografiche, di abbattere le barriere linguistiche e favorire la conservazione delle registrazioni in formato leggibile dalla macchina.

ISBD (G) si compone di otto aree:

- ✿ Area del titolo e della formulazione della responsabilità;
- ✿ Area dell'edizione;
- ✿ Area specifica del materiale (o tipo di pubblicazione);
- ✿ Area della pubblicazione, distribuzione ecc.;
- ✿ Area della descrizione fisica;
- ✿ Area della collana (o serie);
- ✿ Area delle note;
- ✿ Area del numero standard (o equivalente) e delle condizioni di disponibilità;

La descrizione normalizzata di ISBD (G) è organizzata come segue in *tabella*.⁷⁶

Tabella 6: Descrizione normalizzata dell'organizzazione ISBD (G).

		1.1 Titolo proprio
1) Area del titolo e della formulazione della responsabilità	[] = : / ;	1.2 Indicazione generale del materiale 1.3 Titolo parallelo 1.4 Complementi del titolo 1.5 Formulazione di responsabilità Prima formulazione Successiva formulazione
2) Area dell'edizione	= / : , / ;	2.1 Formulazione dell'edizione 2.2 Formulazione parallela di edizione 2.3 Formulazione di responsabilità relative all'edizione Prima formulazione Successiva formulazione 2.4 Ulteriore formulazione 2.5 Formulazione di responsabilità relative a ulteriore formulazione di edizione Prima formulazione Successiva formulazione

⁷⁶ Cfr. *International Federation of Library Associations and Institutions*, ISBD (G), ed. it. a c. di R. Dini, AIB, Roma, 1987, pp. 12-15.

		1.1 Titolo proprio
3) Area specifica del materiale (o tipo di pubblicazione)		
4) Area della pubblicazione, distribuzione ecc.	; : [] , (: ,)	4.1 Luogo di pubblicazione, distribuzione ecc. Primo luogo Successivo luogo 4.2 Nome dell'editore, distributore ecc. 4.3 Formulazione di funzione dell'editore, distributore, ecc. 4.4 Data di pubblicazione 4.5 Luogo di manifattura 4.6 Nome della manifattura 4.7 Data della manifattura
5) Area della descrizione fisica	: ; +	5.1 Indicazione specifica del materiale ed estensione del documento 5.2 Altre particolarità fisiche 5.3 Dimensione del documento 5.4 Indicazioni di materiale allegato
6) Area della collana (o serie) Nota: la formulazione di serie è racchiusa tra parentesi tonde. Se vi sono due o più formulazioni di serie, ciascuna è racchiusa tra parentesi tonde.	= : / ; , . = : / / ;	6.1 Titolo proprio della serie 6.2 Titolo parallelo della serie 6.3 Complementi del titolo della serie 6.4 Formulazione di responsabilità relative alla serie Prima formulazione Successiva formulazione 6.5 International Standard Serial N. della serie 6.7 Numerazione e titolo della serie 6.8 Titolo parallelo della sottoserie 6.9 Complementi del titolo della sottoserie 6.10 Formulazione di responsabilità relative alla sottoserie Prima formulazione Successiva formulazione 6.11 International Standard Serial N. della serie 6.12 Numerazione all'interno della sottoserie
7) Area delle note		
8) Area del numero standard (o equivalente) e delle condizioni di disponibilità	= : ()	8.1 Numero standard (o equivalente) 8.2 Titolo chiave 8.3 Condizioni di disponibilità e/0 prezzo 8.4 Qualificazione (in posizioni variabili)

Le maiuscole sono utilizzate per la prima lettera della parola iniziale di ciascuna area e secondo l'uso della lingua in cui le informazioni sono date nella descrizione⁷⁷.

Quest'ultima prevede l'uso delle parentesi tonde⁷⁸ (), delle parentesi quadre⁷⁹ [], dei punti di omissione⁸⁰ (...) ed altri segni di interpunzione.

ISBD (S), *International Standard Bibliographic Description for Serials*

ISBD (S) fu pubblicato nel 1974 sotto forma di raccomandazione.⁸¹

Si tratta di uno standard che utilizza strumenti quali l' International Standard Serial Number (ISSN)⁸² ed il titolo chiave al posto del titolo proprio. Ai perio-

⁷⁷ Si utilizzano poi le seguenti abbreviazioni: "cm" per centimetri; "mm" per millimetri; "et. al." per et alii (ad altri); "s.l." per *sine loco* (luogo di pubblicazione ecc., sconosciuto) e "s.n." per *sine nomine* (nome dell'editore ecc., sconosciuto).

⁷⁸ Le parentesi tonde e le parentesi quadre sono considerate un unico segno. Ad eccezione del punto e della virgola ciascun segno di punteggiatura che preceda o racchiuda un elemento è preceduto o seguito da uno spazio. Ad eccezione della prima, e a meno che l'area non sia distinta da capoverso, ciascuna area della descrizione è preceduta da punto, spazio, linea lunga, spazio (. _).

⁷⁹ Le parentesi quadre [], oltre che per racchiudere particolari elementi, sono la punteggiatura prescritta all'interno di un elemento e sono utilizzate per indicare integrazioni nella descrizione ed informazioni tratte fuori dalla fonte prescritta. Per la redazione della descrizione, di norma, sono preferite informazioni presenti nel documento.

⁸⁰ Si utilizzano punti di omissione (...) in alcuni casi per indicare l'omissione di una parte di descrizione all'interno di un elemento. Si omette qualsiasi area o elemento (e relativa punteggiatura che lo precede o racchiude) che non si applichi al particolare documento descritto.

⁸¹ Questo standard non seguì le analoghe vicende degli altri ISBD, in quanto si pose il problema della omogeneità fra ISBD (M) e ISBD (S). Oltre all'ISBD (S), si segnala l'*International Serial data System*. (ISDS).

⁸² L'ISSN (*International Standard Serial Number*) è un codice internazionale del materiale librario in commercio e consiste in un numero, stampato sul verso del frontespizio, che identifica univocamente una edizione di un'opera pubblicata da un dato editore. L'ISBN (*International Standard Book Number*) invece è costituito da dieci cifre suddivise in quattro sequenze che riguardano il gruppo linguistico o geografico, l'editore, il titolo o l'edizione, un carattere di controllo.

dici⁸³ sono collegate pubblicazioni in serie⁸⁴ (come ad esempio gli atti degli enti pubblici e congressi, e le collezioni).

A ciascuna pubblicazione è assegnato un **ISSN** (International Standard Serial Number), ovvero un numero internazionale normativo composto da otto cifre (a gruppi di quattro separate da trattino) e ideato allo scopo di semplificare il trattamento automatico delle pubblicazioni in serie.⁸⁵

Per le pubblicazioni seriali la Commissione Tecnica 46 (Documentazione) dell'ISO ha elaborato norme relative alla costituzione dell'ISSN, per la compilazione dei riferimenti bibliografici, per la traslitterazione, per la codifica di lingue, paesi ecc.

Da ricordare è senza dubbio anche la norma UNI 6392⁸⁶ che stabilisce un criterio di rivelazione dei dati catalografici e prescrive un ordine obbligatorio per la presentazione dei dati.⁸⁷

ISBD (A), *International Standard Bibliographic Description for Older Monographic Publications (Antiquarian)*

⁸³ Nella RICA (Regole Italiane di Catalogazione per Autori, 1979) la descrizione dei periodici, che recepisce i criteri di ISBD (S), è oggetto di specifica e distinta trattazione, nella quale si rimarca che è possibile redigere per i periodici o schede ordinate all'interno del catalogo per autori e titoli o un catalogo speciale con le intestazioni sotto il titolo.

⁸⁴ Le *Directive* per l'applicazione degli ISBD alla descrizione della parti componenti - approvate dalle Commissioni permanenti della Sezione IFLA della catalogazione e della Sezione IFLA delle pubblicazioni in serie - sono state pubblicate nel 1991. Da citare è il programma UNISTST, che fornisce una definizione del livello analitico di una registrazione bibliografica per descrivere un'unità bibliografica. Il sistema, che rientra nell'ambito del programma UNISTS dell'UNESCO, ha lo scopo di costituire un registro delle pubblicazioni in serie di ogni disciplina e provenienza.

⁸⁵ ISDS associa le pubblicazioni al titolo chiave (che corrisponde al titolo distintivo, ossia quello principale che è sufficiente ad identificare la pubblicazione a cui, se eventualmente comune a più pubblicazioni, si associano elementi caratterizzanti, quali il luogo o la data di inizio o altri dati), e all'ISSN utilizzato per lo scambio delle informazioni a livello internazionale.

⁸⁶ La norma UNI 6392 -Documentazione e riproduzione documentaria - Cataloghi alfabetici di periodici elaborata dalla commissione Documentaria dell'UNI (UNI-DRD), pubblicata nel 1968 in via sperimentale (UNI 6392-68) e poi rivista e destinata ai cataloghi di periodici di ogni tipo (UNI 6392-76) - intende fornire una metodologia generale per la registrazione delle pubblicazioni in serie incluse nei cataloghi di periodici; recepisce numerose soluzioni particolari di ISBD (S) oltre che lo schema e i criteri fondamentali della normativa internazionale.

⁸⁷ Alcuni dati sono intesi come "essenziali" (livello 1), altri come "facoltativi" ("di interesse primario" - livello 2 e "di interesse secondario" - livello 3).

LISBD (A) riguarda il libro antico ed è organizzato come segue:

- ☀ Area del titolo e dell'indicazione dell'autore contiene: titolo in senso stretto; indicazione del tipo di materiale (facoltativo); titoli paralleli; altre informazioni riguardanti il titolo.
- ☀ Area dell'edizione che contiene: indicazione di edizione; indicazioni di edizioni parallele; indicazioni di responsabilità che seguono le ulteriori indicazioni di edizione.
- ☀ Area della pubblicazione, distribuzione, ecc. che contiene: luogo principale di pubblicazione, distribuzione ecc.; primo luogo, altri luoghi; indicazioni di pubblicazione; data di pubblicazione; luogo di stampa dato in una fonte di informazione diversa dalla principale; primo luogo; altro luogo; altre informazioni sulla stampa date da una fonte di informazioni diversa dalla principale.
- ☀ Area della descrizione fisica che contiene: indicazione della consistenza; indicazione del materiale illustrativo; formato e/o dimensioni; materiale accessorio (facoltativo).
- ☀ Area della collezione che registra: titolo in senso stretto; titoli paralleli della collezione; altre informazioni riguardanti il titolo della collezione (facoltativo); indicazioni di responsabilità relativa alla collezione (facoltativo); ISSN; numerazione interna della collezione; titolo della sotto-collezione (facoltativo); indicazioni di responsabilità relativa alla sotto-collezione; ISSN della sotto-collezione; numerazione interna della sotto-collezione.

Altre aree dello standard sono l'area delle note e l'area dell'impronta.⁸⁸

ISBD(ER), International Standard Bibliographic Description for Electronic Resources

La prima versione delle specifiche degli ISBD(ER) viene pubblicata nel 1997 da K.G Saur, con lo scopo di rinnovare lo standard *ISBD(CF) International*

⁸⁸ Si tratta di un elemento identificativo del libro antico. L'impronta è costituita dalla notazione della data e da 16 caratteri - a gruppi di quattro - rilevati dall'ultima e penultima riga di terminate pagine.

standard bibliographic description for computer files, del 1990. Il precursore era lo standard AACR2 degli anni Ottanta.

I maggiori cambiamenti nell'ISBD(ER) si devono principalmente ai rapidi mutamenti nel mezzo elettronico, in particolare nel caso del materiale ad accesso remoto disponibile in internet. L'indicazione dell'ambito di applicazione è stata allargata per prendere in considerazione i servizi on line, i siti web, ed altre recenti risorse elettroniche in ascesa che in precedenza non erano trattate da ISBD(CF). Le linee guida sono compatibili con le norme dell'IFLA. In particolare vengono specificate le seguenti trattazioni:

- ✿ titolo;
- ✿ designazione generica del materiale;
- ✿ indicazione di responsabilità (ISBD(ER) 1.5.1 stabilisce che “un’indicazione di responsabilità può riferirsi ad ogni singola entità (autore personale o collettivo) responsabile o che abbia contribuito alla creazione del contenuto intellettuale o artistico dell’opera contenuta nel documento descritto”). A tale proposito possono essere indicati come responsabili anche: scrittori di testi, programmatori, ricercatori principali, grafici, compositori ecc., adattatori di opere preesistenti sullo stesso supporto dell’originale o su uno diverso, sviluppatori e designers, webmasters. L’IFLA invece prescrive di non creare indicazione di responsabilità se non è riportata nel documento;⁸⁹
- ✿ edizione, distinguendo tra nuova edizione e aggiornamento (nuova edizione si ha quando: ci sono aggiunte, omissioni, differenze nel linguaggio di programmazione, cambiamenti che aggiornano o migliorano l’efficienza della risorsa, cambiamenti nel linguaggio di programmazione o nel sistema operativo che ampliano la compatibilità);
- ✿ tipo di estensione della risorsa;
- ✿ designazione della risorsa (es. “base di dati bibliografica elettronica”, “giornale elettronico” ecc. nasce però il problema della proliferazione in quanto il catalogatore può supplire dove manca un termine adatto per designare la risorsa);
- ✿ estensione della risorsa (es. numero file che costituiscono il contenuto dei dati o programmi);
- ✿ pubblicazione, distribuzione;

⁸⁹ <<http://archive.ifla.org/VII/s13/pubs/isbd.htm>>.

- ✿ luogo di pubblicazione, produzione e/o distribuzione;
- ✿ nome dell'editore, produttore e/o distributore;
- ✿ data di pubblicazione, produzione e/o distribuzione;
- ✿ note (che riguardano fonti del titolo, autore, luogo ecc.);
- ✿ numero standard, per le risorse che non hanno un ISBN o ISSN di dovrà registrare l'URN⁹⁰ o il DOI.⁹¹

Area	Fonti d'informazione prescritte
1. Titolo e formulazione di responsabilità	Fonti interne; etichette sul supporto fisico; documentazione, contenitore o altro materiale allegato
2. Edizione	Fonti interne; etichette sul supporto fisico; documentazione, contenitore o altro materiale allegato
3. Tipo ed estensione della risorsa	Qualsiasi fonte
4. Pubblicazione, distribuzione etc.	Fonti interne; etichette sul supporto fisico; documentazione, contenitore o altro materiale allegato
5. Descrizione fisica	Qualsiasi fonte
6. Serie	Fonti interne, etichette sul supporto fisico; documentazione, contenitore o altro materiale allegato
7. Note	Qualsiasi fonte
8. Numero standard (o equivalente) e condizioni di disponibilità	Qualsiasi fonte

Figura 16: fonti di informazioni prescritte da ISBD(ER).

Functional Requirement for Bibliographic Records (FRBR)

I Functional Requirements for Bibliographic Records (Requisiti funzionali per la registrazioni bibliografiche)⁹², che costituiscono un modello innovativo per la costruzione di strumenti necessari alla revisione della normativa catalogafi-

⁹⁰ URN, *Uniform Resource Locator* è un identificatore univoco per le risorse che si trovano in Internet, il quale viene utilizzato come riferimento stabile a degli URL (Uniform Resource Locator o - più semplicemente - link). Esso può sostituire gli URL nei cataloghi delle biblioteche ed in altri sistemi d'indicizzazione, ma anche venir utilizzato come riferimento stabile all'interno dei documenti stessi.

⁹¹ DOI, *Digital Object Identifier* è uno standard che consente di identificare persistentemente, all'interno di una rete digitale, qualsiasi oggetto di proprietà intellettuale e di associarvi i relativi dati di riferimento, i metadati, secondo uno schema strutturato ed estensibile.

⁹² Sul tema Cfr. C.Ghilli, M. Guerrini, *Introduzione a FRBR. Functional Requirements for Bibliographic Records, Requisiti funzionali per record bibliografici*, Editrice Bibliografica, Milano 2001.

ca,⁹³ delineando funzioni svolte da un *record* bibliografico rispetto ai diversi media, alle varie applicazioni e ai molteplici bisogni informativi dell'utente.

FRBR “rappresenta l'elaborazione più originale e importante edita dai *Principi di Parigi* dal 1961 a oggi, e si presenta come riflessione per l'analisi e l'individuazione dei requisiti base dei *record* inseriti in bibliografie e cataloghi, e come perno per la costruzione di strumenti utili alla revisione della normativa catalografica attualmente in vigore”.⁹⁴

Considerata la diffusione dell'utilizzo, in tutto il mondo, degli OPAC (*Online Public Access Catalog*), sia da parte dei catalogatori sia da parte degli utenti, FRBR sostituiscono i *Principi di Parigi*⁹⁵ e li estendono dalle sole opere testuali a tutti i tipi di materiali ed a tutti gli aspetti delle registrazioni bibliografiche e di autorità utilizzate nei cataloghi di biblioteca (e non più alla sola scelta e forma di accesso).

FRBR tendono ad incrementare la condivisione internazionale di dati bibliografici e di autorità, e a guidare gli estensori di regole di catalogazione nel loro impegno per l'elaborazione di un codice internazionale di catalogazione, sulla base della tradizioni catalografiche del mondo, ma anche dei modelli concettuali dei documenti IFLA.⁹⁶

L'idea di studiare i requisiti funzionali dei *records* risale al Seminario tenutosi nel 1990 a Stoccolma, sotto la direzione dell'IFLA: nel 1991 è stato diffuso un documento⁹⁷ con lo scopo di stabilire le funzioni dei *records* in modo univoco rispetto ai media, le applicazioni ed i bisogni informativi dell'utente. Dopo la conferenza di Copenaghen (1997), nel 1998 sono stati pubblicati i *Functional*

⁹³ La Bozza di *principi* tratta i seguenti punti: 1) Ambito di applicazione; 2) Entità; attributi e relazioni; 3) Funzioni del catalogo; 4) Descrizione bibliografica; 5) Punti di accesso; 6) Registrazioni di autorità.

⁹⁴ C, Ghilli, M. Guerrini, *op. cit.* p.7.

⁹⁵ La *Dichiarazione dei Principi* - comunemente nota come *Principi di Parigi* - fu approvata nel 1961 dalla Conferenza Internazionale sui principi di catalogazione con lo scopo di servire come base per una standardizzazione internazionale della catalogazione.

⁹⁶ *Functional Requirements for Bibliographic Records (FRBR)* e *Functional Requirements and Numbering for Authority Records (FRANAR)*, estendono i *Principi di Parigi* al campo della catalogazione per soggetti. Nel 1998 l'IFLA ha costituito il gruppo di studio di FRANAR che circoscrive le finalità degli *authority records* FRBR.

⁹⁷ *Terms of reference for a study of functional requirements for bibliographic records.*

Requirements for Bibliographic Records, un modello concettuale elaborato allo scopo di consentire l'identificazione dei requisiti minimi della descrizione catalografica che interessa l'utente, il quale consulta i records bibliografici e consente di stabilire le finalità delle registrazioni e le loro caratteristiche strutturali.⁹⁸

Dopo oltre venticinque anni dalla pubblicazione delle Regole Italiane di Catalogazione per Autori, le mutate esigenze delle biblioteche e dell'utenza⁹⁹ hanno infatti generato il bisogno di istituire una Commissione permanente per la revisione del codice, nato già con la possibilità di accogliere eventuali integrazioni ed ampliamenti successivi.

Sebbene siano emerse la validità e l'applicabilità dei principi sui quali le RICA sono basate, è sorta al contempo l'esigenza di revisionare, estendere e far evolvere il codice, anche alla luce degli standard e dei documenti accettati a livello internazionale, allo scopo di rispondere:

- * all'evoluzione delle pratiche catalografiche;
- * alla necessità di un aggiornamento della terminologia;
- * ampliamento degli esempi legati al contesto automatizzato;
- * nuove tipologie di documenti sempre più presenti nelle biblioteche.

Nei *Functional Requirement for Bibliographic Records*, concetti fondamentali sono quelli di "Entità, attributi e relazioni".

Ai fini della redazione delle registrazioni bibliografiche sono da considerarsi **entità** riguardanti i prodotti del lavoro intellettuale o artistico: l'opera, l'espressione, la manifestazione ed il documento. Le registrazioni bibliografiche devono rispecchiare, tipicamente, le manifestazioni¹⁰⁰ (costituite da un raccolta di opere, una singola opera, una parte componente di un'opera). Infatti le entità sono "oggetti primari di interesse di chi ricerca i prodotti del lavoro intellettuale e artistico. Le entità sono definite focalizzando l'attenzione sulle metodologie e sulle esigenze di ricerca e recupero di chi usa il *record*."

⁹⁸ Esse variano a seconda delle tipologie delle risorse.

⁹⁹ Si segnalano: il cambiamento delle strutture del catalogo con l'uso sempre più ampiamente diffuso delle tecnologie informatiche, il mutamento nelle modalità di ricerca per l'utente, la presenza di tipologie documentarie diverse (per supporto e/o per modalità di accesso) e lo sviluppo della catalogazione derivata.

¹⁰⁰ Le manifestazioni possono presentarsi in una o più unità fisiche. Per ciascun formato fisico (manifestazione), in genere, va creata una registrazione bibliografica separata.

Ricordiamo che FRBR analizza la registrazione tradizionale e che le entità sono essenzialmente gli elementi descritti e di indicizzazione presenti nel *record*¹⁰¹.

Le entità del primo gruppo sono:

✿ *opera*: è la prima delle entità del modello FRBR; è un'entità astratta, una creazione intellettuale o artistica distinta, riconosciuta attraverso le sue varie *espressioni* (realizzazioni).

✿ *espressione*: è la seconda entità del modello, la realizzazione (artistica o intellettuale) di un'opera in forma diversa (alfanumerica, immagine, musicale, notazione coreografica, sonora, oggetto in movimento) o in una combinazione di varie forme.

✿ *manifestazione*: è la terza entità del modello concettuale, la materializzazione fisica dell'espressione di un'opera che, dunque, costituisce la serie di oggetti fisici che diffondono o pubblicano l'espressione di un'opera.

✿ *item*: quarta entità del modello, un esemplare o una copia, un documento della manifestazione.

Le entità del secondo gruppo sono:

✿ *persona*: singoli individui coinvolti nella creazione o nella realizzazione dell'opera.

✿ *ente*: organizzazioni o gruppi di persone identificati con un nome.

Le entità del terzo gruppo sono:

✿ *concetto*: un'ampia gamma di astrazioni, di nozioni astratte, di idee come soggetto di un'opera (ad esempio il soggetto di un'opera filosofica).

✿ *oggetto*: una cosa materiale quale soggetto di un'opera.

✿ *evento*: un'ampia gamma di azioni o avvenimenti quali soggetti dell'opera (ad esempio, il soggetto di un dipinto).

✿ *luogo*: una località soggetto di un'opera (ad esempio, il soggetto di una mappa).

¹⁰¹ C. Ghilli, M. Guerrini, op. cit. p.41.

Gli **attributi** caratterizzano le entità e si configurano come un mezzo fondamentale a disposizione degli utenti per recuperare le informazioni di loro interesse¹⁰², infatti *“l'attributo funge anche da chiave di recupero per l'entità”*. (Questa funzione dell'attributo è stata ripresa anche nella codifica TEI realizzata, vedi sezione Tirocinio Guerini).

Le *registrazioni di autorità* devono documentare forme controllate di nomi; le entità che servono come soggetti di opere comprendono: l'opera, l'espressione, la manifestazione, il documento, la persona, l'ente, il concetto, l'oggetto, l'evento e il luogo.

Gli *attributi*¹⁰³ che identificano ciascuna entità devono essere utilizzati come elementi nelle registrazioni bibliografiche e di autorità; devono essere identificate nel catalogo anche le relazioni bibliograficamente significative fra le entità in cui vi sono relazioni primarie che consentono l'individuazione di collegamenti utili all'utente che consulta la base di dati bibliografica.

Una **relazione** è un'associazione tra entità, essa aiuta l'utente a svolgere le attività di ricerca all'interno di archivi (bibliografie, cataloghi e basi di dati bibliografiche).

Le *relazioni tra entità del primo gruppo* collegano *opera a espressione*, *espressione a manifestazione* e *manifestazione a item*.

Le relazioni esistenti sono: *si realizza nella*, *si materializza nella*, *è rappresentata da*.

Il catalogo assolve a due precise funzioni: il reperimento e la localizzazione delle risorse bibliografiche. Esso pone l'utente in grado di reperire, attraverso una ricerca in cui si utilizzano attributi o relazioni, risorse bibliografiche in una raccolta, reale o virtuale, di localizzare una singola risorsa e/o insieme di risorse, e poter accedere al documento richiesto.

¹⁰² Si dicono concorrenze le materializzazioni di un'entità del mondo reale. Tali occorrenze presentano valori per ciascun attributo; ad esempio le occorrenze dell'entità pittori sono i singoli pittori.

¹⁰³ Gli attributi possono essere multipli e variabili; possono essere collegati all'entità o essere a essa esterni. Vi sono attributi dell'*opera* (titolo dell'opera, data dell'opera ecc.), dell'*espressione* (indice di contenuto, periodicità prevista di pubblicazione seriale, scale cartografica ecc.), della *manifestazione* (formulazione di responsabilità, dimensione del supporto, tipo di audio, indirizzo di accesso ecc.), dell'*item* (identificatore, stato di conservazione). Vi sono poi attributi della *persona* (nome, luogo ecc.), del *concetto* (termine per l'oggetto), dell'*evento* (termine dell'evento), del *luogo* (termine del luogo).

FRBR analizza gli attributi e le funzioni in riferimento alla ricerca dell'utente. Le funzioni utente sono quattro: trovare, identificare, selezionare ed ottenere occorrenze.

I **punti di accesso** di una *registrazione bibliografica* sono costituiti dai titoli delle opere, e di solito dai loro creatori. Secondo FRBR, nella redazione dei *records* le agenzie bibliografiche devono operare in modo da far sì che possano essere attivate diverse funzioni utente.

Ad esempio la ricerca può avvenire per: responsabile dell'opera, titolo, soggetto, serie ecc.

In merito alla **descrizione bibliografica**, FRBR specifica che la parte descrittiva della registrazione bibliografica deve basarsi su uno standard accettato a livello internazionale. Sulla base delle finalità del catalogo o archivio bibliografico, le descrizioni possono presentare vari livelli di completezza.

Gli **elementi descrittivi** riguardano aree:

- ★ area del titolo e responsabilità;
- ★ area dell'edizione;
- ★ area specifica del materiale (o del tipo di pubblicazione);
- ★ area della pubblicazione;
- ★ area della distribuzione;
- ★ area della descrizione fisica;
- ★ area della serie;
- ★ area delle note;
- ★ area del numero standard

Gli **elementi organizzativi** invece sono:

- ★ intestazione al nome;
- ★ intestazione al titolo;
- ★ intestazione alla serie;
- ★ intestazione al soggetto o alla classe

3.2 Automazione e biblioteche, MARC e DC

La gestione e l'organizzazione dei servizi della biblioteca e le modalità di cooperazione fra agenzie bibliografiche nazionali ed internazionali sono state fortemente innovate, come già detto, dall'introduzione dell'automazione che

ha configurato ulteriormente la biblioteca quale centro di documentazione e diffusione culturale in grado di avvalersi dell'innovazione tecnologica.

Fin dagli anni Settanta i progetti di catalogazione centralizzata, tra i quali si segnala quello della Library of Congress di Washington, sono stati finalizzati a consentire la circolazione di dati catalografici on line resa fattibile attraverso l'uso del **formato MARC** (***M**Achine **R**eadable **C**atalogin*), uno schema standardizzato di codifica dei dati catalografici elaborato per convertire le informazioni bibliografiche in linguaggio macchina.

A partire dal formato MARC, poi modificato da diverse agenzie bibliografiche di vari paesi furono elaborati l'USMARK (*United States MARC*, in America), l'UK MARC (*United Kingdom MARC*, nel Regno Unito) e l'ANNA MARC¹⁰⁴ (*Automazione Nella Nazionale MARC*, in Italia).

LIFLA (*International Federation of Library Association and Institutions*), in seguito ha promosso l'introduzione e la diffusione del formato **UNIMARC** (***U**Ni-**v**ersal **M**Achine **R**eadable **C**atalogin*). Si tratta di un formato che, nato nel 1977, è stato adottato quale standard dalla Comunità Europea per i progetti di cooperazione.

UNIMARC è "un formato di scambio allineato alla scansione degli elementi standard di ISBD, consente l'accoglienza e la coesistenza di diversi linguaggi catalografici, essendo stato che la compatibilità tra diversi formati è legata soprattutto alle diverse codifiche assegnate ai dati bibliografici in riferimento alle differenti prassi di catalogazione".¹⁰⁵

A partire dalla seconda metà degli anni Novanta MARC è stato gradualmente sostituito da **EAD** (***E**ncoded **A**rchival **D**escription*), un linguaggio di marcatura che possiede una struttura simile a quella delle risorse elettroniche accessibili in Internet. Il linguaggio EAD è modellato su SGML (Standard General Markup Language), uno standard per la gestione delle informazioni adottato dall'ISO, SGML è oggi rimpiazzato dall'XML.

¹⁰⁴ La versione italiana del MARC, l'ANNAMARC nacque negli anni '70 per automatizzare la Bibliografia Nazionale italiana (BNI). Essa fu utilizzata a tale scopo solo dal 1975 al 1984; dal 1985 la BNI è stata prodotta in formato UNIMARC.

¹⁰⁵ Solimine Giovanni, *Gestire il cambiamento*, Editrice Bibliografica, Milano 2003.

EAD è stato sviluppato principalmente per la creazione di inventari, soprattutto sotto forma di basi dati (spesso molto strutturate), ed ideato per rispondere alle esigenze di analisi a livello di fondo, sottofondo, serie, sottoserie, unità archivistica, etc. fino al singolo documento.

Può anche essere utile per la codifica di descrizioni archivistiche meno analitiche (es. guide ai fondi). E' per sua natura estremamente flessibile e personalizzabile, ciò lo rende adatto alla descrizione di archivi che ricercano un singolo item (fondi fotografici o di manoscritti).

L'ultima versione è del 2002, contiene 146 elementi¹⁰⁶. Ad esempio l'elemento titolo è modellato sullo standard *TEI*.

Esempio di elementi EAD per descrivere un titolo¹⁰⁷:

```
<filedesc>
  <titlestmt>
    <titleproper>
      Inventory of the Kingsley Amis Papers, <date>1941-1995</date>
    </titleproper>
    <author>Processed by Sara S. Hodson; machine-readable
      finding aid created by Xiuzhi Zhou</author>
    </titlestmt> . . .
  </filedesc>
```

In merito ai metadati utilizzati per descrivere raccolte bibliografiche, particolare rilievo riveste ***Dublin Core Metadata***, nato nel 1995, nel corso di un convegno promosso dall'OCLC (*Online Computer Library Center*), quale standard descrittivo di ogni risorsa disponibile in Rete.

Lo standard, che consente agli autori delle risorse di garantire per ciascun documento un livello minimo di autocatalogazione¹⁰⁸, permette la realizzazione di successive eventuali indicizzazioni e l'implementazione di un'architettura per i metadati. Nel 2005 è stato pubblicato il documento *Guidelines for Enco-*

¹⁰⁶ Elenco degli elementi che contiene lo standard EAD
<http://www.loc.gov/ead/tglib/appendix_d.html>.

¹⁰⁷ Fonte: <<http://www.loc.gov/ead/tglib/elements/titlestmt.html>>.

¹⁰⁸ Lo standard del *Dublin Core* individua alcuni elementi descrittivi del documento di tipo bibliografico, quali il titolo, l'autore, il soggetto, le parole chiave, l'abstract, l'editore, la data, la lingua ecc.

ding Citation Information in Dublin Core Metadata come *DCMI Recommendation*. Nello stesso anno si è lavorato allo sviluppo delle *Guidelines for national standardization of Dublin Core*.

Le principali aree di interesse nel 2004-2005 hanno riguardato strategie e priorità individuate nell'architettura DCMI (*Dublin Core Metadata Initiative*), al fine di rendere coerenti le applicazioni dello *standard* con le *DCMI Recommendation*: per implementare *Dublin Core* in XML, sono state revisionate le *Guidelines*, con l'aggiunta di nuovi termini ed il livello di descrizione delle collezioni.

In Italia, l'ICCU ha curato la traduzione italiana della versione del formato e, nel 2000, nell'ambito del progetto ONE-2 (*Opac Network in Europe 2*), ha preparato la mappatura tra i quindici elementi del *Dublin Core Metadata Element Set Reference Description* (versione 1.1), ed il formato UNIMARC.¹⁰⁹

Il set minimo (*Dublin Core Metadata Element Set*), proposto nel 1996 e noto anche come *Core dei 15 elementi della descrizione*, si compone di 15 tag ciascuno dei quali è definito utilizzando 10 attributi ricavati dalla norma ISO 1179. Il *set* è rimasto stabile, sebbene si sia poi anche esteso a sottoelementi o qualificatori. le caratteristiche del set sono:

- ✿ semplicità di utilizzo (in quanto si rivolge sia a specialisti che a non catalogatori)
- ✿ armonizzazione degli standard per i metadati
- ✿ comprensibilità dei descrittori
- ✿ flessibilità formale relativa allo sviluppo di ulteriori applicazioni
- ✿ interoperabilità semantica che consente la descrizione dei documenti appartenenti a qualsiasi disciplina e stabilisce una rete di dati tra loro concordati (quanto a significato e a valore)
- ✿ promozione dell'attivazione di una infrastruttura a livello internazionale

I 15 *metatag* sono ripartiti in 3 gruppi come mostra la *tabella* seguente:

¹⁰⁹ Data la complessità dei problemi sorti nella comparazione dei dati dal Dublin Core al formato bibliografico (la definizione e la formalizzazione dei dati bibliografici si basa su regole di catalogazione), nella mappatura sono stati previsti valori diversi in UNIMARC (per elementi qualificati o non qualificati).

Content	Intellectual Property	Instantiation
Title	Creator	Date
Subject	Publisher	Type
Description	Contributor	Format
Source	Rights	Identifier
Language		
Relation		
Coverage		

Tabella 7: Schema del *Dublin Core Element Set*.

In particolare:

- ▶ *Title*: indica il nome della risorsa
- ▶ *Creator*: indica il responsabile intellettuale della risorsa
- ▶ *Subject*: indica il contenuto concettuale della risorsa
- ▶ *Description*: fornisce una descrizione del contenuto della risorsa
- ▶ *Publisher*: indica un'entità responsabile della produzione della risorsa
- ▶ *Contributors*: elenca gli eventuali autori secondari della risorsa
- ▶ *Date*: indica la data di pubblicazione della risorsa
- ▶ *Type*: indica la natura o il genere della risorsa
- ▶ *Format*: indica la manifestazione fisica o digitale della risorsa
- ▶ *Identifier*: indica un riferimento univoco alla risorsa
- ▶ *Language*: indica la lingua nella quale è espresso il contenuto della risorsa
- ▶ *Relation*: indica eventuali relazioni della risorsa con altre risorse correlate
- ▶ *Coverage*: indica l'estensione o scopo del contenuto della risorsa (copertura spaziale o temporale)
- ▶ *Rights*: indica le informazioni sui diritti esercitati dalla risorsa.

Ciascun elemento è definito mediante l'utilizzo di un insieme di dieci attributi dello standard ISO/IEC 1179 per la descrizione dei dati.

L'**Open Archive Initiative (OAI)** per lo scambio dei dati adotta Dublin Core, che fornisce gli elementi identificatori; si tratta di un sistema che presenta un'architettura *Harvest* in un meccanismo di scambio tra *gatherer* e *brokers*.

Il *gatherer* estrae dalle collezioni le informazioni indicizzate e le trasmette a programmi chiamati *brokers* in un formato standard, attraverso un protocollo standard; il *broker* costruisce un indice di informazioni circa le varie collezioni. Il protocollo *Open Archives Metadata Harvesting (OAMH)*, nel sistema OAI¹¹⁰, definisce il meccanismo per la raccolta dei dati contenenti i metadati dei vari *repositories*.

Fondamentale risulta la costruzione di mappature tra metadati o mappature semantiche (i *data providers* espongono i propri metadati attraverso un protocollo HTTP).

3.3 WorldCat e Google Book

WorldCat è la più grande biblioteca al mondo, realizzata da OCLC con l'obiettivo di connettere le persone alle biblioteche ed alle loro collezioni (non ai solo libri, ma alle collezioni elettroniche e digitali).

WorldCat.org ti permette di ricercare le collezioni delle oltre 10.000 biblioteche che partecipano alla cooperativa in OCLC.

OCLC, si occupa del servizio e l'organizzazione di ricerca senza scopo di lucro, opera costantemente per ridurre i costi di informazioni, identificare i problemi e le tendenze sulla circolazione delle risorse, ed il ruolo sociale delle biblioteche, al fine di sviluppare nuove tecnologie e standard per l'interscambio cooperativo. Possiede numerosi canali di adesione che fungono da orientamento per OCLC, che cerca di impegnarsi in un dialogo con i leader del settore, per favorire la collaborazione e definire il futuro della biblioteconomia.

Alcune statistiche e fatti rilevanti che permettono di comprendere la portata dello strumento realizzato:

- 🌐 Numero di biblioteche rappresentate sul web: 71,000
- 🌐 Numero di record bibliografici contenuti: 150,223,241
- 🌐 Ogni quanto si aggiunge un record: ogni 10 secondi
- 🌐 Numero di holdings: 1,477,319,838
- 🌐 Numero paesi rappresentati: 112

¹¹⁰ OAI distingue tra i *data providers* ed i *service providers*, due settori che dialogano attraverso lo scambio delle informazioni grazie all'approccio *metadata harvesting*: i *servers* che contengono i metadati e i servizi (*service providers*) raccolgono i dati "esposti" dai *data providers* organizzandoli in servizi rispondenti ai bisogni dell'utenza con archivi di documenti.

- 🕒 Numero di lingue e dialetti rappresentati: più di 470
- 🕒 Ogni quanto spesso si inoltra una richiesta per WorldChat Resource Sharing: 4 secondi
- 🕒 Quanto spesso WorldCat è utilizzata dalla prima ricerca: ogni secondo

WorldCat collabora con Google Book: è l'idea di OCLC per aumentare la scoperta e l'accesso alle biblioteche e alle loro collezioni in tutto il mondo. Gli utenti usano WorldCat.org per individuare, valutare, liste di materiali per la biblioteca dal web. Fornisce un modo semplice per connettersi ed avere accesso al full text delle opere ritenute interessanti grazie a Google Book. Con questa nuova funzione, gli utenti saranno in grado di localizzare la risorsa WorldCat.org, visualizzare il testo presentato da Google tramite le sue API, e individuare una copia della voce nella biblioteca più vicina a loro.

OCLC ha iniziato la sua integrazione con il Book Search funzione Dynamic Link, che permette agli utenti di navigare ad un'anteprima Ricerca Libri quando è disponibile. Una volta localizzata, OCLC ha utilizzato il Embedded Viewer API per l'implementazione di un'anteprima della pagina che viene visualizzata sotto l'immagine della copertina del titolo ricercato in WorldCat, (come mostra la figura sottostante). La pagina che mostra il testo mantiene la navigazione WorldCat, le informazioni sul libro, e lo stile del sito.



Da un punto di vista tecnico, l'aggiunta di un visualizzatore integrato per il libro è stato abbastanza veloce cosicché hanno creato la nuova pagina in poche ore: si è trattato di aggiungere un CSS personalizzato nell'intestazione, il passaggio di una dinamica da codice ISBN al numero di OCLC, e l'aggiunta di un linguaggio di configurazione in JavaScript che chiamasse le API.

I record bibliografici sono taggati con XML, proponiamo un esempio che ne esemplifica la struttura:

```
<record>
<url>http://worldcat.org/wcpa/oclc/70043491</url>
<title>For one more day</title>
<author>Mitch Albom</author>
<edition>1st ed.</edition>
<publisher_data>
<location>New York</location>
<publisher>Hyperion</publisher>
<year>c2006</year>
</publisher_data>
<language>eng</language>
<oclc-number>70043491</oclc-number>
<isbn>1401303277</isbn>
<lccn>2006-18824 2006018824</lccn>
<ddc>813/.6</ddc>
<literary-form>Fiction</literary-form>
<genre>Fiction</genre>
<document-type>Book</document-type>
<subject>Mothers and sons</subject>
<subject>Mothers (Death)</subject>
<subject>Suicidal behavior</subject>
<subject>Psychological fiction. (Psychological fiction.)</subject>
<library-count>839</library-count>
<holding-countries>DE, HK, CA, SG, GB, US</holding-countries>
</record>
```

Fonte: <http://www.worldcat.org/partnership/harvestset/default.jsp>

Si può notare che usano per identificativo l'ISBN (che gli serve anche per collegare la risorsa a Google Book) oltre ad un loro codice identificativo "oclc-number". Anche GoogleBook usa una struttura simile, con tag XML per taggare le sue risorse. In particolare il Book Search API dati utilizza il Dublin Core Metadata Element Set sotto il namespace DC per fornire informazioni standard su ogni voce della rubrica.

Esempio: <dc:creator>Jane Austen</dc:creator>¹¹¹.

¹¹¹ DCMI Namespace for the Dublin Core Metadata Element Set, Version 1.1, <<http://dublincore.org/2008/01/14/dcelements.rdf>>.

Part III

Il libro nella rete

“[I]l testo viene prima del libro, sta fuori dal volume stampato. I nuovi media sono mezzi di diffusione della parola scritta. Ma se il mezzo può cambiare, la scrittura resta.”

Emilio Tadini, Fiera del Libro di Torino, Area Interattiva, Maggio 2000

Libro e la rete

La storia del libro e dell'industria libraria ad esso collegata, che è nata in seguito all'invenzione della stampa a caratteri mobili e migliorata attraverso il progresso delle tecnologie di produzione, è segnata da trasformazioni significative che modificano gli atteggiamenti nei confronti del medium e l'azione della lettura. Nel Novecento si assiste ad una crescente integrazione fra la carta stampata e i mass media, fenomeno all'interno del quale si trovano le origini degli sviluppi attuali. Un contenuto è trasmesso tramite molteplici canali, che offrono contributi differenti, espressione delle peculiarità del mezzo. Una storia narrata in un libro può essere oggetto di transcodifica e diventare una trama di un film proiettato al cinema ed in seguito in televisione. Ma può tradursi anche in un fumetto, un videogioco, una trasmissione radiofonica.¹¹²

Il libro ed internet, in base alla soddisfazione di utilizzo, rappresentano due casi primari. E' possibile ipotizzare uno sviluppo sempre maggiore di questo binomio che, se opportunamente alimentato, può produrre degli effetti positivi sui vari segmenti della filiera libraria. La produzione di e-book e di opere on-demand, disponibili on line e visibili attraverso le vetrine digitali, per la loro natura consentono la creazione di contenuti alternativi, che soddisfano

¹¹² “[...] il fenomeno che si va delineando è l'esplosione di bisogni culturali e comunicativi sempre più articolati, in un quadro in cui - dal punto di vista del consumo - sembra imporsi più una logica dell'integrazione tra media e generi di spettacolo, che l'antagonismo competitivo del recente passato. Del resto anche dal punto di vista dell'offerta le logiche predominanti sul mercato si indirizzano sempre più verso prodotti culturali a declinazione multimediale”. Mario Morcellini, *Il secondo Novecento italiano come "Evo medio"*, in Mario Morcellini e Paolo De Nardis, *Società e industria culturale in Italia*, Roma, Meltemi, 1998.

una certa nicchia di pubblico concentrata su testi specifici, poco vantaggiosi per il mercato tradizionale, ma importanti dal punto di vista della conservazione e della circolazione del patrimonio della conoscenza. Per questa ragione sono in grado di orientare con più facilità la proposta editoriale verso obiettivi di carattere culturale piuttosto che economico. I file di testo conservati nei database evitano i problemi legati ai costi di immagazzinaggio, all'incapacità da parte delle librerie di assorbire la quantità di titoli prodotti (nel 2006 sono 61.440, per un totale di 268.097.000 copie, di cui 37.991 sono novità che hanno tiratura complessiva di 153.866.000 copie),¹¹³ alle conseguenti rese e al possibile macero.

La vendita dei libri può essere agevolata dal canale internet sia in termini di rapporto tra il distributore e la libreria (*business to business*) che di relazione tra il venditore e l'utente (*business to consumer*). Nel primo caso i portali dei maggiori distributori sono diventati strumenti di servizio che facilitano l'attività di comunicazione con i clienti. Il comune acquirente, invece può avvalersi delle librerie elettroniche, in grado di soddisfare un ordine senza limiti di spazio o di tempo, scegliendo all'interno di un catalogo che dispone di un ampio numero di titoli, ma anche dei siti degli stessi editori che offrono un servizio di e-commerce o rimandano tramite un link ad un negozio on line. A questo si aggiunge la possibilità di poter acquistare libri, superando i possibili ostacoli derivati dall'insufficiente distribuzione dei punti vendita sul territorio, dalla mancanza di tempo per recarsi in un negozio, fino ad agevolare coloro chi ha impedimenti fisici.

In merito alla fruizione, quindi, i vantaggi legati al canale internet possono essere molteplici: internet può rappresentare un valido strumento per gli editori e le librerie che, oltre alla possibilità di fruire di servizi che agevolano la loro attività (portali dei distributori, software per la gestione degli ordini), comunicano e sponsorizzano attraverso newsletters, blog, forum, i prodotti trattati, l'attività svolta, gli eventi organizzati, le offerte in corso. In particolare, per gli editori la rete può fungere da canale di vendita diretto del proprio catalogo; per le librerie rappresenta un'ulteriore formula di avvicinamento al cliente.

¹¹³ ISTAT, *La produzione libraria*, Anno 2006, Roma 2008, tavola 1, <http://www.istat.it/dati/dataset/20080415_00/>.

4.1 Le case editrici e l'uso delle rete

In Italia i primi esperimenti significativi di utilizzo di internet nel settore dell'editoria libraria risalgono a fine anni Novanta. Fino al 1999 le imprese mettono in rete, a disposizione degli utenti, il catalogo dei loro titoli in commercio e forniscono alcune notizie di carattere generale relative all'azienda. Un esempio è la casa editrice Guaraldi che nel 1996 inizia la sua esperienza sul mondo web, pubblicando on line i propri titoli e sperimentando le tecniche del print on demand. In questo periodo sono poche le aziende editoriali predisposte al commercio elettronico, che implica un sistema organizzativo complesso, e che sfruttano a pieno le peculiarità della rete, in particolare la sua interattività.

A partire dal 2000 la maggiore conoscenza di internet e delle potenzialità ha consentito alle imprese di avviare un'attività di informazione e di comunicazione del libro on line.¹¹⁴ Risale a questo punto la prima ricerca condotta dall'Ufficio studi dell'Associazione Italiana Editori¹¹⁵ con l'obiettivo di analizzare l'uso della rete da parte delle case editrici e le trasformazioni verificatesi nel settore in seguito alla comprensione delle potenzialità del web. Il periodo oggetto di indagine è la seconda metà degli anni Novanta. Nel 1995 le imprese editoriali che dispongono di un sito internet rappresentano un numero irrisorio: 14. Il dato cresce in maniera significativa negli anni successivi: nel 1999 diventano 388, nel 2000 475 (+22,4%).

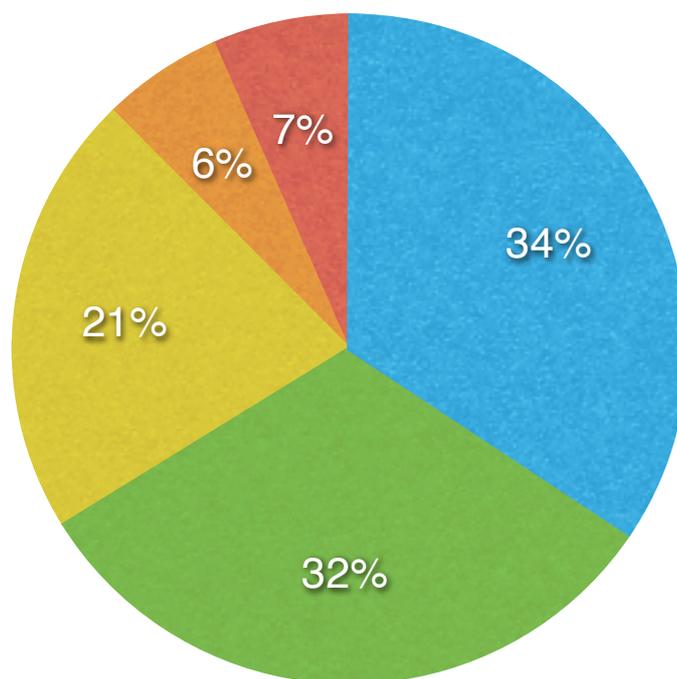
Ci sono cinque modalità di presenza on line delle case editrici:

- ▶ *Home page*, unica pagina che ne annuncia l'esistenza in rete
- ▶ *Vetrina*, presentazione dell'azienda, bassa interattività
- ▶ *Negozi Virtuale*, permette l'e-commerce
- ▶ *Casa editrice interattiva*, medio o alto grado di interattività
- ▶ *Portale*, sito che, oltre a gestire prodotti e servizi, si propone come snodo di indirizzi affini.

¹¹⁴ Giuliano Vigni, Rapporto 2000 sull'editoria italiana, in AIE, Catalogo degli editori italiani 2001, Milano, Editrice Bibliografica, 2000.

¹¹⁵ AIE, *Il futuro dell'editoria in rete. La presenza delle case editrici su internet: editori vs fornitori?* a.c. di Giovanni Peresson e Paola Mazzucchi, Milano 2000.

Tipologia sito internet: presenza delle case editrici in rete



- Vetrina
- Negozio Virtuale
- Interattiva
- Home
- Portale (Assente al momento)

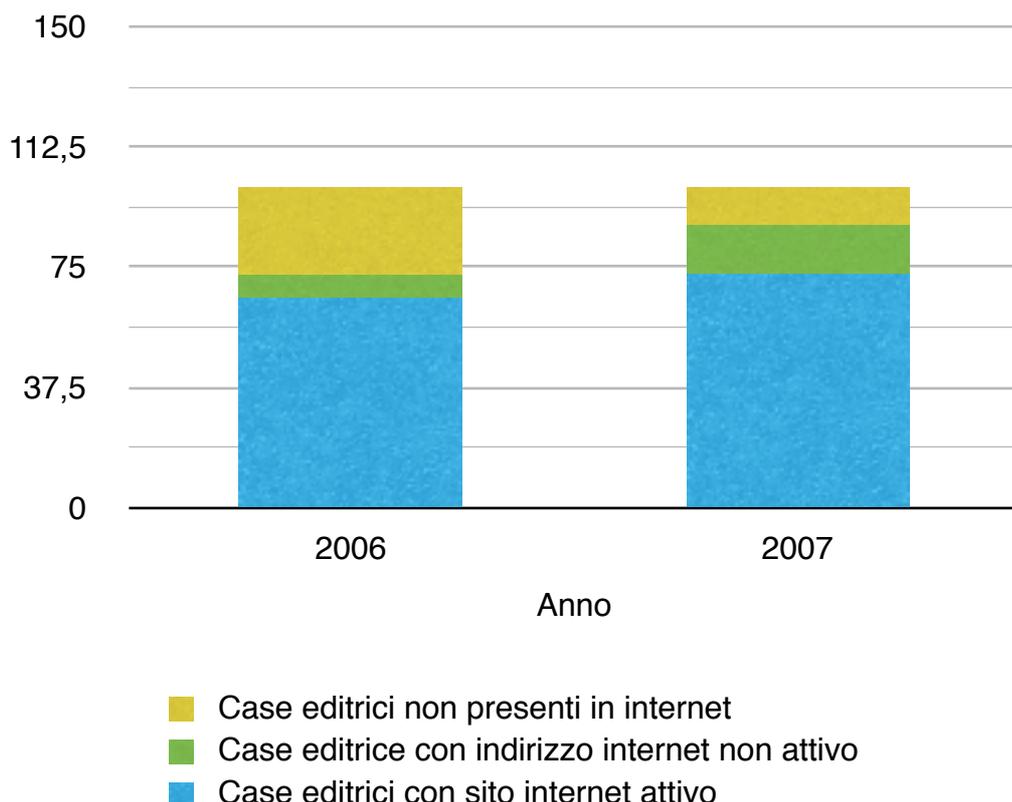
Nel 2000 la tipologia più adottata dalle case editrici è la “Vetrina”, seguita dal “Negozio Virtuale”, il grafico a torta mostra le percentuali per tipologia.

La modalità di esistenza in internet muta anche in base alla dimensione dell’azienda (piccola, media, grande), stabilita in termini di fatturato. Già nel 2000 è evidente una maggiore propensione da parte delle imprese più piccole ad offrire attraverso il loro sito la possibilità di effettuare acquisti. E’ il caso del 45,7% delle case editrici con un giro di affari inferiore a 2,6 milioni di euro, seguita dal 33,3% delle aziende con un fatturato compreso tra 5,2 e 10,3 milioni di euro, contro il 6,7% delle grandi imprese che registrano vendite superiori a 10,3 milioni di euro.

I portali restano poco diffusi, in particolare sono le case editrici con i fatturati più elevati a essere rappresentate in rete con siti che, oltre a gestire prodotti, si propongono come snodo di costellazione di pagine web affini. Il 16,7% dei siti di aziende con un giro di affari superiore a 10,3 milioni di euro è un portale. Seguono il 10,7% delle imprese di piccola grandezza e il 5,2% delle imprese medie.

Il 53,2% delle aziende pratica il commercio elettronico dal sito della casa editrice, offrendo la possibilità di acquistare direttamente i libri, cd-rom, software. Il 5,8% rinvia tramite link a librerie virtuali, come IBS, Internet BookShop. Per ciò che concerne il pagamento on line, la forma più diffusa è il contrassegno (64,6%, seguito dall'uso della carta di credito 47.6%).

Case editrici di narrativa presenti in Internet. Anni 2006 - 2007



Analizzando la situazione attuale si notano delle analogie, sebbene il panorama abbia subito una decisa trasformazione. In pochi anni la presenza delle case editrici in rete è cresciuta considerevolmente. Il grafico mostra le percentuali della presenza o meno delle case editrici in internet nel periodo 2006-2007. Nello specifico nel 2007 gli editori con un sito internet sono 5.143 , su 8.373 aziende, nel 2006 erano soltanto 475. L'incremento è evidente anche nel solo arco di un anno: rispetto al 2006 è del 13,4%.

Da una ricerca svolta nel 2007, che ha considerato gli editori che si occupano della produzione di narrativa, deriva che il 72,9% di essi possiede un dominio attivo.¹¹⁶

Per quanto riguarda la distribuzione, contrariamente quanto si potrebbe pensare, i nuovi canali non costituiscono una minaccia. I canali come le librerie su internet, avendo bisogno di un servizio molto efficiente, devono rivolgersi ai grossisti o ai distributori. Non sono come i nuovi canali di franchising della grande editoria, che creano problemi alla distribuzione perché tendono a scavalarla. La vendita di libri on line spingerà infatti verso una sempre maggiore efficienza della distribuzione: mentre la libreria tradizionale si può accontentare di tempi anche lunghi, la libreria su internet vive solo sugli ordini diretti dei consumatori, è praticamente sempre in “rottura di stock” ed ha quindi sempre bisogno di un rifornimento costante ed immediato.

Molte case editrici che non espongono una vendita diretta di e-commerce sul proprio sito si possono avvalere delle librerie on line, i vantaggi della rete permangono: allargamento della proposta, raggiungimento di migliaia di potenziali acquirenti, vetrina di accesso facile e veloce.

4.2 Il libro: dal cartaceo al digitale

Un libro è “un oggetto formato da un insieme di fogli contenenti o meno un testo e riuniti da una legatura o una broccatura”,¹¹⁷ che si articola lungo un numero preciso di pagine, suddiviso in capitoli e paragrafi, con un indice che permette l’individuazione degli argomenti. Ed è riconoscibile dal suo codice ISBN (International Standard Book Number), da poco ha associato anche un

¹¹⁶ Fonte ISTAT, *La produzione libraria. Anno 2006*, Roma 2008, <http://www.istat.it/dati/dataset/20080415_00/>, tavola 7.

¹¹⁷ “[...] nelle lingue latine, la parola deriva dal latino *liber*, termine che indica la pellicola di un albero compressa fra la corteccia esterna e il legno propriamente detto, ovvero un primo supporto scrittorio. La stessa osservazione è valida per le lingue germaniche, dove la parola deriva dall’antico alto tedesco *bokis* (ingl. *book*, ted. *buch*), termine che designa il faggio. E in greco il termine libro è intesa come *biblon* (βιβλίον) derivato da *bilos* (βιβλος ο βιβλος), il nome del papiro egiziano: da qui innumerevoli altri derivati come *biblioteca* (βιβλιοθήκη, etimologicamente l’armadio dei libri), come il libro la sacro (la Bibbia) e il nome del libraio in latino medievale, *bibliopola*; Frederic Barbier, *La storia del libro. Dall’antichità al XX secolo*, Bari, Edizioni Dedalo, 2005, pp. 11-12.

codice EAN che rappresenta il libro come oggetto per la commercializzazione.



L'ISBN identifica quindi l'asset intangibile, l'opera, mentre l'EAN rappresenta la risorsa dal punto di vista fisico, l'asset tangibile.

L'ISBN in rete viene sostituito dal DOI.

Il DOI

Il **Digital Object Identifier** è uno standard che consente di identificare persistentemente, all'interno di una rete digitale, qualsiasi oggetto di proprietà intellettuale e di associarvi i relativi dati di riferimento, i metadati, secondo uno schema strutturato ed estensibile.

Il DOI è stato definito "il codice a barre per la proprietà intellettuale: analogamente ai codici a barre dei prodotti fisici, l'utilizzo dei DOI costituisce un valore aggiunto e consente il risparmio di risorse lungo l'intera catena produttiva e commerciale.

Il DOI si distingue dai comuni indicatori internet come gli URL in quanto identifica un oggetto direttamente, quale entità di prima classe, e non semplicemente attraverso qualche suo attributo come il luogo dove l'oggetto è collocato.

Il DOI si distingue dai comuni identificatori di proprietà intellettuale come quelli legati agli standard bibliografici (ISBN, ISRC, ecc.) in quanto è imme-

diatamente azionabile in rete e utilizzabile per lo sviluppo di servizi specifici, quali motori di ricerca, certificazioni di autenticità, ecc.

Il sistema di numerazione DOI Name utilizza una sintassi standardizzata: *ANSI/NISO Z39.84* del 2000.

Il DOI Name non dice nulla riguardo l'entità che ha come oggetto (è una stringa opaca), l'unico modo per avere informazioni sull'entità è leggere i metadati. Se si modificassero i metadati o qualche altra proprietà dell'entità, quindi, il DOI Name resterebbe lo stesso (Persistenza del DOI Name).

Il DOI Name è composto da 2 componenti separati da una slash:

 **prefisso:** tutti i nomi DOI iniziano con 10., questo per distinguere un DOI da un'altra qualsiasi implementazione del Handle System. La stringa seguente, delimitata dal "." è l'identificativo dell'organizzazione che ha registrato l'elemento. Ogni organizzazione può avere più di un numero, e non c'è un limite massimo. L'uso di differenti suffissi per una organizzazione può tornare utile per mantenere l'unicità del DOI Name. Il secondo elemento può essere ulteriormente diviso in infiniti prefissi secondari. Il DOI Name non dà nessuna informazione e sarebbe sbagliato derivare il proprietario di un dato elemento a partire dal secondo elemento del prefisso DOI. Infatti in caso di passaggio di proprietà il DOI Name, essendo persistente, non viene modificato.

Esempio di prefisso DOI: 10.11234.12.

 **suffisso:** insieme al prefisso forma una stringa unica (ogni DOI Name deve essere unico). Il suffisso come il prefisso può essere una combinazione qualsiasi di caratteri alfanumerici. Esistono due modi per creare un suffisso: o le entità sono già numerate in qualche modo, o le unità non sono ancora numerate. Nel primo caso il DOI Name può essere utilizzato per entità che hanno già schemi standard di numerazione (es. ISBN per i libri) o può essere che l'ente registrante abbia un proprio sistema di numerazione interno. Si può utilizzare il nome definito dallo standard o dall'ente registrante come suffisso. Ovviamente quando si percorre questa strada le applicazioni che leggono il DOI Name non attribuiscono alcun significato al suffisso, in quando al DOI Name non è associato alcun significato, ma si può creare un'applicazione che estrapoli dal DOI Name il suffisso e lo interpreti secondo gli standard corretti.

Non vi è un limite sulla lunghezza di queste due parti, ciò permette una disponibilità infinita di DOI Name.

L'e-book

Il termine e-book è utilizzato per indicare sia una pubblicazione su supporto digitale che gli strumenti con cui vi si accede. Il prodotto si distingue essenzialmente per il suo contenuto e per le sue caratteristiche tecniche (il formato di codifica utilizzato per rappresentare il testo, il programma di lettura che ne consente l'accesso, il dispositivo hardware usato come interfaccia fisica per acquisire, conservare e leggere l'e-book). Esso può essere fruito attraverso appositi dispositivi oppure strumenti informatici come computer o palmari, sia in modalità on line che offline. Internet è il veicolo privilegiato per la loro circolazione.

Nel 2000 il Consorzio internazionale Open eBook Forum (OEBF), organizzazione no-profit, nato nel 1998 come un'organizzazione che riunisce aziende informatiche e editoriali (tra cui Adobe, IBM, McGrawHill, Microsoft, Mondadori, Random House), centri di ricerca e singoli esperti, con lo scopo di definire gli standard tecnologici per il settore dell'e-book.

Nel 2005 l'OEBF ha cambiato il suo nome in IDPF,¹¹⁸ si basa un file con estensione **.epub**, è scritto secondo il linguaggio XML. Il file è composto da tre open standards:

- ✿ OPS Open Publication Structure
- ✿ OPF Open Packaging Format
- ✿ OCF Open Container Format, prodotto dall'IDPF

Tale formato consente agli editori di produrre e inviare un singolo file di pubblicazione attraverso la distribuzione e offre ai consumatori interoperabilità software/hardware. The Open eBook Publication Structure (OEB), originariamente prodotto nel 1999, è il precursore di OPS. Scaricato il file .epub basterà poi installare un programma per la visualizzazione a schermo come: *Adobe Digital Editions, FB Reader, Mobipocket Reader* e molti altri.

¹¹⁸ <<http://www.idpf.org/>>.

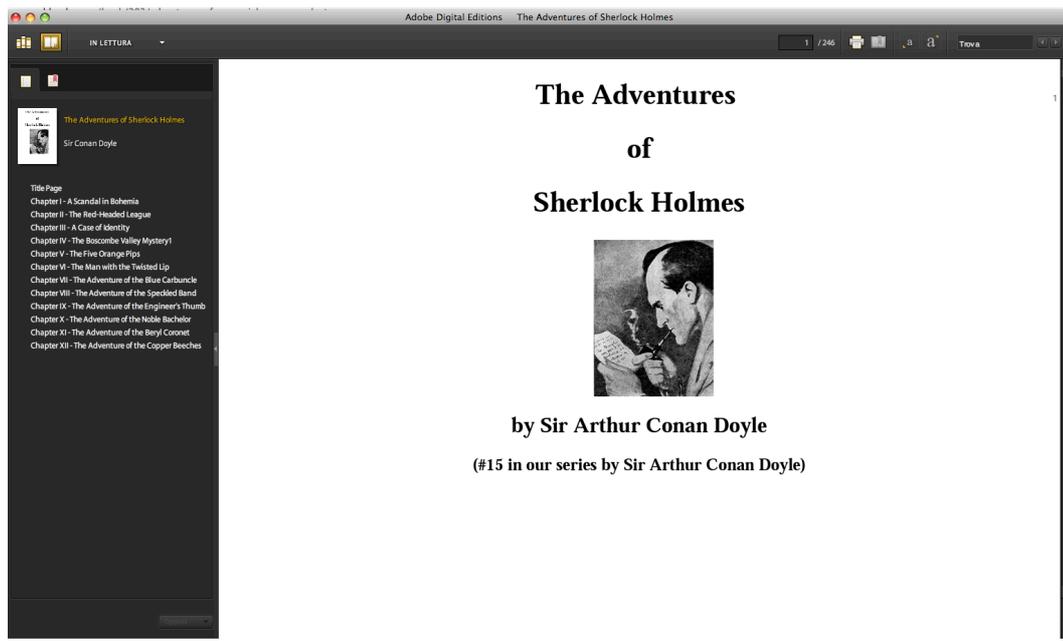


Figura 17: file con estensione .epub visualizzato con Adobe Digital Editions.

La nascita del digital book modifica la tradizionale costruzione di un testo. I contenuti pubblicati in un libro cartaceo sono caratterizzati dalla loro sequenzialità. Un testo in formato elettronico può essere composto da ipertesti, che rimandano a documenti, anche ipermediali, legati tra di loro da collegamenti che il lettore attiva, creando un insieme autonomo di significati, che si snodano secondo un'architettura decisa in fase progettuale. L'interattività interrompe il fluire regolare della narrazione e consente una personalizzazione della lettura e dell'acquisizione dei contenuti.

Nonostante tutti questi aspetti favorevoli il libro cartaceo continuerà ad esistere e l'e-book dopo un primo boom iniziale, alla prima vendita del *Riding the bullet* di Stephen King, nel 2002 gli utenti intasarono la rete per averne una copia (al prezzo di 2,5 dollari), successivamente si è registrato un andamento che privilegia il download di e-book gratuiti. I dati diffusi dall'IDPF¹¹⁹ fanno sperare un andamento positivo che sta crescendo negli ultimi due anni negli Stati Uniti, grazie anche ad un miglioramento delle tecnologie connesse.

¹¹⁹ Nel periodo ottobre-dicembre 2007 il fatturato è di 8.200.000 dollari contro la cifra di 1.556.000 dollari registrata nel gennaio-marzo 2002.

4.2.1 L'editoria on-demand e il self-publishing

Lo sviluppo delle tecnologie digitali ha determinato la nascita della stampa su richiesta e la conseguente diffusione di case editrici che si avvalgono di questa tipologia di produzione. Tale tecnica consente la materializzazione su un supporto cartaceo di libri in formato elettronico, precedentemente archiviati in un database, e pubblicati solamente in seguito alla loro ordinazione. La prima impresa sorta in Italia che ha utilizzato esclusivamente questa modalità di lavoro è Lampi di stampa¹²⁰: ha un catalogo di circa 600 titoli, tra saggistica, narrativa, poesia, letteratura italiana, monografie e dispense generate in ambito universitario. Dedicava spazio agli autori esordienti, ai titoli fuori commercio, alle coedizioni, ai libri in formato elettronico.

La stampa digitale su richiesta, così come la pubblicazione on line di libri in formato elettronico, si presta alla produzione di opere il cui target è circoscritto a numeri contenuti, non essendo vincolata dalle tirature elevate che risultano economicamente vantaggiose nell'ambito della tradizionale tecnologia offset. Rappresenta quindi un'alternativa per quei testi che non sono coinvolti da un pubblico di massa o sono fuori catalogo; costituisce un'occasione per gli autori esordienti ed uno spazio necessario per la saggistica universitaria o per le ricerche riprodotte in ambito accademico, che troppo spesso hanno una scarsa visibilità. A tale proposito, sono stati avviati esperimenti di convergenza tra il mondo imprenditoriale dell'editoria libraria e l'università, sebbene siano necessari ulteriori sviluppi in questa direzione, un esempio è Edizioni ETS di Pisa. Anche Guerini e Associati ha legami con le Università, pubblicando dispense, saggi, ricerche su commissione (un esempio è la collana ad esse dedicata _QDI_ Quaderni del dipartimento di Italianistica, Università di Padova).

Il print on demand si inserisce in un contesto più ampio che riguarda i nuovi modelli editoriali sorti conseguentemente alla diffusione della rete. Sistemi che potrebbero orientare maggiormente le scelte di produzione verso ragioni di carattere culturale piuttosto che economiche ed agevolare la circolazione della conoscenza e del pensiero scientifico. Negli anni Duemila si è verificato un incremento del numero di case editrici che sfruttano questa tecnologia. Da

¹²⁰ <<http://www.lampidistampa.it/>>.

un'indagine effettuata nel 2007 dall'AIE, deriva che le case editrici che utilizzano questa tecnica sono 413, rappresentando l'11% delle imprese attive (3.658) e il 4.9% degli editori (su un totale di 8.373); rispetto al 2006 è stata registrata una crescita del 76%.

La crescita della pubblicazione di titoli riguarda soprattutto le novità sfruttando l'opportunità di verificare il successo: se il titolo ha un buon impatto allora l'editore considera se passare al tradizionale offset.

L'indagine rileva quattro opportunità derivate da questa tecnologia:

-  lavorare su nicchie di domanda altrimenti irraggiungibili
-  ridurre le giacenze di magazzino e i tempi di lavorazione
-  aumentare la flessibilità
-  gestire le code di stampa

Anche in ambito internazionale si assiste ad una crescita dell'interesse sul print on demand. Nel 2005 Amazon.com acquista la società CustomFlix Labs, e dà via al marchio CreateSpace¹²¹, che si occupa di editoria on demand (libri, cd, dvd). L'utente che vuole creare un proprio volume può liberamente crearlo, a colori o in bianco e nero, fissandone il prezzo e ottenendo *royalties* mensili in base alle vendite. La stampa avviene al momento della richiesta da parte dell'acquirente; ogni titolo riceve automaticamente un ISBN ed è reso disponibile attraverso il catalogo di Amazon e di CreateSpace Shop .

Tra i servizi offerti dall'editoria on demand compare il self-publishing. Un esempio di questa modalità di editoria è Lulu.com¹²². Da ottobre 2006 il portale è attivo anche in Italia, consentendo la libera produzione di libri, e-book, cd, dvd, calendari, report nonché la possibilità di una visibilità globale on line. Lulu mette a disposizione gli strumenti per formattare il contenuto digitale (ad un costo fisso da 5 a 10 euro e una commissione pari al 20% sui diritti di autore); la grafica, l'impaginazione e il prezzo sono scelti dall'autore. La spesa di produzione di un libro varia in base alla dimensione e alla tipologia di rilegatura. Lulu stampa su richiesta esclusivamente gli esemplari destinati alla vendita che avviene attraverso il proprio portale, ma anche tramite alcune librerie elettroniche inserite nel normale circuito di distribuzione (Amazon,

¹²¹ <<https://www.createspace.com/>>.

¹²² <<http://www.lulu.com/it>>.

Borders, Barnes&Noble). Nel sito di Lulu sono attivi blog, forum, laboratori tv, radio, funzionanti come una comunità virtuale che condivide, acquista, carica e scarica contenuti di vario tipo, quali testi, immagini, audio e video.

Anche ilmiolibro.it¹²³, del gruppo l'Espresso, dal maggio 2008 permette agli aspiranti scrittori di pubblicare e vendere il proprio manoscritto, e si può subito creare un preventivo in base alle pagine del libro e delle copie che si vuole stampare.

¹²³ <<http://ilmiolibro.kataweb.it/>>.

Part IV

Descrizione parte di progetto

Tirocinio presso Guerini e Associati

Il Tirocinio è stato svolto per la casa editrice Guerini e Associati con sede a Milano. Lo scopo del lavoro da realizzare era quello di creare un catalogo unico storico che riesumasse i tre cataloghi elettronici esistenti in modo da fonderli in uno, il quale mantenesse le informazioni presenti provenienti dalle diverse fonti. Una volta allineato il catalogo unico è stato codificato con gli standard TEI.

I cataloghi analizzati sono:

- ↳ l'anagrafica
- ↳ il catalogo on line della casa editrice
- ↳ il catalogo editori Alice
- ↳ ai quali si aggiungono dei cataloghi cartacei suddivisi per anno di pubblicazione (mi sono stati consegnati solo di alcuni anni, quelli che avevano in duplice copia).

Vi è una descrizione dettagliata riportata di seguito e una *tabella* riassuntiva che analizza nello specifico come sono organizzati tali cataloghi, quali campi possiedono, come sono riempiti e le corrispondenze in comune.

5.1 Catalogo anagrafica

Fornito in formato .xls, il catalogo contiene 3101 records.

Campi presenti: autore, codice alternativo, titolo, sottotitolo, edizione, data (solo anno), collana, collana (identica alla precedente, estende soltanto le sigle), num pagine, prezzo, area, settore, stato.

Descrizione riempimento campi:

- ★ **ISBN:** tutti hanno il codice
- ★ **Descrizione:** tutti hanno descrizione
- ★ **Autore:** riempiti tutti ad eccezione di circa 200 records.
- ★ **Titolo:** tutti hanno titolo
- ★ **Sottotitolo:** solo 111 records hanno il sottotitolo, gli altri sono vuoti.
- ★ **Edizione:** solo 34 elementi compilati, il resto vuoto.
- ★ **Data Pubblicazione:** libri dal 1987 al 2009
- ★ **Collana:** ha solo 288 campi riempiti, gli altri vuoti
- ★ **Collana:** ha solo 288 campi riempiti, gli altri vuoti
- ★ **Num pagine:** 1853 elementi hanno il num 0, 1113 elementi riportano il num della pagine, 135 elementi sono vuoti.
- ★ **Prezzo:** 225 elementi hanno il prezzo uguale a 0,00, una il resto ha un prezzo.
- ★ **Categoria:** ha impostato 1,2,3,4 come valori, a parte gli ultimi 20 elementi che hanno il valore 999.
- ★ **Ean:** 3068 elementi hanno il valore, 36 sono vuoti.
- ★ **Centro:** tutti hanno un valore che varia da 1,2,3,4 8,9.
- ★ **Area:** tutti (meno circa 10 elementi) hanno valore che varia da 0 a 5.
- ★ **Settore:** tutti (meno circa 10 elementi) hanno un valore.
- ★ **Stato:** tutti (meno circa 5 elementi) hanno un valore: F, R, D.

5.2 Catalogo on line del sito

Fornito in formato .xls, tale catalogo contiene 1526 records

Campi: ID, ID_I, edizione_da_ISBN, Edizione_da _catalogo, ISBN_I_simbolo, ISBN 1, ISBN 2, ISBN_Tot_simbolo, ISBN Totale, autore, a cura di, a cura di 2, titolo, sottotitolo, note, collana_codice, anno, edizione, numero pagine, jpg, evergreen, spinta settore, homepage, strillo, abstract, prezzo, prezzo in euro, ILL_rilegato_ecc, area di interesse1, area di interesse2, In_preparazione, copertina, area_di_interesse3, promo, promo2, url

Descrizione riempimento campi:

- * **ID_I:** numeri progressivi, tutti compilati
- * **ID:** numeri non progressivi, 635 hanno il valore, gli altri vuoti.

- * **Edizione_da_ISBN:** tutti gli elementi hanno il valore che varia: Cantieri Italia, Guerini Studio, Guerini e Associati, Guerini Scientifica.
- * **Edizione_da_catalogo:** campo uguale al precedente.
- * **ISBN_I_simbolo:** riempito con simboli come: \$, #.
- * **ISBN_I:** contiene la prima parte del codice ISBN (6 cifre, varia a seconda dell'editore) , compilato per tutti.
- * **ISBN_2:** ultime 4 cifre del codice ISBN, tutti hanno il valore.
- * **ISBN_tot_simbolo:** ISBN con simbolo.
- * **ISBN_tot:** codice ISBN per intero, tutti possiedono il codice.
- * **Autore:** tutti gli elementi hanno un autore.
- * **A_Cura_Di:** riporta solo la scritta (a_cura_di) per alcuni elementi.
- * **A_cura_Di2:** solo 8 elementi hanno il nome del curatore, gli altri sono vuoti.
- * **Titolo:** tutti gli elementi hanno il titolo.
- * **Sottotitolo:** 957 elementi riportano il sottotitolo, gli altri sono vuoti.
- * **Note:** circa 600 elementi hanno il valore nota, con informazioni sulla prefazione o a cura di.
- * **Collana:** 970 elementi riportano informazione della collana, gli altri sono vuoti.
- * **Anno:** tutti hanno il campo riempito, anno che va dal 1987 al 2009.
- * **Edizione:** 222 elementi hanno il valore edizione, gli altri sono vuoti.
- * **N_Pagine:** tutti gli elementi hanno il valore del numero della pagine.
- * **JPG:** codice immagine.
- * **Strillo:** 117 elementi hanno descrizione strillo.
- * **Abstract:** 744 elementi hanno impostato l'abstract, gli altri sono vuoti.
- * **Prezzo:** campo assegnato con simboli #####.
- * **Prezzo_in_euro:** tutti gli elementi hanno un valore in euro.

- * **Ill_rilegatura_ecc**: solo 46 elementi riportano se libro è ill. o ril.
- * **Area_di_interesse1**: tutti gli elementi hanno impostato l'area di interesse (es. Architettura arte e giardini, Economia e Management, Filosofia...).
- * **Area_di_Interesse2**: 270 elementi riportano una seconda area di interesse, diversa dalla prima.
- * **In_preparazione**: solo si o no, circa 20 elementi.
- * **Copertina**: colonna vuota.
- * **Area_di_interesse3**: vuota.
- * **Promo**: solo 15 elementi hanno la dicitura promo, il resto è vuoto.
- * **Promo2**: solo 15 elementi hanno un valore.
- * **url**: ha solo un indirizzo url.

5.3 Catalogo editori Alice

Fornito in formato .xls, contiene 1557 records.

Campi: ISBN-13, autore, titolo, curatore, traduttore, anno, ed.num, pagine, illustrazione, num. Volumi, collana, prezzo, disponibilità, scol., marchio editoriale, codice Alice.

Descrizione riempimento campi:

- ❖ **ISBN-13**: tutti hanno il codice ISBN-13, ovvero l'EAN.
- ❖ **Autore**: 1036 elementi hanno impostato il campo autore, il resto è vuoto.
- ❖ **Titolo**: tutti gli elementi hanno il titolo.
- ❖ **Cur.**: 620 elementi hanno il cur.
- ❖ **Trad**: 77 elementi hanno il trad.
- ❖ **Anno**: tutti hanno il valore anno, che varia dal 1994 al 2009.
- ❖ **Ed.num**: 117 elementi hanno il num ediz, il resto è vuoto.
- ❖ **Pag**: 1534 elementi hanno il valore numero pagine.
- ❖ **Ill**: 647 elementi riportano questa informazione (ril., br., ill...).
- ❖ **Num.Vol**: solo 3 elementi hanno il valore, il resto è vuoto.

- ❖ **Collana:** 597 elementi hanno il simbolo (), circa 900 elementi hanno riportato la collana, 25 elementi hanno il campo vuoto.
- ❖ **Prz:** tutti hanno indicazione del prezzo;
- ❖ **Disp:** tutti hanno indicazione della disponibilità (valori immessi: “C” oppure “E”).
- ❖ **Scol.:** tutto vuoto.
- ❖ **Marchio editoriale:** tutti hanno indicazione del marchio editoriale (Guerini e Associati, Guerini Scientifica, Guerini Studio).
- ❖ **Codice Alice:** 1507 elementi hanno il codice Alice.

5.4 Altre fonti

Quarta fonte sono le schede di produzione (consultazione in sede, sono cartacee). Mi sono fatta spedire una scansione, ho notato che contiene molte informazioni già presenti nei cataloghi sopra descritti (titolo, sottotitolo, autore, curatore, collana, marchio editoriale, data, prezzo..) più altri dati più dettagliati (recapito autore, tiratura, dimensioni, estetica..). Comunque mi hanno detto che non tutte sono compilate in ogni dettaglio e quelle più vecchie riportano meno dati.

Quinta fonte: catalogo cartaceo, possiedo solo alcuni anni che ricoprono neanche la metà dei 22 anni di attività.

5.5 Modifiche effettuate sui cataloghi

Uso di macro Find_Matches() per trovare duplicati.

Applicato al titolo, tra file Catalogo on line e catalogo IE Alice, risultati: quasi tutti i titoli sono in comune. Purtroppo apre una finestrella dove elenca tutti i titoli comuni ma non si possono contare né copiare in una nuova colonna.

Svolgimento, passi svolti

Siamo partiti da un punto di forza, un codice univoco che identifica ogni libro, il codice ISBN, EAN dal 2007. L'ISBN non solo identifica univocamente libri ma anche eBooks, software, e media.

Interventi prima dell'allineamento, pulizia dei dati:

☀ conversione dei numeri che apparivano in cifre arabe ma in realtà erano testo (ciò avviene quando si esportano dati da database). In particolare la tecnica utilizzata è stata la seguente:

- Immettere il valore 1 in una cella vuota.
- Selezionare la cella in cui digitato 1 e scegliere **Copia** dal menu **Modifica**.
- Selezionare le celle contenenti i valori da convertire in numeri.
- Scegliere **Incolla speciale** dal menu **Modifica. Operazione**, scegliere **Moltiplica**. Sotto **Incolla**, fare clic su **valori** e scegli **OK**.

☀ eliminare le celle vuote (nel catalogo anagrafica, una riga sì e una o era vuota, ciò avrebbe comportato un lungo lavoro manuale che avrebbe portato al possibile errore di cancellazione di qualche riga che conteneva un dato utile). Per questa operazione è stata utilizzata una macro:

- **Macro to delete rows**

- `'/=====/'`

`' Sub Purpose: Delect all blank ROWS within the active cell's`

`' Used Range`

`'`

`Public Sub DeleteBlankRows()`

`Dim dbMaxRow As Double, dbMinRow As Double, i As Double`

`Dim dbMaxCol As Double`

`Dim rng As Range`

`On Error Resume Next`

`'only look in used area of the worksheet where active cell is`

`Set rng = Selection.Parent.UsedRange`

`'calculate area to be searched for blank rows`

`dbMaxRow = rng.Rows.Count '# of rows in used area`

```
dbMinRow = rng.Cells(1, 1).Row '1st row in used area
dbMaxCol = rng.EntireColumn.Count '# of columns in used area
```

```
For i = dbMaxRow To dbMinRow Step -1
If IsError(rng.Cells(1, 1).Offset(i - 1, 0).EntireRow. _
SpecialCells(xlCellTypeBlanks).Count) Then
Else
If rng.Cells(1, 1).Offset(i - 1, 0).EntireRow. _
SpecialCells(xlCellTypeBlanks).Count = dbMaxCol Then
rng.Cells(1, 1).Offset(i - 1, 0).EntireRow.Delete
End If
End If
Next i
```

```
Set rng = Nothing
```

```
End Sub
```

```
'/=====/'
```

```
'
```

- ✿ Per verificare che tutti gli items del catalogo on line fossero presenti nel catalogo anagrafica, è stata utilizzata una formula:

-CERCA.VERT(B1;A2:A3104;FALSO) restituisce #VALORE se ISBN della colonna catalogo on line è presente in una qualsiasi posizione della colonna A (ISBN anagrafica), NON Valore altrimenti.

Tale formula è stata copiata e modificata manualmente per ogni riga del catalogo on line (cella B1, B2, B3... ecc). Il match è risultato al 100%, ovvero ogni items del catalogo on line era contenuto nel catalogo anagrafica.

- ✿ Una volta puliti i dati ho stilato una tabella che descrivesse dettagliatamente come erano riempiti i campi, quante caselle vuote, oppure quanti dati inesistenti. Questa operazione di preparazione è stata fondamentale per decidere come allineare i dati per poi definire il modello di codifica. Vi è allegato il documento.
- ✿ Per verificare invece che ogni items del catalogo Alice fosse compreso in anagrafica ho dovuto convertire il codice EAN (cat. Alice riporta solo il codice EAN mentre gli altri hanno prevalentemente il codice ISBN). Trovato il codice EAN per ogni items ho fatto una ricerca per

singolo dato e ho riportato una tabella che diceva se era presente in anagrafica e allo stesso tempo se era contenuto anche nel catalogo on line. Il match con il catalogo anagrafica è stato anch'esso al 100%.

- ✿ Il passo successivo è stato quello di allineare su ogni singola riga, le informazioni provenienti dai 3 cataloghi che si riferivano a quel dato oggetto. Le colonne che riportavano campi comuni come: num pagine, prezzo autore, titolo, isbn ecc. sono state allineate ed è stato controllato che tutto risultasse in linea con ciò che era scritto nelle diverse fonti. Di seguito una nota sulla incongruenze riscontrate e come ho operato nella risoluzione dei singoli casi.
- ✿ Conversione degli ISBN in EAN, si aggiunge il prefisso (978) e poi va calcolato il check-number, ovvero il numero di controllo che si calcola dalle altre nove cifre del codice ISBN, varia da 0 a 9 e per indicare il numero 10 viene utilizzata la cifra romana x.

Campi presi in considerazione nel catalogo unico finale: ISBN, AUTORE, COD. ALTERNATIVO, TITOLO, DATA, COLLANA, NUM PAGINE, PREZZO, CATEGORIA, EAN, CENTRO, AREA, SETTORE, NOME EDITORE, A CURA DI, NOTE, COD. COLLANA, RILEGATURA, AREA INTERESSE 1, AREA INTERESSE 2, TRADUTTORE, ED. NUM, NUM. VOLUMI, CODICE ALICE.

Incongruenze nei cataloghi

✍ Annotazione 1: alla riga 1663 di anagrafica, si trova il codice ISBN 8878029584 con EAN 9788878029583. Tale codice EAN se riconvertito dà il seguente ISBN 8878029580. Quindi abbiamo 2 ISBN per lo stesso EAN. Il primo ISBN che finisce con il 4 è presente in on line mentre gli altri 2 cataloghi hanno quello che finisce con lo 0. Questo è un problema del checknumber.

✍ Annotazione 2: nel catalogo on line il codice 8883354001 riporta i seguenti titoli: “Quaderni del piano territoriale” su una riga e sotto con lo stesso codice presenta anche il titolo “Linee guida per interventi di ingegneria naturalistica lungo i corsi d'acqua”, nel catalogo Alice lo stesso codice ISBN ha titolo: “Quaderni del piano territoriale. Vol. 20: Linee guida per interventi di ingegneria naturalistica lungo i corsi d'acqua.” E nel catalogo anagrafica: “LINEE

GUIDA PER INTERV.DI ING.”. Accorpo in un unico item, poiché è lo stesso libro ma con il titolo spezzato in 2.

 Annotazione 3: l'ISBN 8883356288 presente nel catalogo on line riporta il titolo “Quattro studi sulla vulnerabilità sociale”, mentre questo codice non è presente né in anagrafica né in Alice. Tale titolo esiste ma riporta il codice ISBN 8883352688. Pertanto il primo codice con suffisso 288 non verrà preso in considerazione perché errato.

Osservazioni

È interessante capire come organizzare al meglio un modello di codifica: i dati possono essere tutti codificati e poi una volta creata la struttura si possono utilizzare per ricerche diverse: ad esempio, dati utili per la gestione del magazzino; dati per sapere quali libri sono fuori catalogo; quali libri sono in catalogo pronti per la vendita e via dicendo.

Si può osservare che non tutti questi aspetti sono presi in considerazione dagli standard internazionali e invece sono utili per la gestione, il riutilizzo la fruizione su largo spettro. Per questo è sempre possibile definire nuovi schemi che arricchiscono quelli pre-esistenti ma che si affacciano su orizzonti ben più ampi.

Catalogo Anagrafica campi (3103 records):

ISBN	Tutti
Descrizione	Campo uguale al titolo
Autore	Tutti meno 200 item hanno il valore
Titolo	Tutti
Sottotitolo	Solo 200 campi (è comunque parte del titolo che è stato abbreviato)
Edizione	34 elementi
Data pubblicazione	Tutti
Collana	288 elementi
Num pagine	1835 hanno campo=0 1113 elementi hanno il num 135 vuoti
Prezzo	225 hanno prezzo = 0,00 Tutti gli altri hanno il valore

Categoria	Hanno impostato il valore 1,2,3,4 Ultimi 20 hanno valore = 999
Ean	3068 hanno il valore (n.b. quelli mancanti si possono ricavare dall'ISBN)
Centro	Tutti hanno valore che varia (1,2,3,4,8,9) (dati area amministrativa)
Area	Tutti valore da 1 a 5 (dati area amministrativa)
Settore	Tutti hanno valore (dati area amministrativa)
Stato	Tutti valore: F, R, D

Catalogo on line del sito (1526 records)

ID1	Tutti (num progressivi, gestione sito)
ID	635 valore, (gestione sito)
Edizione_da_ISBN	Tutti valori: Cantieri Italia, Guerini Studio, Guerini e Associati, Guerini Scientifica
Edizione_da_Catalogo	Tutti (uguale al campo precedente)
ISBN_1_simbolo	Tutti. Simboli \$ % #...
ISBN_1	Tutti. Contiene la prima parte del codice ISBN (6 cifre varia a seconda dell'editore)
ISBN_2	Tutti. Contiene ultima 4 cifre del codice ISBN
ISBN_tot_simbolo	Tutti. ISBN con simbolo
ISBN_tot	Tutti. Codice ISBN
Autore	Tutti
A_cura_di	Riporta scritta a_cura_di
A_cura_di2	8 compilati
Titolo	Tutti
Sottotitolo	957 elementi compilati
Note	600 elementi (info sulla prefazione o cura di)
Collana	970 elementi
Anno	Tutti (varia da 1987 a 2009)
Edizione	222 elementi
N_pagine	Tutti
JPG	Codice immagine
Strillo	117 hanno descrizione strillo (serve per il sito, mette notizia nella home)
Abstract	744 elementi
Prezzo	Hanno tutti ####
Prezzo in euro	Tutti
Ill_rilegatura	46 (dice se libro è ill o ril)
Area_interesse_1	Tutti (es. architettura, economia, filosofia...)
Area_interesse2	270 elem. Area diversa dalla prima

In_preparazione	20 (riporta solo si o no)
copertina	vuoto
Area_interesse3	vuoto
Promo	15 elementi
Promo2	15 elementi
url	1

Catalogo editori Alice, (1557 records)

ISBN_13	Tutti
Autore	1036
Titolo	tutti
Cur.	620
Trad	77
Anno	Tutti (da 1994 a 2009)
Ed. num	117
Pag.	1534
Ill.	67 hanno valore: ril. br. ill
Num. Vol	3
Collana	597
Prz	Tutti hanno prezzo
Disp	Tutti (valori "C" o "E")
Scol.	Vuoto
Marchio editoriale	Tutti, valori: Guerini e Associati, Guerini Scientifica, Guerini studio
Cod Alice	1507

Tutti i libri del catalogo on line sono presenti in Anagrafica

Match Catalogo Alice con Anagrafica e catalogo on line risultati:

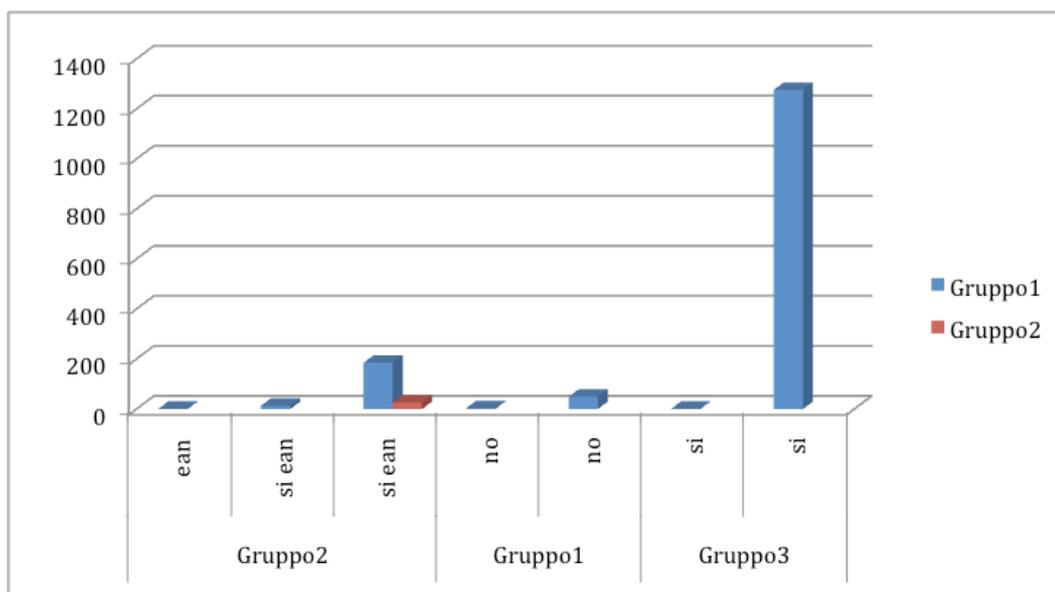
Conteggio di ISBN		Match anagrafica2	Match anagrafica	Totale complessivo
Match cat on line2	Match cat on line	Gruppo1	Gruppo2	
Gruppo2	si ean	1199	26	1 13 211
Gruppo1	no	3		3
	no	49		49
Gruppo3	si	1		1
	si	1276		1276

Totale complessivo	1528	26	1554
--------------------	------	----	------

Lettura grafico: Tutti i gli elementi di Catalogo Alice sono presenti in anagrafica, 26 con il codice EAN gli altri riportano l'ISBN.

1502 elementi di Catalogo Alice sono presenti anche in catalogo on line.

52 elementi non sono presenti in entrambi i cataloghi.



Codifica XML/TEI

Prima di dare una spiegazione dettagliata degli elementi utilizzati, entrando così nel vivo della codifica, ritengo necessario fare una panoramica sull'XML e dare alcuni cenni sulla TEI. Premetto che non scenderemo in dettagli troppo approfonditi, poiché questa si prefigura come una semplice presentazione e non vogliamo scrivere una dispensa esaustiva sul linguaggio in questione.

6.1 XML

LXML, acronimo di eXtensible Markup Language, nasce come metalinguaggio nel 1998, prima ed unica versione, uscito dopo una prima bozza del 1996.

Ideato e realizzato dal W₃C (World Wide Web Consortium): un consorzio fondato dal padre del web Tim Berners Lee¹²⁴ con lo scopo di migliorare gli esistenti protocolli e linguaggi per il WWW e di aiutare il Web a sviluppare tutte le sue potenzialità.

I nostri scopi che prefiguravano la creazione di una codifica digitale, si sono ben conciliati con questa tipologia di marcatura. LXML è in grado di risolvere i principali problemi legati al supporto digitale; capace nel superare barriere di portabilità. In sostanza l'XML: non necessita di dispositivi hardware e software specifici, cerca di fare fronte all'elevata obsolescenza, funziona su tutte le piattaforme, è uno standard che ingloba la proliferazione dei sistemi di codifica, mette a disposizione una facile condivisione dei dati e dei risultati.

¹²⁴ Tim Berners-Lee (Londra, 8 giugno 1955) il co-inventore del World Wide Web (WWW), insieme a Robert Cailliau e ad un migliaio di altri amici. Nel 1980 trascorse sei mesi al CERN come consulente nel campo dell'ingegneria del software. Lì realizzò, per uso privato, il primo programma per immagazzinare informazioni usando associazioni casuali. Tale programma (chiamato Enquire e mai pubblicato) formerà la base concettuale per il futuro sviluppo del World Wide Web.

Inoltre i punti forti dell'XML, essendo un linguaggio dichiarativo, come lo era l'SGML¹²⁵, il padre dell'XML, sono:

- ✦ è orientato al testo;
- ✦ è un insieme di istruzioni per annotare il significato degli elementi costitutivi;
- ✦ trascura l'aspetto che assumerà nell'output;
- ✦ permette comunque di progettare una sua grafica, servendosi di un ulteriore livello procedurale (fogli di stile: CSS e XSLT);
- ✦ il markup è definito dall'utente;
- ✦ cattura della semantica del testo;
- ✦ flessibilità e possibilità di essere usato per molteplici scopi;
- ✦ libertà nella costruzione della propria grammatica (DTD).

La sigla **DTD**, sta per Document Type Definition, è una grammatica costituita da un insieme di dichiarazioni che regolano la struttura del documento XML, la DTD permette di controllare che il file XML sia valido; la validazione avviene se il file XML è conforme alla DTD. Se invece è rispettata la sintassi si dice che il documento è ben formato.

Scelto il linguaggio va pensato il **modello di codifica**. Un modello è il risultato di un procedimento di astrazione necessario per sfuggire all'imprevedibilità del continuo. Un modello deve essere qualcosa di "più piccolo" dell'oggetto che si vuole analizzare, deve essere isomorfo, e si dovrà quindi fare una scelta delle caratteristiche da codificare. Nella progettazione del nostro modello di codifica ci siamo chiesti quali fossero le caratteristiche da mettere in evidenza, quali parti del testo considerare e quali altre cose tralasciare poichè non erano funzionali alla fruizione che avevamo in mente.

Con l'XML possiamo inventare tutti i tag che desideriamo... ciò se da una parte lascia ampio spazio alla fantasia e alla libertà di ognuno, genera però una proliferazione degli schemi di codifica. Inoltre non si era ancora previsto un modello condiviso da tutti adatto ai testi umanistici o linguistici, che permettesse loro la diffusione a livello digitale. Se il modello è uguale per tutti al-

¹²⁵ Acronimo di Standard Generalized Markup Language è anch'esso un meta-linguaggio per definire linguaggi di markup, ma è più generale e complesso di XML. XML è in realtà derivato da SGML.

lora si possono fare dei confronti costruttivi. Sono questi i motivi che nel 1986 hanno portato alla nascita della TEI.

6.2 La TEI

Nel novembre 1987 i rappresentanti di oltre quaranta istituzioni accademiche si riunirono al Vassar College, nello stato di New York. Tutti avevano lo stesso intento: la trasformazione del patrimonio letterario mondiale in formato elettronico. Il motivo principale che li spinse era la mancanza di standard internazionali condivisi per la creazione di quelli che al tempo venivano denominati “machine-readable text”.

Nei primi anni ‘90 la TEI iniziò ad avere finanziamenti da vari istituzioni¹²⁶ che ne incentivavano lo sviluppo. Lo scopo del progetto era creare una descrizione completa delle pratiche correnti e la realizzazione di raccomandazioni o *Guidelines* per la codifica di testi elettronici per le discipline umanistiche e linguistiche.

Nel maggio 1994 si ha la prima pubblicazione di una versione completa e stabile delle *Guidelines for Text Encoding and Interchange*, chiamata TEI P₃, il linguaggio utilizzato era l’SGML. La TEI P₃ divenne lo standard di riferimento per la costruzione di biblioteche digitali.

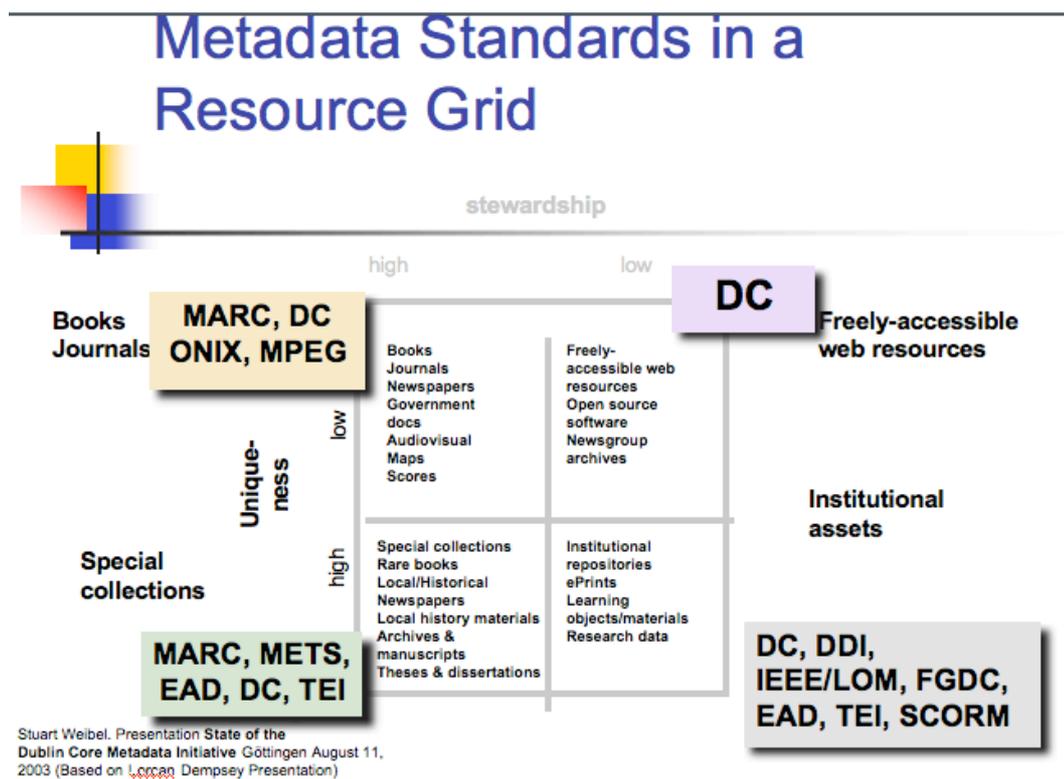
Nel 2000 la TEI si è costituita in consorzio internazionale a partecipazione libera, esse unisce un gruppo internazionale di studiosi, comprendendo tre associazioni: l’Association for Computing and the Humanities (ACH), l’Association for Computational Linguistic (ACL) ed infine l’Association for Literary and Linguistic Computing (ALLC). Possiede una sede congiunta in quattro università, sparse per il mondo: al CNRS di Nancy in Francia, all’Università di Oxford, alla Brown University, ed all’Università della Virginia. Lo standard si è evoluto e nel 2002 è uscita la nuova versione: TEI P₄ con l’adozione dell’XML standard promosso del W₃C, pubblicata dall’University of Virginia Press per conto del TEI Consortium.

¹²⁶ In particolare coloro che parteciparono al finanziamento furono: Nationale Endowment for the Humanities statunitense, l’Unione Europea, Canadian Social Science Research Council e la Andrew W. Mellon Foundation. La TEI venne inoltre sponsorizzata in modo congiunto da tre affermate associazioni professionali internazionali, le quali stabilirono un comitato direttivo e designarono due “curatori” al fine di dirigere circa cento studiosi provenienti da tutto il mondo.

Nel 2005, dopo circa tre anni di lavoro, è uscita un'ulteriore versione, la TEI P5, ed è tuttora lo schema di riferimento.

Infine esiste una versione denominata TEI Lite, la quale fornisce un sottoinsieme delle raccomandazioni della TEI, cercando di avere un numero abbastanza ricco di elementi tali da supportare diverse elaborazioni testuali per lo produzione di documentazione on line. E' stata concepita per dimostrare che uno schema TEI può essere personalizzato in base alle proprie esigenze.

La TEI possiede una DTD molto vasta che comprende centinaia di elementi e molti pacchetti che caratterizzano le varie tipologie testuali: teatro, trascrizione del parlato, poesia, prosa ecc. La versione completa della DTD TEI P5 può essere scaricata alla pagina del W3C sotto la forma di un file compresso in formato zip. Essa ha una complessa struttura modulare.



5

Figura 18: dove si colloca la TEI rispetto agli altri standard.

6.3 Scelta dello standard TEI

Il catalogo deve assolvere a due precipue funzioni: il reperimento e la localizzazione delle informazioni. Il catalogo codificato pone l'utente in grado di reperire, attraverso una ricerca in cui si utilizzano attributi o relazioni, risorse bibliografiche presenti nel catalogo, di localizzare una singola risorsa e/o insieme di risorse che corrispondono alle specifiche richieste.

Data la grande versatilità e la vasta offerta di tag che offre si presta bene anche in previsione di un nuovo progetto: l'inserimento del full text on line con possibilità di vendita, integrale o parziale dei capitoli.

Se è codificato in TEI anche il testo (non solo con elementi strutturali, quali capitolo, fine pagine ecc.), l'utente può reperire il materiale di cui necessita. Si possono fare query più raffinate rispetto "a trova quella parola nel testo", perché se ho aggiunto un elemento specificando ad esempio tutti i <persName> con un attributo che ne specifica la professione o il ruolo i risultati saranno molto dettagliati. A tale proposito vorrei citare, per capire quale può essere un uso appropriato, il mio lavoro di tesi della laurea triennale.

Il titolo della tesi era *Edizione digitale degli Annales Pisani di bernardo Maragone nell'edizione critica di Michele Lupo Gentile*. In questo lavoro, scritto nel XI in tardo latino, è stato scelto un modello di codifica. Gli elementi scelti, in base al tipo di ricerca che poi si voleva effettuare, sono stati:

- ▶ <date>, differenziazione del calendario comune e del calendario pisano (un anno in avanti), utilizzo dell'attributo *value*, nel formato AAAAMMGG es. <date value="11450000">MCXLV</date>;
- ▶ <abbr>, normalizzazione delle abbreviazioni, es. <abbr expan="kalendae">kal.</abbr>
- ▶ <persName> normalizzazione del nome in nominativo del latino attraverso l'attributo *reg* e *type* che specificava il ruolo, es. <persName reg="Federicus rex Romanorum, Federico I" type="imperatore">imperatoris Frederici</persName>
- ▶ <placeName><name> con gli attributi *reg* e *type* che ne specificavano rispettivamente la regolarizzazione (nome attuale), es. <placeName reg="Ecclesia Sancta Maria, Cattedrale di Santa Maria, Pisa"><name

```
type="edificio religioso">canonicum Sancte Marie Maioris ecclesie</persName></name>
```

- ▶ `<num>` irregolarizza i numeri, es. `<num value="7">VII</num>`
- ▶ `<add>`
- ▶ `<supplied>` utilizzato per codificare correzioni o aggiunte da parte dell'editore, poichè la carta può essere deteriorata, es `quar<supplied>o</supplied>`
- ▶ `<hi>` funzione di enfaticizzazione, serve per descrivere un aspetto che si diversifica graficamente nel testo, come il corsivo, es `<hi rend="apice">III</apice>`
- ▶ `<note>` le note erano di due tipologie: note al testo e note storiche, sono state codificate e poi allegate al testo tramite xslt che le visualizzava come apice linkabile.
- ▶ `<text>`, `<body>`, `<head>`, `<div>`, `<p>`, `<pb/>`

Tali codifiche sono utili in ambito accademico per un ricercatore, un professore o uno studioso più in generale. Può essere utile in ambito filologico per ricostruire le varie annotazioni e le versioni di un manoscritto.

6.3.1 Elementi utilizzati per la codifica del catalogo storico editoriale

Va premesso che le informazioni che abbiamo scelto di mantenere erano molte e specifiche del settore. Oltre ai dati descrittivi della risorsa libro vi sono anche dati amministrativi (settore, area, disponibilità) che non abbiamo cancellato poiché utili all'azienda (una volta inseriti si possono sempre ignorare se non vogliamo utilizzarli). Sono stati utilizzati quindi alcuni espedienti che permettessero di inserire tutte le informazioni, usando gli elementi standard delle *Guidelines*. Elenco completo degli elementi, che rispettano le specifiche di TEI-c:

✓ **<bibl>** contiene la risorsa, ingloba anche gli altri elementi. Ha un attributo `n` che indica il numero del titolo inserito. E' l'elemento utilizzato per le citazioni bibliografiche. Gli **elementi core** di `<bibl>` secondo le *Guidelines*, sono:

abbr, add, address, author, biblScope, cb, choice, corr, date, del, distinct, editor, email, emph, expan, foreign, gap, gloss, hi, index, lb, measure, measureGrp, meeting, mentioned, milestone, name, note, num, orig, pb, ptr, pubPlace, publisher, ref, reg, relatedItem, respStmt, rs, series, sic, soCalled, term, time, title, unclear.

La sua dichiarazione:

```
element bibl
{
  att.global.attributes,
  att.declarable.attributes,
  att.typed.attributes,
  (
    text
  | model.gLike
  | model.highlighted
  | model.pPart.data
  | model.pPart.edit
  | model.segLike
  | model.ptrLike
  | model.biblPart
  | model.global
  )*
}
```

✓ **<idno>** utilizzato con l'attributo type per identificare tutte quelle informazioni che non avevano un elemento corrispondente, come ad esempio il codice ISBN. La scelta è stata dettata dalle *Guidelines* che mostrano come esempio, per rappresentare il codice ISSN: `<idno type="ISSN">0143-3385</idno>`.

La sua dichiarazione:

```
element idno
{
  att.global.attributes,
  attribute type { data.enumerated }?,
  text
}
```

✓ **<author>** in un riferimento bibliografico contiene il nome dell'autore (o degli autori), personale o collettivo, di un'opera; è la dichiarazione di responsabilità primaria di ciascuna unità bibliografica.

La sua dichiarazione:

```
element author
{
```

```

    att.global.attributes,
    att.canonical.attributes,
    macro.phraseSeq
}

```

✓ **<editor>** è la dichiarazione di responsabilità secondaria di ciascuna unità bibliografica, ad esempio il nome di un individuo, di un'istituzione o organizzazione che agisce come curatore, compilatore, traduttore, ecc. Utilizzo dell'attributo `role`, `@role` specifica la natura della responsabilità intellettuale.

La sua dichiarazione

```

element editor
{
    att.global.attributes,
    attribute role { data.enumerated }?,
    macro.phraseSeq
}

```

Utilizzati i valori `editor` e `translator`.

✓ **<title>** contiene il titolo completo di una qualsiasi opera. L'attributo `@level` indica il livello bibliografico di un titolo, ovvero se identifica un articolo, un libro, una rivista, una collana, o materiale non pubblicato.

La sua dichiarazione

```

element title
{
    att.global.attributes,
    att.canonical.attributes,
    attribute level { "a" | "m" | "j" | "s" | "u" }?,
    attribute type { data.enumerated }?,
    macro.paraContent
}

```

✓ **<date>** contiene una data in qualsiasi formato, può contenere `@calendar` che ne specifica la tipologia di calendario (gregoriano, islamico, cinese ecc.)

La sua dichiarazione

```

element date
{
    att.global.attributes,
    att.dateable.w3c.attributes,
    att.dateable.iso.attributes,
    att.duration.w3c.attributes,
    att.duration.iso.attributes,
    att.editLike.attributes,
}

```

```

att.dimensions.attributes,
att.ranging.attributes,
att.typed.attributes,
attribute calendar
{
  "Gregorian"
| "Julian"
| "Islamic"
| "Hebrew"
| "Revolutionary"
| "Iranian"
| "Coptic"
| "Chinese"
| xsd:Name
}?,
(text | model.gLike | model.phrase | model.global)*
}

```

✓ **<series>** (informazioni sulla collana) contiene indicazioni sulla collana in cui un libro o un'altra unità bibliografica sono stati pubblicati.

La sua dichiarazione

```

element series
{
  att.global.attributes,
  (
    text
  | model.gLike
  | title
  | editor
  | respStmt
  | biblScope
  | model.global
  )*
}

```

✓ **<publisher>** fornisce in nome dell'editore

La sua dichiarazione

```

element publisher { att.global.attributes, macro.phraseSeq}

```

✓ **<biblScope>** (estensione del riferimento bibliografico) definisce l'estensione di un riferimento bibliografico, per esempio mediante una lista di numeri di pagina, o il titolo di una parte di un'opera più ampia. Possiede l'attributo *type*, per il quale le *Guidelines* suggeriscono alcuni valori:

vol (volume) l'elemento contiene il numero del volume

issue l'elemento contiene l'indicazione del numero della pubblicazione

pp (pagine) l'elemento contiene l'indicazione di pagina o pagine

ll (lines) l'elemento contiene un numero di righe o un intervallo di righe

chap (capitolo) l'elemento contiene un'indicazione di capitolo (numero e/o titolo)

part l'elemento identifica una parte di un libro o di una raccolta

La sua dichiarazione

element biblScope

```
{
  att.global.attributes,
  attribute type
  {
    "vol" | "issue" | "pp" | "ll" | "chap" | "part" | xsd:Name
  }?,
  attribute from { data.word }?,
  attribute to { data.word }?,
  macro.phraseSeq
}
```

✓ <note> è l'elemento jolly, ovvero con un corretto attributo può contenere tutto ciò che non ha un elemento corrispettivo vero e proprio, infatti contiene una nota o un'annotazione. @type descrive il tipo di nota.

La sua dichiarazione

element note

```
{
  att.global.attributes,
  att.placement.attributes,
  attribute type { data.enumerated }?,
  attribute resp { data.pointer }?,
  attribute anchored { data.truthValue }?,
  attribute target { list { data.pointer+ } }?,
  attribute targetEnd { list { data.pointer+ } }?,
  macro.specialPara
}
```

6.3.2 Il `teiHeader` e il tag TEI

✓ **<teiHeader>** (intestazione) fornisce informazioni descrittive e dichiarative che costituiscono un frontespizio elettronico che precede qualsiasi testo TEI-conforme.

La sua dichiarazione

```
element teiHeader
{
  att.global.attributes,
  attribute type { data.enumerated }?,
  ( fileDesc, model.headerPart*, revisionDesc? )
}
```

Di seguito il `<teiHeader>` del catalogo:

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/css" href="teixlite.css"?>
<?oxygen RNGSchema="schema.xml"?>
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Conversione cataloghi, formato TEI standard</title>
        <author>Alice Guidi</author>
      </titleStmt>
      <publicationStmt>
        <p>Catalogo storico editoriale</p>
      </publicationStmt>
      <sourceDesc>
        <p>Catalogo unico della casa editrice Guerini e Associati</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
```

✓ **<TEI>** (documento TEI) contiene un documento TEI-conforme, comprendente un'intestazione e un testo, sia esso isolato o parte di un elemento.

La sua dichiarazione

```
element TEI
{
  att.global.attributes,
  attribute version { xsd:decimal }?,
```

```
(teiHeader, ((model.resourceLike+, text?) | text))
}
```

L'elemento TEI nel catalogo:

```
<TEI xml:id="Indbibliografiche">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Catalogo storico editoriale, Guerini e Associati</title>
        <author>Alice Guidi</author>
      </titleStmt>
      <publicationStmt>
        <p>Catalogo storico editoriale</p>
      </publicationStmt>
      <sourceDesc>
        <p>Catalogo unico</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <div1>
```

teiHeader e MARC

Nel corso del tempo, l'evolversi del teiHeader è stato verso una maggiore coerenza e compatibilità con le librerie tradizionali di catalogazione e di una maggiore congruenza sintattica con MARC. Il teiHeader nasce con l'idea di una codifica accessibile senza aver bisogno di competenze catalografiche specifiche, richieste ad es. da MARC. Non è escluso che l'intestazione TEI si evolverà in modo tale da creare una codifica ancora più dettagliata, descrivendo specificatamente il profilo e le informazioni di revisione, puntando a un record MARC, che dovrebbe contenere la descrizione bibliografica.

L'intestazione TEI supporta una serie di categorie di campo che non possono essere catturati in MARC, ad esempio, la sezione cronologia delle modifiche fornisce una struttura per la registrazione dei cambiamenti apportati a un testo elettronico, compresa la data, i responsabili e la natura dell'aggiornamento. La sourceDesc all'interno del file desc consente una dettagliata descrizione ricca di contenuti, designati in particolare per le fonti non di stampa. L'enco-

dingDesc prevede una lunga e dettagliata descrizione della codifica dei file elettronici contenenti i dati del progetto, lo scopo per il quale è stato creato, le decisioni editoriali che sono state fatte, e le pratiche di trascrizione che sono state utilizzate.

A differenza di METS, nelle *Guidelines* TEI non si specifica un particolare approccio al problema della fedeltà al testo di partenza e di recupero degli originali, ma si prevede una trattazione per le caratteristiche tipografiche e linguistica del testo, piuttosto che una codifica del layout o distinzioni di qualità sui manoscritti.

TEI non limita la combinazione di informazioni oggettive e soggettive nella codifica. Le linee guida forniscono un mezzo per la codifica per la rappresentazione del testo così come per l'interpretazione del testo e di analisi.

Esempio di codifica di un titolo presente nel catalogo:

```
<bibl n="8"><idno type="Isbn">8878020087</idno>
<idno type="CodAltern">887802008</idno>
<idno type="Ean">9788878020085</idno>
<idno type="codAlice">£</idno>
<author>D'AGOSTINI MARIA</author>
<editor role="editor">£</editor>
<editor role="translator">£</editor>
<title>LA LETTERATURA DI VIAGGIO</title>
<date>1987</date>
<series>£</series>
<publisher>£</publisher>
<biblScope type="pp">0</biblScope>
<biblScope type="num_volumi">1</biblScope>
<note type="Prezzo">20.66</note>
<note type="Categoria">1</note>
<note type="Centro">1</note>
<note type="Area">1</note>
<note type="Settore">114</note>
<note type="Stato">F</note><
note type="Note">£</note>
<note type="Collana_codice">£</note>
<note type="Ill_rilegatura">£</note>
<note type="Area_interesser">£</note>
<note type="Area_interesse2">£</note>
<note type="EdNum">£</note></bibl>
```

6.4 Scelta del motore di ricerca per XML: Tauro-core

Una volta codificato il catalogo era necessario un motore di ricerca che mostrasse i risultati delle interrogazioni. Esistono alcuni motori di ricerca che per XML, tra tutti ho scelto Tauro-core.

Breve premessa sui motori di ricerca. I motori di ricerca tradizionali, come Google per il web, considerano principalmente la parte testuale dei documenti permettendo la ricerca di parole al suo interno. I motori di ricerca XML offrono invece la possibilità di ricercare sia nel contenuto testuale del documento che nella *struttura* codificata, appunto, tramite il formato XML. Ciò consente inevitabilmente all'utente di progettare interrogazioni più sofisticate e selettive.

Attualmente il panorama scientifico e industriale offre diversi motori di ricerca per l'indicizzazione dei documenti XML: la maggior parte di essi sono orientati all'utilizzo in modalità **“data centric”**, ovvero consentono di gestire documenti strutturati o campi con forte regolarità quali i dati di tipo tabellare, ordini di vendita, dati scientifici e simili. Poche sono le invece le soluzioni per la gestione di XML in modalità **“document centric”**, (sono un esempio eXist e GalaTeX), ovvero adatti a trattare documenti con struttura irregolare, pochi campi e molto testo, quali appunto le fonti letterarie. TauRo si concentra su quest'ultima tipologia, individuando i requisiti fondamentali che un motore di ricerca moderno dovrebbe offrire per essere considerato utile per gli studiosi umanisti. TauRo è un motore di ricerca per documenti XML, sviluppato dal *Signum*, il centro di ricerca informatico per documenti umanistici della Scuola Normale Superiore. Il nome nasce dall'acronimo di Text Retrieval componendo il nome delle corrispondenti lettere dell'alfabeto greco: *Tau* e *Ro*. Nella fase di progettazione di “motore di ricerca ideale” gli implementatori di TauRo hanno tenuto conto di un aspetto non meno rilevante da considerare riguardante la modalità di restituzione dei risultati, in quanto questa è strettamente legata ad eventuali operazioni di post-processing che potrebbero essere effettuate su di essi al fine di estrarre ulteriori informazioni, possibilmente grazie ad altri strumenti software tipici del *text mining*.

Per offrire questa flessibilità è necessario prevedere un meccanismo di restituzione dei risultati che consenta una efficiente contestualizzazione, attraverso una serie di informazioni aggiuntive che includono, oltre al risultato vero e proprio (documento, numero di linea, posizione assoluta in parole e caratteri, . . .), anche il testo che lo circonda (chiamati *snippet*.) possibilmente privo della marcatura XML per consentirne una sua più efficace visualizzazione.

Nelle figure 19 e 20 si può notare la sua potenzialità, paragonato agli altri motori di ricerca per XML

Compression Level and Indexing Time

Collections	TauRo		BaseX		eXist		Qizx/open		MonetDb	
	index dim. (Kb)	Time creation (s)								
28	28	0,02	172	0,06	80	0,41	74	0,39	94	0,11
12000	8300	4,53	29000	8,65	30000	26,7	10700	10,2	12000	1,9
23000	18000	9,22	57000	16,16	70000	52,8	21500	18,33	24000	5,25
40000	33000	15,86	96000	27,93	109000	93,29	37000	30,57	41000	8,25
56000	46000	22,43	137000	41,24	151000	154,6	52000	42,29	57000	10,37
112000	80000	47,46	274000	85,39	299000	351,46	102000	86,43	105000	21,22
112000	85000	47,27	274000	89,97	297000	539,88	102000	82,72	105000	26
168000	111000	71,79	403000	131,24	470000	683	152000	127,75	151000	32,4
224000	156000	97,33	547000	187,88	630000	858	201000	172,99	194000	52,72
258000	186000	110,18	626000	238,54	700000	1411	231000	199,07	219000	47,93

Caratteristiche a confronto

	XXL	Elixir	Galatex	XIRQL	eXist	JuruXML	Amberfish
Esatta	X	come XML-QL	X	X	X	X	X
Stemming	X	come XML-QL	X	-	-	-	X
Full text-> Proximity	-	come XML-QL	X	-	X	-	-
Errore	-	come XML-QL	-	-	fuzzy matching	-	-
RegExp	-	come XML-QL	X	-	X	-	solo prefisso
Boolean	-	come XML-QL	X	-	X	-	con ranking
Ranking	X	X	X	X	in futuro???	X	X
Ricerca in struttura	X	X	X	X	X	X	X
Piattaforma	Java	-	Windows/Mac OS X/Linux Solaris	-	Java	-	Unix
Licenza	-	-	AT&T's NON-COMMERCIAL SOURCE CODE AGREEMENT	-	LGPL	-	GPL
Requisiti	Oracle 8i	XML-QL, WHIRL	PCRE 5.0, O'CAML	-	-	-	-
Indicizzazione	-	-	solo parole	solo un query language	parole, tag, full-text	-	Parole (path nell'indice)
Compressione	EPI,ECl,OI	solo un query language	-	solo un query language	EPI, ECI	ECI	-
Interfaccia	-	solo un query language	Command-line, APIs (C, Java, O'caml), web	solo un query language	webAdmin Database Interface, Java Admin Client	-	-
Query Language	XXL	XML-QL + Similarity	Xquery+TexQuery	XIRQL	estensione di Xquery	XML fragment && text	proprietario
Note	Ontol. Similarity (alla WordNet); supporto di join	syntatic similarity (WHIRL)	XIRQL est. XQL	XML Native DB	XML Native DB	-	-

EPI=Element Path Index
ECl=Element Content Index
OI=Ontology Index

Il motore di ricerca era già presente on line dal 2007, sotto il nome di TauRo e permetteva di caricare la propria collezione all'indirizzo <<http://tauro.signum.sns.it/>> dove una volta memorizzato nel database era possibile fare query per ricercare nei tag inseriti nel documento. Avevo già te-

stato il suo funzionamento caricando la codifica degli *Annales Pisani*, svolta per la tesi della triennale, scegliendo la modalità privata sotto login. Avevo quindi già verificato la sua efficacia, ed era interessante che si potesse consultare on line e non come altri motori installati in remoto sul proprio computer.

Da giugno 2008 è nato TauRo-core specificatamente realizzato per gestire testi XML complessi tipici del dominio umanistico: esegue ricerche sulla struttura dei documenti XML e ricerche full-text incluse quelle per espressioni regolari, per approssimazione e per vicinanza tra termini. Il grande vantaggio della nuova versione Tauro-core è quello di poter installare il motore di ricerca sul proprio computer e poi sfruttarne il funzionamento su un sito web, accessibile a tutti quindi e non solo a pochi utenti.

Si presenta con un'architettura modulare e flessibile e assicura elevate prestazioni nelle interrogazioni e nell'indicizzazione.

Caratteristiche tecniche:

- ↳ gestisce gli smart-tag (che si suddividono in *Jump-tag*, *Soft-tag* e *Split-tag*)
- ↳ opera su qualsiasi documento XML senza bisogno di specificare DTD o schema
- ↳ supporta Unicode
- ↳ *TRQL* - TauRo Query Language
- ↳ ha un'indicizzazione compressa
- ↳ disponibile open-source con *licenza Creative Commons*
- ↳ realizzato in C
- ↳ disponibile per Linux (e in futuro Windows)

La sua Architettura

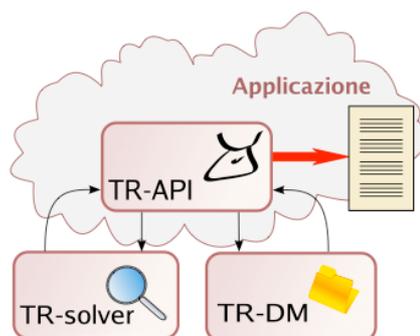


Figura 22: schema dell'architettura di TauRo

Il primo modulo è **TR-API** – TauRo Application Programming Interface – che riceve l’interrogazione in TRQL dall’applicazione con cui l’utente interagisce. L’interrogazione ricevuta viene validata e sottoposta al modulo **TR-Solver** che si occupa della sua risoluzione e restituisce a TR-API una serie di risultati utili alla successiva estrazione delle occorrenze dal testo (*snippet*). Tali risultati sono poi inviati al modulo **TR-DM** – TauRo Document Manager – che dispone della collezione di documenti (possibilmente memorizzata in formato compresso), estrae gli snippet secondo il formato indicato nella clausola Return dell’interrogazione, e li restituisce all’applicazione che dovrà presentarli all’utente.

In dettaglio, il modulo TR-Solver riceve l’interrogazione validata e la trasforma in una propria rappresentazione attraverso la quale è possibile scomporre l’interrogazione completa in interrogazioni elementari (risolte attingendo alle informazioni presenti nell’indice relativo alla collezione di documenti). Successivamente i risultati delle interrogazioni elementari vengono ricombinati e filtrati per ottenere quelli finali e definitivi.

Nel caso in cui i documenti non siano stati ancora indicizzati (o nel caso in cui le risorse disponibili siano tali da non risultare conveniente la creazione di un indice), il sistema è in grado di accedere direttamente ai documenti ed estrarne le informazioni necessarie (per scansione): a parità di tempo, però, la mole di dati elaborabile è inferiore rispetto al caso in cui la collezione di documenti sia stata indicizzata a priori.

I tre moduli possono essere fusi in un’unica applicazione, oppure possono far parte di applicazioni distinte che comunicano attraverso la rete: in quest’ultimo caso le applicazioni e i dati possono risiedere su nodi diversi, anche distanti tra loro. In questo scenario i moduli comunicano attraverso un protocollo proprietario denominato TRIL, descritto qui di seguito.

Protocollo di comunicazione e configurabilità

TRIL – TauRo Interface Language – è il protocollo di comunicazione tra i moduli di TauRo. Il protocollo è basato sul linguaggio XML e ha lo scopo di interfacciare i vari moduli nel caso in cui questi non facciano parte di una singola applicazione. Le sue funzionalità principali possono essere raggruppate in quattro categorie:

★ *Invio interrogazione*: viene creato un messaggio incapsulando l'interrogazione in TRQL e aggiungendo la richiesta di invio dei risultati. Tale richiesta può prevedere la restituzione di tutti i risultati, oppure la suddivisione dei risultati in blocchi. Questa comunicazione avviene da TR-API verso TR-Solver.

★ *Restituzione dei risultati*: il formato di un risultato include diverse informazioni tra cui l'identificativo del documento in cui appare, e la posizione dell'occorrenza al suo interno. Questa comunicazione avviene da TR-Solver a TR-API.

★ *Richiesta di formazione snippet*: il protocollo mette a disposizione vari tipi di richieste per implementare le funzionalità descritte relativamente alla clausola Return. Questa comunicazione avviene da TR-API a TR-DM, utilizzando le informazioni ricevute da TR-Solver.

★ *Invio snippet* : gli snippet sono scritti in XML in modo da essere incapsulati in TRIL. Questa comunicazione avviene da TR-DM a TR-API.

6.5 TRQL, TauRoCore Query Language

Il motore TauroCore permette di effettuare le query per:

- * parola esatta, trova *filosofia*
- * prefisso, cerca filo*
- * suffisso, cerca *sofia
- * contenuto, cerca *filosof*
- * max error, imposto errore es=2, ceca *filosofico* trova anche "filosofia"
- * espressione regolare: cerca per espressioni regolari come AND, OR, NOT.

Vediamo il codice del linguaggio:

```
TRQLELEMENT ::= TAG | TRQLTAG | TRQLWORD;
```

```
TAG ::= startTag TRQLELEMENT* endTag | pointTag;
```

```
TRQLTAG ::= (TRQLTAG)+ | TRQLWORD |  
    <xml_next [stag] [binding]? > pointTag </xml_next> |  
    <xml_prev [stag] [binding]? > pointTag </xml_prev> |  
    <xml_proximity maxprox [binding]? > (TRQLWORD)(TRQLWORD)+  
</xml_proximity> |  
    <xml_anyvalue [binding]? > TRQLELEMENT </xml_anyvalue>;
```

```
TRQLWORD ::= <xml_exact [binding]? [case]? > Literal </xml_exact> |
```

```

<xml_prefix    [binding]? [case]? > Literal </xml_prefix> |
<xml_suffix    [binding]? [case]? > Literal </xml_suffix> |
<xml_contained [binding]? [case]? > Literal </xml_contained> |
<xml_error maxerr [binding]? > Literal </xml_error> |
<xml_regexp    [binding]? > RegularExpression </xml_regexp>;

```

```

TRQLRETURN ::=      bindingName | startTag bindingName(TRQLRETURN)* endTag |
Literal bindingName(TRQLRETURN)* Literal;

```

```

ATTS ::= Literal='Literal' | TRQLATT;

```

```

TRQLATT ::= binding | dist | maxerr | maxprox | stag;

```

```

binding ::= 'xml_var=' '''bindingName''';
case    ::= 'xml_case=' '''caseValue''';
dist    ::= 'xml_dist=' (-1|1|2|3|4|5|6|7|8|9)(0|9)*;
maxerr  ::= 'xml_maxerr=' (0|9)(0|9)*;
maxprox ::= 'xml_maxprox=' (1|9)(0|9)*;
stag    ::= 'xml_stag=' '''Literal''';

```

```

startTag ::= <tagName ATTS* >;
endTag    ::= </tagName>;
pointTag  ::= <tagName ATTS*/>;

```

```

caseValue ::= sensitive | insensitive;
bindingName ::= $Literal;
tagName    ::= Literal;
Literal    ::= string;

```

```

RegularExpression ::=      RegularExpression '*' |
                           RegularExpression '+' |
                           RegularExpression '.' RegularExpression |
                           RegularExpression '|' RegularExpression

```

6.6 Sito web realizzato

Il catalogo una volta codificato doveva pur essere reso in una forma tale da poter essere consultato. Invece di creare un sito con php che si interfacciasse con un database, ho creato un sito html e php potenziato dal motore di ricerca Tauro-Core, che indicizza la collezione e crea le query. Sono necessari alcuni fogli di stile, gli XSLT che permettono una visualizzazione grafica con stile dei tag XML. Guerini e Associati possiede già un sito web dinamico che sfrutta php. La grafica non è delle più accattivanti e per mostrare la potenzialità della codifica sia dei libri presenti in catalogo e del testo (quando lo renderanno disponibile) abbinata ad un motore di ricerca ho pensato di crearne uno ex-novo. Riporto di seguito gli screenshot del sito realizzato:

GUERINI E ASSOCIATI

HOME

CATALOGO

RICERCA



Eventi

29/10/2009

Presentazione del libro *Dalla maschera al volto. Piccole storie, immagini e suoni per ritrovare se stessi*, Alessandro Rinadi

30/10/2009

Seminario sul libro *Russia: democrazia europea o potenza globale? A vent'anni dalla fine della guerra fredda*, Maurizio Massari

Novità in libreria



Leggere la crisi per rimettere in moto l'economia

Deaglio Mario, Arfaras Giorgio, Caffarena Anna, Frankel Giorgio S., Russo Giuseppe

Alla scuola della crisi. XIV Rapporto sull'economia globale e l'Italia

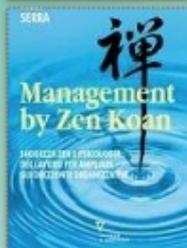
In collaborazione con il centro di Ricerca e Documentazione "Luigi Einaudi" e Ubi Banca



Educare e istruire nel terzo millennio

Cominelli Giovanni

La scuola è finita... forse. Per insegnanti sulle tracce di sé
Introduzione di Giorgio Vittadini



La saggezza Zen al servizio dell'impresa

Verza Stefano, Serra Tetsugen

Management by Zen Koan. Saggezza zen e psicologia del lavoro per ampliare gli orizzonti organizzativi

Vetrina



Francesco Varanini

Le parole del manager. 108 voci per capire l'impresa



Figura 22: home page del sito, in formato html riporta una sezione con gli eventi, una delle novità in libreria ed infine una che mostra i libri in vetrina.

Catalogo

HOME CATALOGO RICERCA



Collane

Università: Economia e Management, Sociologia e comunicazione, Psicologia e Pedagogia, Lingue e Letteratura, Filosofia, Arte e Architettura

Svago: Giardinaggio, cucina, musica

News

Tutte le novità

Iscriviti alla Newsletter

Libri presenti

« precedente 2 successivo »

ISBN: 8878020028 **EAN:** 9788878020023

Ferraris Maurizio, *ERMENEUTICA DI PROUST*, 1987
 Collana: Saggi
 Editore: Guerini e Associati , € 10.33

« precedente 2 successivo »

Figura 23: è la pagina, in formato php, che presenta tutti i record in catalogo. Delle informazioni abbiamo scelto di visualizzare solo quelle più interessanti per l'utente (tramite fogli di stile xsl che seleziona ciò che si vuole scrivere sulla pagina), il proprietario può sempre prelevare quelle di cui necessita.

Ricerca

HOME CATALOGO **RICERCA**



Ricerca

Ricerca libera: ricerca in tutto il catalogo

Ricerca nei tag TEI: puoi fare ricerche incrociate specificando più campi contemporaneamente, a cui puoi associare il codice ISBN oppure EAN

Sapere per fare



Guerini Vent'anni
1987 - 2007

Ricerca nel catalogo

Ricerca libera
In tutto il catalogo

Ricerca nei tag TEI

Autore Titolo

Area di interesse

Codice ISBN/EAN ISBN

Figura 24: è la pagina, in formato php, che permette all'utente una ricerca all'interno del catalogo. Si può fare una ricerca libera per parola, oppure una ricerca più avanzati nei tag XML. Per semplicità non mostriamo tutti i campi su cui poter ricercare, essi si possono sempre aggiungere e modificare in base ai bisogni.

Risultati

[HOME](#)

[CATALOGO](#)

[RICERCA](#)



Collane

Università: Economia e Management, Sociologia e comunicazione, Psicologia e Pedagogia, Lingue e Letteratura, Filosofia, Arte e Architettura
Svago: Giardinaggio, cucina, musica

News



Tutte le novità

Iscriviti alla Newsletter

Libri presenti

sono state trovate **4** occorrenze

[« pagina precedente | 1 | pagina successiva »](#)

» [voce completa nel catalogo](#)
EAN: 9788862500746

Leader e media. Campagna permanente e,

» [voce completa nel catalogo](#)
Coaching alle nuove sfide. **Leader**, coach e non solo manager,

» [voce completa nel catalogo](#)
Coaching **leader**. Dal potenziale individuale ai risultati,

» [voce completa nel catalogo](#)
di volo ad uso dei **leader** di domani, 2007

Figura 25: è la pagina, in formato php, dei risultati. La parola evidenziata in giallo è quella ricercata. Lo snippet mostra una porzione di testo che sta intorno all'occorrenza richiesta. Se si vogliono visualizzare tutte le informazioni relative a quel record basterà accedere alla voce completa nel catalogo, tramite link.

Conclusioni

Scrivere le conclusioni di un lavoro su tematiche *in fieri*, che per definizione non si consertano con l'elaborazione di un quadro ermeneutico definitivo, è impresa ardua, che a ben vedere, rischia di essere riduttiva.

Si tratta infatti di argomenti complessi, di un *mare magnum* di spunti di riflessione sulle tematiche suscettibili di essere sviluppate attraverso articolate trattazioni monografiche.

Vorrei ad ogni modo proporre un interrogativo a fine lavoro:

un archivio di testi digitalizzato, correlato da metadati, comporta una qualità maggiore rispetto ad un semplice catalogo, oppure è solo un modo per rappresentare i testi sul web e facilitarne l'accesso?

L'informatica umanistica cerca di rispondere a questi quesiti: unendo la parte letteraria alle tecnologie informatiche traccia che cerca di unire punti di vista diversi.

L'informatica umanistica non è una disciplina ancillare o di supporto non vuol dire farsi abbagliare dallo scintillio delle macchine o delle mode, giacché «l'informatica che ci interessa è più un episteme che una tecnologia» e tale nuovo orizzonte epistemico disegnato (e impostoci) dall'informatica è campo d'azione obbligato per chi studia come i segni si producono, trasmettono e comunicano. Nelle parole di Wendell Piez, obiettivo delle *Digital Humanities* allora è costruire una «media consciousness in a digital age»¹²⁷: la consapevolezza cioè che i formati della nostra memoria stanno cambiando e che è necessario una rifondazione delle discipline del documento e dei loro supporti per garantire una valida e “culturalmente sostenibile” trasmissione del sapere. L'informatica umanistica si colloca fra quelle «*not sufficiently laboured*» terre di mezzo e come

¹²⁷ Piez, Wendell, Something called Digital Humanities, in “Digital Humanities Quarterly”, <<http://www.digitalhumanities.org/dhq/vol/2/1/000020.html>>.

vedremo ambisce, con le sue riflessioni e i suoi strumenti, a occupare un posto di rilievo nelle nuove metodologie di edizione dei testi.

L'idea di creare un catalogo storico editoriale è sicuramente una fonte di informazioni, l'editore può accedere ai dati che interessano la gestione del magazzino, organizzare la ristampa di libri fuori catalogo. Dal lato utente si ha invece un potenziamento dato dai risultati della ricerca che permette di reperire risorse con argomenti affini. Lo studioso, il ricercatore, il professore ma anche un utente generico può così crearsi una propria ontologia su quel contenuto.

La Trans-codifica altro non è se non una “rappresentazione” e una “particolare traduzione” di un oggetto in/attraverso un altro linguaggio o codice.

Codificare un testo è un processo a strati che inizia dal singolo carattere alfabetico e che può arrivare a descrivere (rappresentare) le forme e strutture più “interne” e complesse di un testo, ivi comprese le figure retoriche, le strutture narrative, ecc. Dunque si tratta di un processo di ‘traduzione’ moltiplicato nel tempo e nello spazio e realizzato attraverso linguaggi diversi e di complessità crescente.

La trascrizione di un documento attraverso la scelta e l'utilizzo di marcatori, come avviene con il linguaggio XML, produce un proprio e indipendente “strato” semiotico che si affianca e sovrappone alla fonte.

Il markup sembra aver fornito, almeno fino ad oggi, lo strumento e il modello più vicino alla concezione del testo come «struttura» e «modello», nel senso di «organizzazione formale morfo-sintattica che ingloba un'articolazione dentro cui, senza scarti possibili, sta un senso, una possibilità di applicazione al concreto»¹²⁸.

Si ha un cambiamento nelle svolgimento di operazioni prima distinte, ma che oggi vengono concentrate in un mezzo (e spesso in una persona sola), esse erano suddivise tra varie figure e mezzi: l'autore, l'editore, l'editor, il composi-

¹²⁸ De Mauro, Tullio, *Minisemantica dei linguaggi non verbali e delle lingue*, Roma-Bari, 1997 Laterza.

tore, lo stampatore, ecc. Ciascuno di questi «agenti del testo»¹²⁹ era e in parte ancora è responsabile di una parte del processo di produzione del documento. Uno degli aspetti principali della digitalizzazione dunque consiste nel cambiamento del processo “produttivo” e “distributivo” della conoscenza scritta. È chiaro che oltre a implicare nuove (e non minori) sfide da parte di linguisti, filologi, paleografi, traduttori, editor, ecc., questo processo è in grado di ridisegnare i «confini identitari»¹³⁰ dei vari attori coinvolti: autori, editori, biblioteche, istituzioni scientifiche e ogni altra forma o luogo di scambio e trasmissione del sapere scritto.

Ciò che appare chiaro, al di là dell'elemento tecnologico, è che questi linguaggi possono essere considerati, sia dal punto di vista intrinseco che estrinseco, vere e proprie “metalingue” capaci di rappresentare e tradurre la conoscenza. Se le culture comunicano fra loro con l'ausilio di meccanismi metaculturali¹³¹, fra cui le traduzioni, si può proporre l'ipotesi di considerare la codifica (e a maggior ragione una codifica attraverso metalinguaggi di markup) un «sistema modellizzante secondario», cioè un tipo di (meta) traduzione culturale.

Per rimanere nel campo delle analogie, possiamo dire che la codifica è un processo semiotico che ha con la sua fonte un rapporto simile a quello che la scrittura ha con il parlato. La scrittura infatti non è mera trascrizione del parlato, ma fornisce un “modello concettuale” della dimensione verbale¹³²: «la scrittura è in via di principio metalinguistica»¹³³. Essa cioè è “rappresentazione” e non trascrizione del parlato. In modo simile, il passaggio dall'analogico al digitale comporta un duplice processo di rappresentazione che insieme “presuppone” e “conduce” alla definizione di un modello.

Come le lingue naturali, anche i linguaggi informatici proiettano sull'oggetto rappresentato le loro ragioni e la loro visione del mondo. E se «la lingua è in

¹²⁹ Schmidt, Desmond, *A Graphical Editor for Manuscripts*, «Literary and Linguistic Computing», 2006 pp. 341-351.

¹³⁰ Marramao, Giacomo, *Passaggio a occidente. Filosofia e globalizzazione*, Bollati Boringhieri, 2003, Torino.

¹³¹ Lotman, Jurij M. e Uspenskij, Boris A., *Eterogeneità e omogeneità delle culture. Postscriptum alle tesi collettive*, in J. M. Lotman 2006.

¹³² Duranti, Alessandro, *Linguistic Anthropology*, Cambridge University Press, 2000, Cambridge (trad. it. *Antropologia del linguaggio*, Meltemi, Roma).

¹³³ Olson, David, *On the relations between speech and writing*, in Pontecorvo, 1997.

realtà una “metalingua”, un sistema per parlare del mondo (corsivo mio)¹³⁴, la codifica “metalinguistica” contribuisce a creare una specifica “condizione meta-testuale” che produce, garantisce e al tempo stesso regola la fruizione e l’accesso al documento.

¹³⁴ Cardona, Giorgio Raimondo, *I sei lati del mondo. Linguaggio ed esperienza*, 1985, Laterza, Roma - Bari.

Tabelle

Crosswalk

Schermi di conversione

I. Marc21 to EAD, alcuni elementi

Marc21	EAD
041 Language	LangCode attribute in <language>
100 Main entry -- personal name	<origination> <persname> <origination> <famname>
110 Main entry -- corporate name	<origination> <corpname>
111 Main entry -- meeting name	<origination> <corpname>
130 Main entry -- uniform title	<unittitle>
240 Uniform title	<controlaccess> <title>
245 Title statement	<unittitle>
245\$f Title statement/inclusive dates	<unitdate type="inclusive">
245\$g Title statement/bulk dates	<unitdate type="bulk">
260\$c Date	<unitdate>
300 Physical description	<phystech>
500 Generate note	<odd> <note>
510 Citation/references	<bibliography>

II. Marc21 to Dublin Core Element set

MARC fields	DC Element	Implementation Notes
100, 110, 111, 700, 710, 711	Contributor	
720		
651, 662	Coverage	
751, 752		
	Creator	Creator element not used.

MARC fields	DC Element	Implementation Notes
008/07-10	Date	
260\$c\$g		
500-599, except 506, 530, 540, 546	Description	
340	Format	
856\$q		
020\$a, 022\$a, 024\$a	Identifier	
856\$u		
008/35-37	Language	
041\$a\$b\$d\$e\$f\$g\$h\$j		
546		
260\$a\$b	Publisher	
530, 760-787\$o\$t	Relation	
506, 540	Rights	
534\$t	Source	
786\$o\$t		
050, 060, 080, 082	Subject	
600, 610, 611, 630, 650, 653		
245, 246	Title	Repeat dc:title for each. Some applications may wish to include 210, 222, 240, 242, 243, and 247.
Leader06, Leader07	Type	See Appendix 2 for Leader-Type rules.

III. DublinCore to teiHeader (elementi del METS)

DC:ELEMENT	LOCATION OF DATA IN TEI HEADER
dc:title	tei.2/teiHeader/fileDesc/titleStmt/title
dc:creator	tei.2/teiHeader/fileDesc/titleStmt/author
dc:subject	tei.2/teiHeader/profileDesc/textClass/keywords /list/item
dc:description	tei.2/teiHeader/encodingDesc/projectDesc
	tei.2/teiHeader/encodingDesc/editorialDesc
	tei.2/teiHeader/encodingDesc/refsDecl
	Others?
dc:publisher	tei.2/teiHeader/fileDesc/publicationStmt/publis her/publisher
	and pubplace
dc:contributor	tei.2/teiHeader/fileDesc/titleStmt/editor
	tei.2/teiHeader/fileDesc/titleStmt/funder
	tei.2/teiHeader/fileDesc/titleStmt/sponsor
	tei.2/teiHeader/fileDesc/titleStmt/principle
	tei.2/teiHeader/fileDesc/respStmt/name/@reg
	(Names from the respStmt will only be pulled in the DC
	record when the catalog form of the name has been entered
	in the @reg)
dc:date	tei.2/teiHeader/fileDesc/publicationStmt/date
dc:type	tei.2/teiHeader/@type (for CDL TEI Form)
dc:format	text/xml

DC:ELEMENT	LOCATION OF DATA IN TEI HEADER
dc:identifier	tei.2/teiHeader/fileDesc/publicationStmt/idno/@type= ARK
dc:source	tei.2/teiHeader/sourceDesc/bibful/publicationS tmt/publisher, pubPlace, and date
	tei.2/teiHeader/sourceDesc/bibl
dc:language	tei.2/teiHeader/profileDesc/langUsage/language
dc:relation	tei.2/teiHeader/fileDesc/seriesStmt/title OR title/xref/@href= url
dc:coverage	N/A
dc:rights	tei.2/teiheader/fileDesc/publicationStmt/availab ility
	Or does CDL boilerplate this?

IV. teiHeader to MARC

TEI Tags	Recommendations	USMARC Tags
<teiHeader type="***">	Standards to which the header applies, e.g., <teiHeader type="ISBD(ER)">, <teiHeader type="AACR2">	
<fileDesc>		

TEI Tags	Recommendations	USMARC Tags
<p><titleStmt> <title type="****"></p>	<p>Only uniform title and main title should be entered here, e.g., <titleType="uniform"> or <titleType="main">. See <sourceDesc> for other title forms for documents where where a user might seek the documents under titles other than those assigned. Where a title is provided by the header creator rather than the document creator, the title should be enclosed in square brackets using standard English language conventions for editorial insertion.</p>	<p>130 240 245 246's</p>
<p><titleStmt> <author></p>	<p>Author of original source (electronic or print) should be entered into the <author> tag before the <respStmt>. Use discrete tags within <author> tag for "last name", "first name", "middle name", "date", "position title" to allow future flexibility in display, indexing, and in transferring to MARC. Whenever possible, establish or use nationally established forms of names. The name should be inverted and entered in the established form.</p>	<p>1XX 1st author 534 la 1st author 7XX 2nd and 3rd authors</p>
<p><titleStmt> <editor></p>	<p>Editor of original source (electronic or print) should be entered into the <editor> tag before the <respStmt>.</p>	<p>7XX all editors</p>
<p><respStmt> <resp>editor, etc. <name></p>	<p>The editor (also compiler, illustrator, etc.) of an <i>electronic</i> version should be entered into the appropriate tag in the <respStmt>. The name should be inverted and entered in the established form.</p>	<p>500 7XX le</p>
<p><editionStmt> <p></p>	<p>Caution: Remind users that the edition statement refers to the electronic piece, not the original item. This field should be used sparingly as there are currently no standards as to when versions become editions. Users should refer to the instructions in the TEI manual.</p>	<p>250</p>
<p><extent></p>	<p>Use the standard text: "ca. **** kilobytes".</p>	<p>256</p>

TEI Tags	Recommendations	USMARC Tags
<publicationStmt>	Caution: This statement describes the electronic file.	
<publisher>	The publisher is whoever has collected the electronic text and has made decisions concerning it.	260 lb
<pubPlace>		260 la
<distributor>	The distributor is whoever makes the electronic text available.	260 lb
<idno>	Any unique identification number determined by the publisher	90,099
<availability><p> <p n="restriction, public, etc.">	Use specialized elements when anticipating sharing of the header or free text if only local usage is expected. Caution: Know your audience.	500 590 local restriction note 856 4X l3 lz lu
<date>	Refers to the date of the publication of the electronic document.	Fixed field: Date1 260 lc
<seriesStmt> <p>	Whenever possible, establish the national authority file authorized form for the electronic locally created series.	440
<notesStmt> <note>	Optional, depending on display decisions. Should be used for indicating questionable attributions for title, author, etc.	500
<SourceDesc>	In order to effectively represent the source(s) when many documents are represented by the TEI header, we see the need for structured elements that minimally allow us to identify parent-child and component relationships. In the absence of these structures, we suggest that multiple source descriptions be employed with relationships described in free text. Relationships also could be useful in other portions of the TEI header. Cataloger may need to do research to establish the original source.	
<bibl> <biblStruct> <biblFull>	Prefer <biblFull> to allow searching on parts of the description.	

TEI Tags	Recommendations	USMARC Tags
<biblFull> <titleStmt> <title>	It is possible to have multiple <titles> in <biblFull>. Alternative titles (cover, running, spine titles) should be entered in separate <title> fields in the <biblFull> field in the <sourceDesc> where they are searchable.	246 700 X2 la lt 730 740 534 lt
<biblFull> <titleStmt> <author>	If the name of the author(s) in the originating source differs from the established form, include here the form from the source tagged <author type="alternate">.	
<biblFull> <editionStmt> <p>	Enter edition statement as found on the original source.	534 lb
<biblFull> <extent>	Enter physical description for the original source.	534 le
<biblFull> <publicationStmt> <publisher>	Don't repeat field. Enter multiple publishers divided by semicolons.	534 lc_2
<biblFull> <publicationStmt> <pubPlace>	Don't repeat field. Enter multiple publishers divided by semicolons.	534 lc_1
<biblFull> <publicationStmt> <date>	Imprint date for the original source.	534 lc_3
<biblFull> <publicationStmt> <idno type="***">	In this location, <idno> refers to identification numbers for the source document. They can be used to indicate the source's location in an individual institution's collection. If a formal standard location system is being used, indicate the nature of the system, e.g., <idno type="LC call number">.	534 ln
<biblFull> <seriesStmt> <p>	Establish via national authority file the series statement of original document.	534 lf
<biblFull> <notesStmt> <note>	Caution: Notes made here should refer to the original source.	534 ln
<encodingDesc>		
<projectDesc> <p>	Enter a description of the purpose for which the electronic file was encoded.	500 550

TEI Tags	Recommendations	USMARC Tags
<editorialDecl> <p>	Enter general and specific statements about how the electronic file has been treated. Record here editorial decisions made during encoding.	516 500
<refsDecl> <p>	<refsDecl> seems a possibility for administrative metadata, e.g., pagination and page sequencing.	500 516
<classDecl> <taxonomy id= "*" >	If used, identify the appropriate taxonomy definitions or descriptive sources in the <taxonomy> tag followed by id, e.g., <taxonomy id=LCSH>, <taxonomy id=AAT>.	
<profileDesc>		
<creation> <date>	Use the date as it comes from the creator.	
<langUsage> <language id="*" >	Language usage is specified by document creators. Use standard language names. Use the ISO 639-2 standard (which is the same as the MARC language codes).	041 546
<textClass> <keywords> <term> <keywords scheme="LCSH"> <term type="*" >	True classification numbers as opposed to call numbers can be entered here. Use for controlled vocabulary as specified in <encodingDesc> taxonomy id.	653, 654 (for AAT), 655, 690, 691 600, 610, 611, 630, 650, 651
revisionDesc		
<change> <date> <resp> <name> <item>	Use the specific codes to note revisions rather than prose description. Include the entire date.	9XX local processing note

Fonti cartacee Catalogo

Guerini e Associati:

- ✦ **schede di produzione**
- ✦ **catalogo cartaceo**

SCHEDA MOVIMENTAZIONE EDITORIALE

A. Area editoriale

Commessa N° _____ Data _____ ISBN C.C.A. _____

Editor _____ Area _____ Settore _____

Linea editoriale (Concept) _____

Collana _____ N° _____

AUTORE/I _____

CURATORE/I _____

TITOLO _____

Sottotitolo _____

Notizie autore/i _____

Notizie curatore/i _____

Caratteristiche e obiettivi di mercato _____

Inedito Nuova edizione [.....^a] ^a Ristampa [della^a edizione]

Marchio: Guerini e Associati Guerini Studio Guerini Scientifica Cantiere Italia

Data accettazione originali _____ Data produzione richiesta _____

Data ricevimento contratto _____ Cedola _____

B. Categoria di copertura

(in collaborazione con Direzione Commerciale)

Data _____

Locomotiva

Vagone

Adozione

Servizio

Note _____

Recapito autore/curatore _____

Particolari esigenze relative alla data di pubblicazione _____

Particolari avvertenze circa l'autore _____

Inserimento giro di vendite di _____

Lancio previsto nel mese di _____ Settimana _____ (1, 2, 3 o 4)

Prenotato Messaggerie _____ Scorte _____

Tiratura proposta _____ Prezzo al pubblico proposto _____ euro

Tiratura definitiva _____

Visto Editor proponente _____

Primo visto Direzione Editoriale _____

C. Caratteristiche tecniche proposte

Data _____

DIMENSIONI

	Max	Min	Finali		Max	Min	Finali
Formato	_____	_____	_____	Battute	_____	_____	_____
Tabelle, Figure	_____	_____	_____	Totale pagine	_____	_____	_____

Materiali in _____

Parti mancanti _____

Apparati richiesti _____

Quarta e «Dal Catalogo» _____

INTERNI

Progetto grafico

Stampa da pellicola

Carta gr. _____

Carta gr. _____

Composizione tipo _____

CTP

Tipo _____

Tipo _____

COPERTINA

Progetto grafico

Con alette

Supporto

Altre lavorazioni

Colori _____

Con sovracoperta dorso _____

patinato uso mano altro _____

vernice plastificazione lucida plastificazione opaca

LEGATORIA

Brossura fresata

Bollino SIAE

Brossura cucita Cartonato Punto metallico

Altre lavorazioni _____

DATI CONTRATTI

Contratto autore

Contratto editore

Approvazione traduzione

Termini pubblicazione

Royalty

Contratto curatore

Contratto traduttore

Contratti sussidiari

Diciture per colophon

Data: _____ Scadenza: _____

Data: _____ Scadenza: _____

Data: _____ Scadenza: _____

Data: _____

Visto Ufficio Diritti _____

Visto Editor proponente _____

Secondo visto Direzione Editoriale _____

CONTROLLI FINALI

DATA _____

REDATTORE _____

- sequenza testo da una pagina all'altra + sillabazione

successione numerica

- numeri di pagina
 capitoli sottocapitoli
 tabelle figure

testatine

- corrispondenza titoli
 allineamento a destra o sinistra
 posizione numeri di pagina

altezze

- testo, inizio capitoli
 spazio fra testatina/numero e testo
 epigrafe (o esergo)
 titolo parti

prime pagine

- occhiello collana
frontespizio
 autore
 titolo
 sottotitolo (segnalare eventuale mancanza)
 curatela
 prefazione/postfazione/introduzione
 editore

copyright

- ISBN
 edizione originale titolo originale
 traduttore
 eventuali copyright di altri (autore, istituzioni)
 segnalazione eventuali contributi (CNR, Università ecc.)
 progetto copertina

indice: corrispondenza

- titoli
 autori
 numeri di pagina

verifica elenchi collana

- elenco collana aggiornato
 dal catalogo Guerini

corrispondenze

- tabelle/figure con riferimenti nel testo
 note con esponenti di nota

uniformazione interna

- occhielli/gerarchia titoli
 tabelle, figure

uniformazione alla collana

- occhiello collana
 frontespizio
 occhielli/titoli
 gabbia
 carattere
 impostazione indice

- Bibliografia: verifica traduzioni in italiano
Guerini dei volumi esteri

COPERTINA

verifica col frontespizio

- autore
 titolo
 sottotitolo (segnalare eventuale mancanza)
 curatela
 prefazione

marchio

- editore
 Guerini e Associati
 Guerini Studio
 Guerini Scientifica
 eventuale logo di sponsor

verifica con volumi della stessa collana

- impostazione grafica/corpi/box quarta
 appartenenza alla collana
 numero di collana (in costa o sul piatto)
 titoli sull'aletta

verifica con la scheda

- prezzo di copertina
 dicitura «copia fuori commercio»
 copie senza prezzo
 isbn

misure (verifica con gabbia testo e dati scheda)

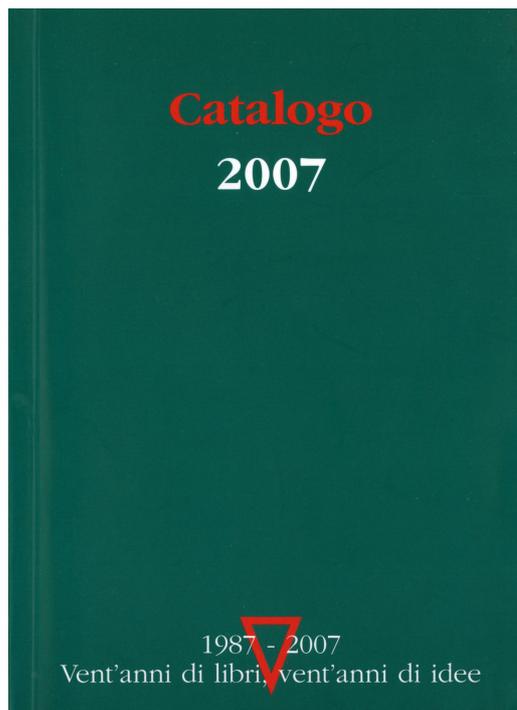
- base
 altezza
 dorso alette

colori

supporto

- cartoncino
 plastificazione
 cordonatura

- Verifica visto si stampi autore



62 *sapere per fare*

VARANINI FRANCESCO, **Le parole del manager. 108 voci per capire l'impresa**



GA, 2006, pp. 238, € 9,50,
88-8335-737-X

VARANINI FRANCESCO, **L'irresistibile ascesa del direttore marketing cresciuto alla scuola del largo consumo**

«... penso che questo libretto sia un messaggio. Regalandolo riuscirete forse a dire con un gesto quello che non siete mai riusciti a far capire al vostro capo, ai vostri collaboratori...»



GA (BP), 2003, pp. 94, € 9,50, 88-8335-472-9

VARCHETTA GIUSEPPE, **Ascoltando Primo Levi. Organizzazione, narrazione, etica**



GA (NF), 1998*, pp. 107, € 11,50,
88-7802-895-9

Primo Levi

sapere per fare 63

VARCHETTA GIUSEPPE, **Liste. Storie dall'organizzazione**



GA (M), 2005, pp. 175, € 17,50,
88-8335-651-9

Tredici storie di vita quotidiana tratte dal variegato mondo dell'organizzazione. Che si ritrova, ancora oggi, pensata per pochi, carente di riflessioni e ricolma di operatività a breve termine. Tredici racconti che portano in sé il lodevole tentativo di trattare temi urgenti per il vivere aziendale trasmettendo dal consueto linguaggio saggistico a quello narrativo.

VARCHETTA GIUSEPPE, **La solidarietà organizzativa**



GA (LA), 1993, pp. 222, € 23,00,
88-7802-369-8

Il libro si fonda sull'ipotesi di una nuova vivibilità degli obiettivi e di un territorio nuovo in cui non siamo condannati a un ineluttabile moltiplicarsi confuso di punti di vista, ma invitati a ricercare la coesione dell'esperienza in una nuova intersoggettività degli attori organizzativi.

VARCHETTA GIUSEPPE, **Emergenze organizzative. Testi e storie**



GST, 1997, pp. 150, € 15,00, 88-7802-739-1

Le riflessioni proposte in questo volume – la leadership e la crisi della gerarchia, la partnership, il conflitto, l'apprendimento organizzativo e lo sviluppo delle competenze, la necessità di mutare i paradigmi dei meccanismi operativi – privilegiano un'interpretazione soggettuale dell'esperienza e dell'azione organizzativa.

Scansione del catalogo cartaceo, in particolare quello rappresentato è del 2007, e sotto vi è un esempio di illustrazione dei titoli in vendita.

MetaMap

MetaMap (Febbraio 2003) propone uno strumento grafico per la gestione dei metadati sviluppato sul modello di mappa dei servizi di metropolitana: questa metafora aiuta gli utenti a navigare in un *metaspazio*, l'ambiente dei metadati. Si tratta infatti di un grafico di utilità per la comunità scientifica nella comprensione degli standard, dei set e delle iniziative di metadati.

MetaMap è stato prodotto presso l'Università di Montreal, il progetto impiega la tecnica Scalable Vector Graphics (VGS). Nella banca dati sono raccolte le informazioni sugli standard, i set, le iniziative di metadati (Metadata Standards, Sets, Initiatives - *MSSI*) e le organizzazioni coinvolte nella loro realizzazione, ed ognuno di essi è rappresentato come una stazione della metropolitana sulla linea appropriata: le stazioni sono denominate dall'acronimo del MSSI e ogni linea rappresenta un ambito tematico diverso di metadati. Il click del mouse sul nome della stazione porta in evidenza una spiegazione del MSSI e il collegamento al sito web ufficiale e ai siti correlati. Attualmente il progetto registra 184 stazioni e viene aggiornato periodicamente per riflettere lo sviluppo dei metadati.¹³⁵

Come è stato più volte sottolineato, lo sviluppo del web comporta una proliferazione di standard e di set di metadati e il problema è esacerbato dal fatto che le loro denominazioni sono rappresentate spesso da acronimi: MetaMap aiuta a raccogliere in un ambiente unico queste informazioni mostrando le relazioni tra di esse e connettendole con i vari *attori* coinvolti nella loro creazione e nel loro uso.

Ogni linea rappresenta uno dei temi scelti ed ogni MSSI appare come una stazione sulla linea. Quando diversi MSSI sono collegati a più di una linea (ossia a più di un ambito tematico) il progetto li rappresenta come nodi nella rete.

Nella stessa mappa sono inclusi diversi approcci:¹³⁶

¹³⁵ <<http://mapageweb.umontreal.ca/turner/meta/english/index.html>>.

¹³⁶ Una linea speciale enfatizza il ruolo nello sviluppo dei MSSI di organizzazioni quali W3C, IFLA, IETF e IEEE.

- 🌻 processi della gestione dell'informazione (creazione, disseminazione, conservazione);
- 🌻 istituzioni con esperienza nella gestione dell'informazione (biblioteche, archivi, musei);
- 🌻 tipi di file di informazione gestiti (testi, immagini fisse, immagini in movimento, suoni).

I processi sono rappresentati secondo la seguente logica:

- 🌻 le linee orizzontali sono usate per i processi delle informazioni;
- 🌻 le linee verticali sono usate per le istituzioni con esperienza nella gestione dell'informazione;
- 🌻 le linee che rappresentano i tipi di informazione gestiti sono raggruppate insieme in un'area della mappa e parti di esse passano in parallelo per riflettere gli elementi che hanno in comune;
- 🌻 le linee relative alle organizzazioni girano attraverso il centro della mappa per potersi connettere a più linee possibili.

Altro principio organizzativo è che i nodi sono creati dove un MSSI è chiaramente collegato ad uno o più temi. Per esempio TEI è parte delle linee delle organizzazioni che incrocia il trattamento di testi.

Bibliografia

[1] AA. VV., Atti della III Conferenza Nazionale delle Biblioteche, *La Bibliothec@ digitale, produzione, gestione e conservazione della memoria nell'era digitale*, Ministero per i Beni e le Attività Culturali, Roma, 2002, Tiellemmedia Editore.

[2] Bassi Maria Cristina, *La catalogazione delle risorse informative in Internet*, Milano, 2002, Editrice Bibliografica.

[3] Battles Matthew, *Biblioteche: una storia inquieta. Conservare e distruggere il sapere da Alessandria a Internet*, traduzione di Idolina Landolfi, Roma, 2004, Carocci (*Library: an unquiet story*, London, 2003, William Heinemann).

[4] Benvenuti Nicola, *L'importanza dei metadati nella costruzione della rete globale: riflessioni a margine della Conferenza mondiale del Dcmi, Firenze, 13-17 ottobre 2002*, «Bollettino Aib», 43 (2003), n. 3, pp. 331-350.

[5] Burnard Lou, O'Brien O'Keeffe Katherine and Unsworth John, *Electronic Textual Editing*, The Modern Language of America, New York, 2006.

[6] Brenner Diane and Rowland Marilyn, Medford-Phoenix, *Beyond book indexing*, Information today-American society of indexers, 2000.

[7] Cesana Roberta, *Editori e librai nell'era digitale. Dalla distribuzione tradizionale al commercio elettronico*, Milano, 2002, editore Franco Angeli.

[8] Chu Heting, *Information representation and retrieval in the digital age*, Medford: Information today-Asist, 2003.

[9] De Mauro Tullio, *Minisemantica dei linguaggi non verbali e delle lingue*, Roma-Bari, 1997, Laterza.

- [10] De Robbio Antonella, *Metadati: parola chiave per l'accesso alla biblioteca ibrida*, pp. 103-128 , in *La biblioteca ibrida: verso un servizio informativo integrato*, Milano, 2003, Editrice Bibliografica.
- [11] Dini Rossella, *ISBD(G) / International Federation of Library Associations and Institutions*. - Ed. italiana - Roma, 1987, Associazione italiana biblioteche.
- [12] Fadini B., Savy C., *Informatica per le scienze umane*, Milano, 1999, editore Franco Angelini.
- [13] Gambari Stefano e Guerrini Mauro, *Definire e catalogare le risorse elettroniche: un'introduzione a ISBD(ER), AACR2 e metadati*, Milano, 2002, Editrice Bibliografica.
- [14] Ghilli C., Guerrini M., *Introduzione a FRBR. Functional Requirements for Bibliographic Records*, Requisiti funzionali per record bibliografici, Milano, 2001, Editrice Bibliografica.
- [15] Gorman Gary Eugene, *Metadata applications and management (International yearbook of library and information management 2003/2004)*, London, 2004, Facet.
- [16] Guercio Maria, *Archivistica informatica: i documenti in ambiente digitale*, Roma, 2002, Carocci.
- [17] Haynes David, *Metadata for information management and retrieval*, London, 2004, Facet.
- [18] Henninger Maureen, *The hidden Web: finding quality information on the net*, Sidney, 2003, University of New South Wales press.
- [19] Lesley M. Moyo, *Collections on the Web: some access and navigation issues*, «Library collections, acquisitions and technical services», 2002
- [20] Licklider J. C. R, *Libraries of the future*, Cambridge, 1965, M.I.T. Press.

- [21] Longo Brunella, *La nuova editoria: mercato, strumenti e linguaggi del libro in Internet*, Milano, 2001, Editrice Bibliografica.
- [22] Lynch Clyfford, *Building digital libraries for the metadiversity: federation across disciplines*, Pre-prints of the II Metadiversity Conference Proceedings, Charleston, SC, June 2002.
- [23] Maltese D., Guerini M., *Il futuro della descrizione bibliografica*, Atti della giornata di studio, Firenze 13 Novembre 1987, AIB, Roma, 1988.
- [24] Marquardt L. e Sararelli A., *Comunicare la conoscenza ai tempi del web: ruolo e responsabilità delle biblioteca ibrida*, Roma, 2005, AIB.
- [25] Morcellini Mario, *Il secondo Novecento italiano come "Evo medio"*, in Mario Morcellini e Paolo De Nardis, *Società e industria culturale in Italia*, Roma, 1998, Meltemi.
- [26] Nanetti M., Pistarelli V. e Santoro L., *La biblioteca digitale: produzione, gestione e conservazione della memoria nell'era digitale*, atti della III Conferenza nazionale delle biblioteche, Padova, Biblioteca del Monumento nazionale di Santa Giustina, 14-16 febbraio 2001, Roma, 2002, Ministero per i Beni e le Attività Culturali-Tiellemedia.
- [27] Parise Stefano con la collaborazione di Massimiliano Spinello, *La biblioteca nel web: tecniche e strategie per un sito Internet di qualità*, atti del seminario, Rozzano, Centro culturale cascina grande, 28 giugno 2002, Milano: Aib. Sezione Lombardia, 2003.
- [28] Peresson G. e Mazzucchi P., *Il futuro dell'editoria in rete. La presenza delle case editrici su internet: editori vs fornitori?*, Milano, 2000, AIE.
- [29] Smith K. Wayne (editor), OCLC, *1967-1997: thirty years of futhering access to the world's information*, New York, 1998, Haworth Press.

- [30] Solimine Giovanni, *Controllo bibliografico universale*, Roma, 1995, AIB.
- [31] Solimine Giovanni, *Gestire il cambiamento*, Milano, 2003, Editrice Bibliografica.
- [32] Tidwell Doug, *XSLT*, Sebastopol, CA, 2001, O'Reilly & Associates.
- [33] Vigni Giuliano, *Rapporto 2000 sull'editoria italiana*, in AIE, Catalogo degli editori italiani, Milano, 2001, Editrice Bibliografica.

Webliografia

[1] Bacha Murtha, *Introduction to metadata: pathways to digital information*, Getty research institute, last updated 7-5-2000, <http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html>. Visitato il 10 Settembre 2009.

[2] Bib: bibliografia italiana delle biblioteche, del libro e dell'informazione, a cura di Alberto Petrucciani e Giulia Visintin, in Aib-Web, ultimo aggiornamento dicembre 2002, <<http://www.aib.it/aib/bib/bib.htm>>. Visitato il 1 Settembre 2009.

[3] Cathro Warwick, *Metadata: an overview, standards Australia seminar «Matching discovery and recovery»*, August 1997, <<http://www.nla.gov.au/nla/staffpaper/cathro3.html>>. Visitato il 3 Agosto 2009.

[4] CENN/ISSS - Information Society Standardization System), <<http://www.cen.eu/cenorm/homepage.htm>>. Visitato il 10 Settembre 2009.

[5] Commission on Preservation and Access and Research Libraries Group, *Report of the task force on archiving and digital information*, Washington, DC, 1996, <<http://www.oclc.org/community/rlg/transitions/>>. Visitato il 10 Luglio 2009.

[6] Conway Paul, *Preservation in the digital world*, Washington, DC, Commission on Preservation and Access, 1996, <<http://www.clir.org/pubs/reports/conway2/>>. Visitato il 10 Settembre 2009.

[7] Covey Denise, *Usage and usability assessment*, Washington, DC, CLIR, 2002, <<http://www.clir.org/pubs/reports/pub105/contents.html>>. Visitato il 20 Settembre 2009.

- [8] EAD, Encoded Archival Description, <http://www.loc.gov/ead/tglib/appendix_d.html>. Visitato il 15 Settembre 2009.
- [9] DC-2002. *Metadata for e-communities: supporting diversity and convergence*, proceedings of the international conference on Dublin core and metadata for e-communities, Florence, Italy, October 13-17, 2002, Firenze University Press, oppure <<http://www.bncf.net/dc2002>>. Visitato il 1 Ottobre 2009.
- [10] DC-1: CLC/NCSA Metadata workshop. *The essential elements of network object description*, <<http://dublincore.org/workshops/dci/>>. Visitato il 1 Ottobre 2009.
- [11] DCMI. Dublin Core Metadata Initiative. <<http://www.dublincore.org>>. Visitato il 1 Ottobre 2009.
- [12] Day Micheal, *The metadata challenge for libraries: a view from Europe, Metadiversity - A Call to Action. Responding to the Grand Challenge for Biodiversity Information Management through Metadata*, Natural Bridge, Virginia, November 1998, <<http://www.ukoln.ac.uk/metadata/presentations/metadiversity/challenge.html>>. Visitato il 10 Ottobre 2009.
- [13] DCMI Namespace for the Dublin Core Metadata Element Set, Version 1.1, <<http://dublincore.org/2008/01/14/dcelements.rdf>>. Visitato il 1 Luglio 2009.
- [14] DELOS Network of Excellence on Digital Libraries <<http://delos-noe.iei.pi.cnr.it/>>. Visitato il 20 Ottobre 2009.
- [15] Dempsey L. and Rachel Heery, *Specification for resources description methods. Part 1: A review of Metadata: a survey of current resources description formats*, deliverable A review of metadata - Development of a European Service for Information on Research and Education, March 1997,

<<http://www.ukoln.ac.uk/metadata/desire/overview/>>. Visitato il 20 Agosto 2009.

[16] Digital Library Initiative-phase 2, <<http://www.nsf.gov/pubs/1998/nsf9863/nsf9863.htm>>. Visitato il 20 Agosto 2009.

[17] Dillon Martin, *Metadata for web resources: how metadata works on the web*, Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenge of Networked Resources and the Web, <http://www.loc.gov/catdir/bibcontrol/dillon_paper.html>. Visitato il 30 Agosto 2009.

[18] DOI, Digital Object Identifier, <<http://www.doi.org/>>. Visitato il 20 Settembre 2009.

[19] DSpace, open source software, <<http://www.dspace.org>>. Visitato il 2 Ottobre 2009.

[20] EDItEUR, international group coordinating development of the standards infrastructure for electronic commerce in the book and serials sectors, <<http://www.editeur.org>>. Visitato il 25 Agosto 2009.

[21] ERA, Electronic Records Archives, <<http://www.archives.gov/era/>>. Visitato il 20 Luglio 2009.

[22] OCLC Digital Archive, <<http://www.oclc.org/digitalarchive/>>. Visitato il 20 Luglio 2009.

[23] Erpanet, Erpaworkshop - *Trusted repositories for preserving cultural heritage*, ERPANET Workshop, Rome, November 2003, <<http://www.erpanet.org/events/2003/rome/RomeFinalReport.pdf>>. Visitato il 20 Settembre 2009.

- [24] FedoraCommons, open-source digital object repository system, <<http://www.fedora-commons.org/>>. Visitato il 2 Luglio 2009.
- [25] FRBR, Functional requirements for bibliographic records, <<http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>>. Visitato il 12 Luglio 2009.
- [26] GEM, Gateway to Educational Metadata, <<http://67.23.47.252/about/documentation2/schemas>>. Visitato il 4 Ottobre 2009.
- [27] Gilliland-Swetland, *Setting the stage. Introduction to metadata*, Pathways to Digital Information, May 2000, <<http://www.slis.kent.edu/~mzeng/metadata/Gilliland.pdf>>. Visitato il 20 Luglio 2009.
- [28] Google Book, code, <<http://code.google.com/intl/it-IT/apis/books/docs/gdata/reference.html>>. Visitato il 27 Ottobre 2009.
- [29] Greenstein Daniel, Next Generation digital libraries, Victorian Association for Library Automation, February 2002, <www.vala.org.au/vala2002/2002pdf/01Grnstn.pdf>. Visitato il 20 Luglio 2009.
- [30] Greenstein Daniele, *Next-generation digital libraries*, Joint Conference on Digital Libraries, July 2002, Portland, Oregon, <<http://www.dlib.org/dlib/september02/rasmussen/09rasmussen.html>>. Visitato il 20 Luglio 2009.
- [31] Greenstone, Digital Library Software, <<http://www.greenstone.org/>>. Visitato il 15 Settembre 2009.

[32] Gruppo di lavoro sui metadati dell'Istituto centrale per il catalogo unico, <<http://www.iccu.sbn.it/genera.jsp?id=101#nota%201>>. Visitato il 15 Settembre 2009.

[33] Hildreth Charles R, *On Line catalog design models: are we moving in the right direction?* A report submitted to the Council on library resources, August 1995, updated March 2000, <<http://myweb.cwpost.liu.edu/childret/clr-opac.html>>. Visitato il 15 Settembre 2009.

[34] IFLA, Functional Requirements for Bibliographic Records, <<http://www.ifla.org/VII/s13/frbr/frbr.htm>>. Visitato il 30 Settembre 2009.

[35] InterPARES (International Research on Permanent Authentic Records in Electronic System) 1 report: *The long-term preservation of Authentic electronic records: findings of the InterPARES project*, Vancouver, Canada, September 2002, <<http://www.gseis.ucla.edu/us-interpares/pdf/InterPARES1FinalReport.pdf>>. Visitato il 15 Ottobre 2009.

[36] IMS Global Learning Consortium, <<http://www.imsglobal.org/>>. Visitato il 5 Settembre 2009.

[37] ISBN, <<http://www.isbn.it/>>. Visitato il 18 Settembre 2009.

[38] ISO 14721:2003, Space Data and information transfer system - Open archival information system - Reference model, <http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683>. Visitato il 18 Settembre 2009.

[39] ISTAT, *La produzione libraria*, Anno 2006, Roma 2008, tavola 1, <http://www.istat.it/dati/dataset/20080415_00/>. Visitato il 15 Settembre 2009.

[40] Leombroni Claudio, *Appunti per un'ontologia delle biblioteche digitali: considerazioni sulla Biblioteca digitale italiana*, «Bollettino Aib», 44 (2004), n. 2, pp. 115-

131, oppure <<http://www.aib.it/aib/boll/2004/0402115.htm>>. Visitato il 23 Settembre 2009.

[41] Lee Tim Beerns, *Metadata architecture*, January 1997, updated September 2000, <<http://www.w3.org/DesignIssues/Metadata>>. Visitato il 23 Luglio 2009.

[42] Library of Congress National Digital Library Program, “*American Memory*” project: *digital library users*, <<http://memory.loc.gov/ammem/dli2/html/lcndlp.html>>, Cornell University and the University of Michigan, *Making of America*, <<http://quod.lib.umich.edu/m/moagrp/>>, <<http://digital.library.cornell.edu/m/moa/>>. Visitato il 23 Luglio 2009.

[43] LOC. METS - Metadata Encoding & Transmission Standard, <<http://www.loc.gov/standards/mets/>>. Visitato il 15 Luglio 2009.

[44] LOC. ZING, Z39.50-International: Next Generation, <<http://www.loc.gov/z3950/agency/zing/>>. Visitato il 18 Ottobre 2009.

[45] LOC. Z39.50 Maintenance Agency, <<http://www.loc.gov/z3950/agency/>>. Visitato il 18 Ottobre 2009.

[46] LOM, <http://ltsc.ieee.org/wg12/files/LOM_I484_I2_I_VI_Final_Draft.pdf>. Visitato il 24 Ottobre 2009.

[47] Lougee Wendy Pradt, *Diffuse Libraries: emergent roles for the research library in the digital age*, Council of Libraries and Information, <<http://www.clir.org/pubs/reports/pub108/contents.html>>. Visitato il 30 Ottobre 2009.

[48] MetaMap, <<http://mapageweb.umontreal.ca/turner/meta/english/index.html>>. Visitato il 28 Ottobre 2009.

[49] N.McLean and C.Lynch, A Joint, White Paper on behalf of the IMS Global Learning Consortium and the Coalition for Networked Information, Joint White Paper - *Interoperability between information and learning environments. Bridging the gaps*, DRAFT, June 2003, <http://www.imsglobal.org/dlims_white_paper_publicdraft_1.pdf>. Visitato il 18 Settembre 2009.

[50] NEDLIB. <<http://www.konbib.nl/nedlib>>. Visitato il 10 Ottobre 2009.

[51] *New South Wales recordkeeping metadata standard*, September 2000, <<http://www.records.nsw.gov.au/recordkeeping/dirks/doing-a-dirks-project/specify-and-apply-recordkeeping-metadata#what-is-recordkeeping-metadata>>. Visitato il 18 Ottobre 2009.

[52] OAI, Open Archives Initiative, <<http://www.openarchives.org/>> . Visitato il 29 Ottobre 2009.

[53] OAF, Open Archives Forum, <<http://www.oaforum.org/>>. Visitato il 27 Settembre 2009.

[54] OCLC, The World's libraries Connected, <<http://www.oclc.org/research/projects/archive/wcp/>>. Visitato il 27 Settembre 2009.

[55] OCLC E-Learning Task Force, *White Paper -Libraries and the Enhancement of E-learning*, October 2003, <<http://www5.oclc.org/downloads/community/elearning.pdf>>. Visitato il 27 Settembre 2009.

[56] ONIX, ONline Information eXchange, <<http://libraries.mit.edu/guides/subjects/metadata/standards/onix.html>>. Visitato il 18 Luglio 2009.

[57] Paci Augusta Maria, *Orientare i sistemi di informazione all'utente: gli obiettivi del progetto di ricerca europeo SESAM*, Nuovi annali della Scuola speciale per archivisti e bibliotecari, a. XII, 1998. Sito web del progetto: <<http://cordis.europa.eu/libraries/en/projects/sesam.html>>. Visitato il 7 Settembre 2009.

[58] Piez Wendell, *Something called Digital Humanities*, in "Digital Humanities Quarterly", <<http://www.digitalhumanities.org/dhq/vol/2/1/000020.html>>. Visitato il 30 Ottobre 2009.

[59] SCHEMAS, <<http://www.schemas-forum.org/>> e SCHEMAS Design, <<http://www.schemas-forum.org/info-services/info-mats.html#Metadata>>. Visitato il 12 Settembre 2009.

[60] UNIMARC, direttive dell'IFLA <<http://www.ifla.org/en/unimarc>>. Visitato il 20 Settembre 2009.

[61] Swearingen Kirsten, *How Much Information 2003?* Researchers: Peter Charles, Nathan Good, Laheem Lamar Jordan, Joyojeet Pal, <<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>>. Visitato il 27 Ottobre 2009.

[62] TauRo-Core, Signum, Scuola Normale Superiore, Pisa, <<http://tauro.signum.sns.it/>>. Visitato il 25 Luglio 2009.

[63] TEI-c initiative: <<http://www.tei-c.org/index.xml>>. Visitato il 7 Luglio 2009.

[64] TEI@Oxford Summer Camp, <<http://tei.oucs.ox.ac.uk/Oxford/2009-07-oxford/>>. Visitato il 7 Luglio 2009.

[65] T. O'Neill Edward, Brain F. Lavoie, Rick Bennett, *Trends in the evolution of the public web*, "D-Lib Magazine", April 2003,

<<http://www.dlib.org/dlib/aprilo3/lavoie/o4lavoie.html>>. Visitato il 27 Settembre 2009.

[66] Valacchi Federico, Vitali Stefano e Zorzi Andrea, *Archivi storici e archivi digitali tra ricerca e comunicazione*, materiali del workshop, Università di Firenze. Dipartimento di studi storici e geografici in collaborazione con l'Archivio di Stato di Firenze, 20-21 ottobre 2000, <http://www.storia.unifi.it/_storinforma/Ws/archivi/ws-archivi-alia.htm>, visitato il 1 Settembre 2009.

[67] W3C. Web Services Activities, <<http://www.w3c.org/2002/ws/>> . Visitato il 7 Settembre 2009.

[68] Working Group on Digital Libraries Metrics, <<http://www.dlib.org/metrics/public/>>. Visitato il 27 Agosto 2009.

[69] XML, eXtensible Markup Language, <<http://www.w3.org/XML/>>. Visitato il 20 Ottobre 2009.

[70] XSLT, stands for XSL Transformations, <<http://www.w3.org/TR/xslt>>. Visitato il 3 Settembre 2009.