

# Appunti di informatica per le biblioteche

di Giovanna Frigimelica e Andrea Marchitelli

Settembre 2011

## SOMMARIO

Definizioni .....	4
Sistemi informativi, sistemi informatici e database.....	4
Formati di file.....	5
Information Retrieval (IR) .....	6
Metadati .....	11
Linguaggi di marcatura e standard bibliografici.....	15
eXtensible Markup Language (XML) .....	16
MAchine-Readable Cataloging (MARC) .....	17
Protocollo Z39.50.....	19
STRUMENTI.....	21
Sistemi integrati di automazione per biblioteche.....	21
Online Public Access Catalog (OPAC).....	24
Nuove tecnologie per la ricerca.....	27
Digitalizzazione .....	29
Open Access (Accesso aperto) .....	30
Risorse .....	34
Biblioteca digitale .....	34
Periodici elettronici (E-JOURNAL) .....	35
E-book.....	35
Banche dati bibliografiche e indici citazionali.....	38
Il web .....	38
L'AUTOMAZIONE DELLE BIBLIOTECHE IN ITALIA .....	41
Breve storia dell'automazione nelle biblioteche italiane .....	41
SBN.....	43
L'automazione delle biblioteche italiane oggi .....	46
BIBLIOGRAFIA .....	50
Indice delle immagini.....	51
Indice analitico.....	51

L'impostazione e la redazione del testo nel suo complesso sono frutto della collaborazione fra i due autori. A Marchitelli vanno ascritti i capitoli Definizioni, Strumenti (Digitalizzazione e *Open access*), Risorse (biblioteca digitale, periodici elettronici, e-book). A Frigimelica i capitoli Strumenti (Sistemi integrati, OPAC, nuove tecnologie per la ricerca), Risorse (Banche dati bibliografiche e indici citazionali, il Web), L'automazione delle biblioteche in Italia.

## DEFINIZIONI

### SISTEMI INFORMATIVI, SISTEMI INFORMATICI E DATABASE

Spesso usati in modo interscambiabile, i termini *dato*, *informazione* e *conoscenza* hanno in realtà significati diversi:

- il **dato** è un singolo elemento informativo: il nome o il formato di un file, la data di edizione di un volume, una transazione di commercio elettronico, il titolo di un libro, una cella di Excel, un'immagine, il numero di abitanti di una città, il numero di accessi a un sito web. La rappresentazione dei dati in formati leggibili dall'elaboratore permette la gestione degli stessi attraverso il calcolatore e li rende utilizzabili in quantità e con velocità sempre maggiori;
- l'**informazione** è un insieme di dati (numeri, immagini, parole) collocato in un contesto di riferimento con rilevanza e obiettivo: la voce di un glossario o di un dizionario, un'immagine accompagnata da una notizia, un certificato anagrafico, una tabella di dati, un riferimento bibliografico, un articolo, una e-mail, un volantino. Sempre più le informazioni nascono ed esistono in forma digitale;
- la **conoscenza** è saper usare e produrre informazioni: saperle selezionare, saperne valutare la qualità e l'utilità, saperle gestire, saperle collegare, saperle manipolare per produrne di nuove allo scopo, ad esempio, di progettare una ricerca, scrivere un libro, redigere un articolo, fare una tesi di laurea, decidere una strategia d'azione, tradurre un testo. L'aumento delle informazioni non si traduce di per sé in aumento della conoscenza.

La produzione di conoscenza da dati e informazioni è legata ai cosiddetti **sistemi informativi**, cioè l'insieme delle attività di gestione delle informazioni, delle relative modalità e degli strumenti tecnologici usati a tale scopo, che oggi possono essere automatizzati (**sistemi informatici**) oppure no. Essendo un complesso di informazioni, personale specializzato e attività e tecniche per gestire tali informazioni, la biblioteca stessa costituisce un esempio di sistema informativo.

I sistemi informativi possono poi fare uso di sistemi informatici per automatizzare alcune procedure di gestione (*sistemi ibridi*) o addirittura tutte (*sistemi digitali*).

Un esempio particolarmente efficace e diffuso di sistema informatico è quello realizzabile attraverso basi di dati (**database**), o banche dati, che presentano determinate caratteristiche generali. Un database (**DB**) è una raccolta di dati omogenei relativi a uno specifico **dominio**, organizzati in modo strutturato secondo precise convenzioni.

Essi gestiscono processi di:

- **amministrazione** (progettazione, creazione ed eliminazione del DB);
- **transazione** (inserimento, aggiornamento, cancellazione di dati);
- **interrogazione** (consultazione di dati).

Le fasi di transazione e interrogazione, in particolare, vengono gestite mediante una famiglia di software definiti **DBMS** (*DataBase Management Systems*), fra i quali alcuni di quelli più noti e utilizzati attualmente sono MS Access e Oracle (proprietary), MySQL e PostgreSQL (*open source*).

La progettazione di un database si basa su **tre fasi successive**, di concretezza sempre maggiore:

- *progettazione concettuale*;
- *progettazione logica*;

- *progettazione fisica*.

La **progettazione concettuale** porta alla creazione del *modello concettuale*, lo schema che rappresenta la realtà del corpus/raccolta di riferimento e che contiene una descrizione dettagliata dei dati disponibili, delle relazioni e dei vincoli tra di essi, senza alcun riferimento ai dettagli implementativi destinati ad essere presi in considerazione solo nelle fasi successive della progettazione.

Il **modello concettuale** è tipicamente basato su uno schema del tipo E-R (entità-relazioni), che descrive:

- **entità**, definibili come una classe di oggetti (astratti o tangibili) della realtà di interesse, distinguibili dagli altri. Ad esempio, in un catalogo di biblioteca, le entità *libro*;
- **attributi**, ovvero le particolari proprietà che caratterizzano ciascuna entità (istanza) nel contesto di interesse. Riprendendo l'esempio del catalogo, attributi dell'entità *libro* sarebbero il *titolo*, il *numero delle pagine* etc.;
- **relazioni**, da intendersi come il rapporto che lega diverse entità tra di loro all'interno dell'universo di riferimento. Riprendendo ancora una volta l'esempio del catalogo, esisterà una relazione dotata di caratteristiche specifiche che colleghi una certa istanza dell'entità *libro* contraddistinta da alcuni attributi specifici con determinati valori (il libro il cui titolo sia *I promessi sposi*) con una specifica istanza di un'entità di tipo diverso, ma sempre identificabile per mezzo di determinati attributi di valore definito (l'autore il cui nome sia Alessandro Manzoni).

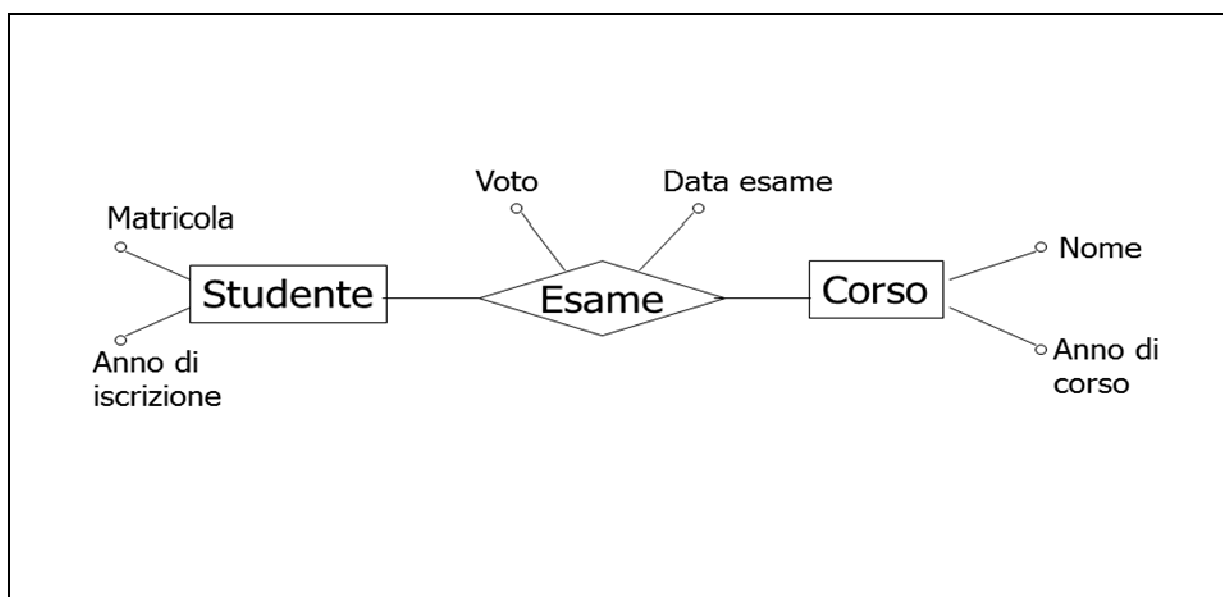


Figura 1: Rappresentazione schematica di entità, attributi e relazioni

Un modello concettuale particolarmente importante nel mondo delle biblioteche è quello definito dai *Functional Requirements for Bibliographic Records (FRBR)*, che si propone come schema di analisi delle entità e delle relazioni intercorrenti in ambito bibliografico e ha segnato un fondamentale punto di svolta nella teoria catalogafica.

## FORMATI DI FILE

Il **formato dei file** è una convenzione utilizzata dagli elaboratori elettronici per leggere, scrivere e interpretare i contenuti di un file; infatti, i file non sono altro che insiemi ordinati di byte, cioè successioni di 0 e 1, che codificano un certo contenuto informativo. Essendo il costituente fondamentale sempre identico, diventa necessario utilizzare una qualche convenzione che legghi i byte a un significato specifico, in modo da poter veicolare contenuti anche profondamente diversi tra loro. Ad esempio, un formato di file per immagini può stabilire che i primi due byte sono l'altezza e la larghezza dell'immagine, mentre i seguenti sono i relativi colori secondo uno schema preordinato. I file di

testo usano vari sistemi di codifica, chiamati **character encoding** (come ASCII o UTF-8), per rappresentare lettere e formattazioni diverse.

Per molti formati sono state rese pubbliche le specifiche che descrivono esattamente come i dati devono essere codificati e che possono essere usate per stabilire se un programma specifico tratti correttamente o meno un determinato formato: tali formati si definiscono **aperti**. Non sempre, però, queste specifiche sono disponibili, anzitutto perché alcuni formati sono considerati segreti industriali e le loro specifiche non vengono distribuite pubblicamente, come avviene, ad esempio, per molti dei formati usati dalla suite Microsoft Office: tali formati si definiscono **proprietary**. In tal modo i dati salvati con quel programma non possono essere letti con altri programmi simili. Risalire ai dati originali salvati in un formato sconosciuto è possibile, attraverso un lavoro di *reverse engineering*, ma di solito si tratta di un processo assai lungo e costoso.

In ogni caso la scelta del tipo di **markup**, il formato finale, determina e veicola le caratteristiche della risorsa digitale prodotta, dunque i suoi utilizzi successivi.

Gli elaboratori possono utilizzare tre diversi metodi per stabilire il formato di un file e decidere, quindi, quale sia lo strumento software necessario per aprirlo e modificarlo:

- a) l'**estensione**: il formato di un determinato file è comunemente indicato attraverso l'estensione, da intendersi come una serie di lettere (in genere tre, per motivi storici) unita al nome del file mediante un punto. Ad esempio, «prova.txt» è un file di testo (o meglio, il suo contenuto va interpretato come testo), mentre «prova.jpg» è un'immagine;
- b) il **magic number**: il formato di un file può anche essere identificato dai primi due o più byte del file, comunemente detti *magic number*. Ad esempio, #! identifica gli script nei sistemi Unix e Unix-like (vedi anche *shabang*), mentre 0xffd8 identifica le immagini in formato jpg;
- c) i **metadati espliciti**: i file system HFS e HFS+ usati sui computer Macintosh affiancano ogni file con informazioni dettagliate riguardanti il suo formato, il programma che l'ha creato etc. Un approccio simile è adoperato con i tipi MIME, usati per identificare il formato dei file trasferiti tramite Internet.

Il tipo **MIME** (*Multipurpose Internet Mail Extensions*) è uno standard proposto dai laboratori Bell Communications nel 1991 per estendere le capacità della posta elettronica (*mail*), cioè permettere di inserire documenti (immagini, suoni, testo etc.) in un messaggio. Da allora il tipo MIME viene utilizzato per individuare sia i documenti allegati a una mail, sia i documenti trasferiti via web con il protocollo HTTP. Pertanto, durante una transazione fra un server web e un navigatore Internet, il server web invia in primo luogo il tipo MIME del file inviato al navigatore, affinché questi possa sapere come visualizzare il documento.

Un tipo MIME è costituito nel seguente modo:

Content-type: type\_mime\_principale/su\_type\_mime

Ad esempio, un'immagine GIF ha il seguente tipo MIME:

Content-type: image/gif

## INFORMATION RETRIEVAL (IR)

L'*Information Retrieval* è l'insieme delle tecniche utilizzate per il recupero mirato dell'informazione elettronica nell'universo informativo (*docuverso*). Si tratta di un ambito di studi interdisciplinare, che coinvolge la psicologia cognitiva, l'architettura informativa, la filosofia (cosiddetta «ontologia»), il design, il comportamento umano sull'informazione, la linguistica, la semiotica, la scienza dell'informazione e l'informatica.

Numerosi e diversificati sono i sistemi di *information retrieval*: dall'indice di un libro a un catalogo cartaceo a schede, fino ai sistemi informatici più evoluti.

L'utente che abbia uno specifico **bisogno informativo** lo veicola attraverso le **query** («interrogazioni»), stringhe di parole-chiave che rappresentano l'informazione richiesta e vengono inserite nel sistema, ad esempio digitandole con la tastiera nella casella di ricerca di un catalogo elettronico. A seguito di tale input il sistema esegue una ricerca nei suoi indici interni (che rappresentano, in formato leggibile dalla macchina, tutta l'informazione disponibile) e fornisce in risposta uno o più **item** («oggetti», «risorse»), che a loro volta rappresentano le entità contenenti le informazioni (ad esempio, i record del catalogo stesso). Se l'utente non è soddisfatto dei risultati ottenuti, perché troppo scarsi («**silenzi**o») oppure troppo abbondanti o contenenti elementi errati e di disturbo («**rumore**»), esegue una nuova *query* con nuove parole-chiave, o raffina la precedente inserendo ulteriori **filtri**.

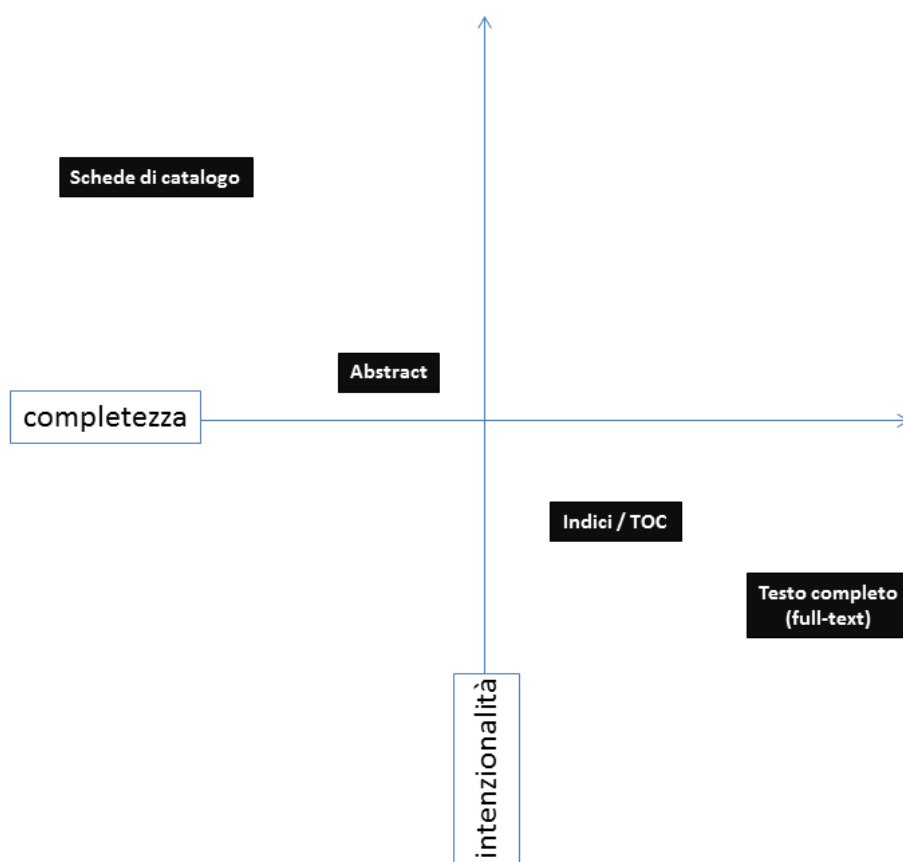


Figura 2: Rapporto intenzionalità/completezza delle risorse informative

L'informazione disponibile può essere rappresentata attraverso differenti item che mostrano diversi livelli di completezza (quantità dell'informazione rappresentata rispetto al contenuto originale), in relazione inversa con quelli di intenzionalità (la volontà di chi ha creato quello specifico item di utilizzarlo come simbolo dell'informazione in un sistema IR).

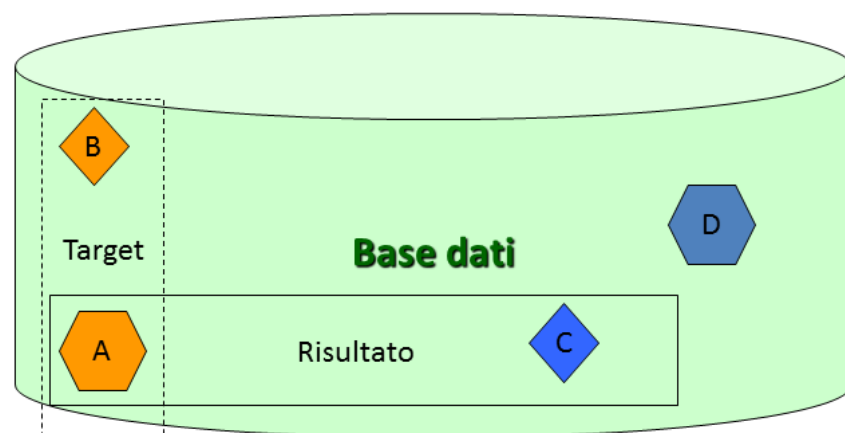
In ogni sistema di *information retrieval* si confrontano due grandezze fondamentali, anch'esse in rapporto di inversa proporzionalità:

1. il **richiamo** (*recall*), indicante la proporzione tra il numero di documenti rilevanti recuperati e il numero di tutti i documenti rilevanti disponibili nella collezione considerata;
2. la **precisione** (*precision*), cioè la proporzione di documenti pertinenti e tutti quelli recuperati.

$$R = \frac{\text{N di documenti rilevanti recuperati}}{\text{N di documenti rilevanti}}$$

$$P = \frac{\text{N di documenti pertinenti recuperati}}{\text{N di documenti recuperati}}$$

In pratica, eseguendo una *query* su un sistema di *information retrieval* (la base dati schematizzata nella figura che segue) si otterrà un risultato che potrà essere, a seconda dei parametri utilizzati, più o meno vicino al target prefissato, ossia al soddisfacimento del bisogno informativo iniziale.



- A = Pertinenti ritrovati
- B = Pertinenti NON ritrovati (silenzio)
- D = NON pertinenti NON ritrovati
- C = NON pertinenti ritrovati (rumore)

Migliore sarà il sistema, migliori saranno i parametri impostati, più facile sarà la possibilità che risultato e target si identifichino, riducendo al minimo silenzio e rumore.

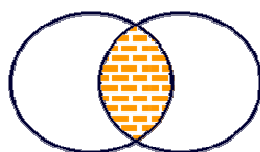
Fra i vari sistemi utili per rappresentare il *docuverso* (o una parte di esso) all'interno di un sistema di *information retrieval*, quello che è stato maggiormente impiegato risulta basato sulla **logica booleana**: questo perché, ancora oggi, il modo più sicuro per la memorizzazione dell'informazione su un supporto è quello di registrare la variazione tra due soli stadi (0/1; acceso/spento; vero/falso; bianco/nero etc.). La logica booleana, a sua volta, fornisce un insieme di regole per la manipolazione e le operazioni su espressioni logiche, cioè espressioni che possono risultare in due soli valori possibili: vero/falso.



A questo punto è facile intravedere il rapporto tra la logica di Boole e l'architettura di funzionamento dell'elaboratore, basata sulla logica binaria: qualunque sequenza di cifre binarie, dunque qualsiasi tipo di informazione rappresentata nel calcolatore, può essere elaborata attraverso le operazioni della logica booleana. L'uso delle sue regole nei sistemi di *information retrieval* si concretizza nella possibilità di far ricorso ai cosiddetti **operatori booleani**, particolari stringhe di caratteri che vengono riconosciute dalla macchina come istruzioni per associare i termini inseriti nella *query*. La ricerca ha inizio con un confronto tra gli elementi posti nella *query* e quelli presenti nel database, dopodiché vengono applicate le operazioni di insieme, secondo la successione prevista dagli operatori.

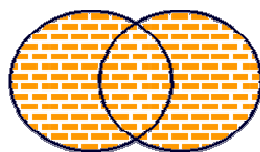
Gli operatori booleani più diffusi sono:

- **AND (intersezione)**: recupera solo i record contenenti entrambi i termini specificati. Più AND si immettono in fase di ricerca, più si restringe il numero dei risultati.



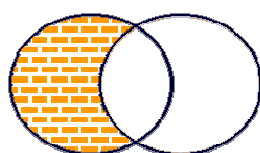
**PROMESSI AND SPOSI**

- **OR (unione)**: recupera i record che presentino almeno uno dei termini forniti. Più OR si immettono in fase di ricerca, più si amplia il numero dei risultati.



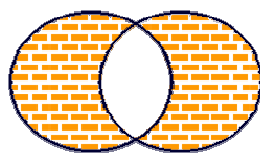
**PROMESSI OR SPOSI**

- **NOT (complemento)**: rintraccia i record che soddisfano un determinato criterio, escludendone altri espressi nella *query*. È utile per eliminare l'ambiguità fra due termini (ad esempio: *lavagna NOT luminosa*).



**PROMESSI NOT SPOSI**

- **XOR (esclusione)**: recupera i record contenenti esclusivamente uno dei criteri immessi. Ad esempio, specificando *promessi XOR sposi* si ottengono tutti i record che contengono la sola parola *promessi* e tutti quelli che contengono la sola parola *sposi*, ma non quelli che le contengono entrambe.



**PROMESSI XOR SPOSI**

(immagini tratte da: F. Metitieri - R. Ridi, *Biblioteche in rete*, Roma-Bari: Laterza, 2003)

Altri **operatori** sono quelli **di prossimità** (NEAR, WITHIN etc.), i quali permettono di eseguire una ricerca specificando a che distanza i termini devono essere tra loro; ad esempio, si può indicare che debbano essere uno accanto all'altro, oppure entro un certo numero di parole, o nell'ordine specificato.

Così ricerca *Titolo = impatto NEAR ambientale* restituirà i record che contengono le parole *impatto ambientale* vicine fra loro, ma non i record che le contengono in punti lontani, come ad esempio *L'impatto dei pesticidi sull'equilibrio ambientale*. Tuttavia, è talvolta possibile richiedere che le due parole siano adiacenti o separate al massimo da un determinato numero di altre parole: ad esempio, la ricerca *Titolo = impatto NEAR3 ambientale* trova record in cui le parole *impatto* e *ambientale* sono separate al massimo da altre tre parole.

Alcuni sistemi permettono la ricerca esatta di frasi chiudendo i termini tra virgolette. Ad esempio, volendo cercare pagine web sull'AIB, se si scrivesse *Associazione italiana biblioteche* si troverebbe molto rumore, poiché i termini presi singolarmente sono piuttosto generici. Inserendoli tra virgolette, invece, si avrebbe una risposta precisa, poiché i termini immessi verrebbero cercati come frase.

Di solito questi linguaggi consentono una serie di trucchi per perfezionare la ricerca. Gli operatori booleani, in particolare, corrispondono alle operazioni matematiche: AND alla moltiplicazione, OR alla somma, NOT alla sottrazione; inoltre, spesso è possibile scrivere direttamente i simboli matematici al posto degli operatori linguistici. Volendo alterare l'ordine naturale delle operazioni (NOT - AND - OR), è possibile utilizzare le parentesi, così come, sempre per raffinare la ricerca, è possibile troncare i termini per rintracciare parole simili.

I **caratteri cosiddetti "jolly"** più usati sono \$, %, \* (ad esempio, immettendo il termine *giornal\** verranno recuperati documenti contenenti giornale, giornali, giornalista, giornalismo, giornaliero etc.). Altri caratteri jolly possono essere inseriti al posto di una lettera all'interno della parola per mascherare un carattere, mentre per ricercare frasi esatte i termini vanno racchiusi tra virgolette (ad esempio, «*nel mezzo del \* di nostra vita*»). Ulteriori funzionalità avanzate di ricerca possono essere offerte dai vari strumenti, come limitazioni per data, lingua, tipologia di documento etc.

### George Boole

George Boole (Lincoln, 2 novembre 1815 - Ballintemple, 8 dicembre 1864) fu un matematico e logico britannico, considerato il fondatore della logica matematica, la cui opera influenzò anche settori della filosofia. A causa dello stato di povertà della sua famiglia fu praticamente un autodidatta e studiò il greco, il latino, il francese, il tedesco, l'italiano, oltre che, naturalmente, la matematica, alla quale si dedicò fin da giovane sui testi di Laplace e Lagrange. Assillato da problemi economici svolse varie attività. Morì all'età di soli 49 anni, per una grave forma febbrile (polmonite) provocata da un banale raffreddore.

Incoraggiato e indirizzato da Duncan Gregory, curatore del «Cambridge Mathematical Journal», Boole si dedicò allo studio di metodi algebrici per la risoluzione di equazioni differenziali e la pubblicazione dei suoi risultati sulla suddetta rivista gli fece ottenere prima una medaglia della Royal Society, poi, nel 1849, la nomina alla cattedra di matematica al Queen's College di Cork, dove avrebbe insegnato per il resto della sua vita.

Con l'opera *The Mathematical Analysis of Logic* (1847), scritta sulla scia della polemica sorta tra Augustus De Morgan e Sir William Hamilton circa la quantificazione del predicato, propose una propria interpretazione del rapporto fra matematica, logica e filosofia che prevedeva l'associazione tra logica e matematica al posto di quella tra logica e metafisica. In particolare, Boole considerava la logica alla stregua della scienza delle leggi dei simboli attraverso i quali si esprimono i pensieri, applicando parte della filosofia algebrica cantabrigense ad un settore inesplorato come quello della logica formale.

Nel 1854 pubblicò la sua opera più importante, indirizzata alle leggi del pensiero, con la quale propose una nuova impostazione della logica. Il suo scopo fu da un lato quello di studiare le leggi delle operazioni

mentali alla base del ragionamento esprimendole nel linguaggio simbolico del calcolo, dall'altro quello di istituire, conseguentemente, una disciplina scientifica della logica sorretta da un metodo, finché, dopo aver rilevato le analogie tra oggetti dell'algebra e oggetti della logica, ricondusse le composizioni degli enunciati a semplici operazioni algebriche. Con questo lavoro fondò la teoria di quelle che attualmente vengono definite **algebre di Boole** (o, semplicemente, *algebra booleana*). Pur mantenendo distinte le operazioni mentali da quelle algebriche, la scienza della logica nella forma algebrica dall'algebra vera e propria in quanto settore della matematica, nonché le leggi logiche dai settori delle scienze naturali, l'obiettivo di Boole fu quello di travestire la logica con un abito matematico algebrico.

Successivamente si dedicò alle equazioni differenziali, argomento a cui riservò, nel 1859, un apposito testo che ebbe molta influenza in materia. Inoltre studiò sia il calcolo delle differenze finite, pubblicando nel 1860 il trattato *Treatise on the Calculus of Finite Differences*, sia problemi generali del calcolo delle probabilità. Fu anche tra i primi ad esaminare proprietà fondamentali dei numeri, come la proprietà distributiva, in quanto proprietà capace di caratterizzare alla base alcune teorie algebriche.

I suoi lavori matematici gli procurarono numerosi riconoscimenti, ma coltivò pure diversi interessi nella letteratura e nella filosofia: Aristotele, Cicerone, Dante e Spinoza furono i suoi autori preferiti.

Primo estimatore e continuatore della sua opera fu Augustus de Morgan, mentre grazie a Claude Shannon, che riconobbe la coincidenza tra il funzionamento dei circuiti commutatori e la logica proposizionale, l'opera maggiore di Boole arrivò a costituire la base per gli studi sui circuiti elettronici e sulla commutazione, consentendo di compiere un passo determinante verso la concezione dei moderni computer.

Nel campo della logica i suoi più grandi meriti sono stati l'applicazione del calcolo simbolico alla logica e il superamento del modello aristotelico e della Scuola, perché insufficienti a sorreggere l'impalcatura della logica stessa.

(Wikipedia)

## METADATI

I **metadati** sono informazioni strutturate che presentano, descrivono e localizzano una risorsa informativa, ne permettono il reperimento, l'uso e la gestione e garantiscono la sua conservazione a lungo termine. Si applicano a qualsiasi tipo di oggetto, digitale o non digitale.

In pratica i metadati sono «dati sui dati» o «informazioni sulle informazioni», finalizzati a rendere possibile:

- la descrizione della risorsa, intesa come documento o anche come set di dati;
- il reperimento della risorsa dalle fonti informative, tramite adeguati strumenti di ricerca (motori, metamotori etc.);
- la gestione delle risorse depositate in un *repository* digitale.

Nella biblioteca tradizionalmente intesa è possibile parlare di descrizione delle risorse o dei documenti, mentre nella biblioteca digitale occorre parlare anzitutto di presentazione delle risorse. I metadati, a loro volta, sono sempre esistiti, anche prima dell'avvento dell'informatica, alla quale il concetto è ora fortemente legato.

Un insieme di metadati è rappresentato, ad esempio, dalla descrizione di una risorsa affidata alle schede del catalogo cartaceo, oppure dalla presentazione affidata ai frontespizi delle pubblicazioni monografiche a stampa. Tuttavia, ai fini della presente esposizione i metadati verranno intesi sempre in senso ristretto, come *«informazione comprensibile alla macchina su risorse web o altri oggetti»*, secondo una celebre definizione di Tim Berners Lee (1987).

Gli **schemi di metadati** sono set di metadati progettati per un obiettivo specifico, come quello di presentare un tipo particolare di risorsa informativa. I più noti e utilizzati sono i seguenti: MARC, nato negli anni Settanta e largamente adoperato per la descrizione delle risorse nei cataloghi di biblioteca; Dublin Core, molto utilizzato per la descrizione di risorse rese disponibili sul web; nonché TEI, METS, MODS, EAD, LOM e altri ancora.

Esistono diverse **modalità di classificazione** degli schemi di metadati, che prendono in considerazione la **collocazione** del metadato (che può essere *interna* alla risorsa, in una modalità particolarmente diffusa tra le risorse disponibili su web, o *esterna*, in appositi *services* contenenti anche un collegamento tra i metadati e la risorsa alla quale si riferiscono) oppure il tipo di **strutturazione**.

In tal senso si possono individuare:

- **metadati generici e non strutturati**, fascia nella quale rientrano metadati di formato proprietario relativamente non strutturati, di solito estratti automaticamente dalle risorse e indicizzati per la ricerca. Il tipico ambiente che ospita questo genere di strutture è il web, all'interno del quale gli attuali motori di ricerca creano formati di metadati semplici attraverso un'indicizzazione su tutto il testo pieno della risorsa. Termini e parole-chiave vengono assegnati in modo assai generico. In questo caso i dati hanno poca semantica esplicita e non permettono la ricerca per campo;
- **metadati in formati generici a struttura semplice**, ovvero metadati strutturati in campi definiti, i quali consentono una ricerca per campi. I dati sono strutturati in modo estremamente semplice, tanto che possono essere creati da utenti non specialisti perché non richiedono conoscenze specifiche della disciplina e generalmente, ma non necessariamente, vengono creati manualmente. Il loro vantaggio consiste nel fatto che i dati contengono una descrizione sufficientemente dettagliata da consentire ad un utente di valutare la potenziale utilità o interesse di una risorsa senza doverla acquisire o connettersi ad essa; per questa ragione sono gli standard preferiti dai servizi di accesso all'informazione. Solitamente tali modelli sono flessibili e consentono aperture verso descrizioni di svariate tipologie di risorse. In quest'ottica è stato creato ed ha riscosso ampio successo il **Dublin Core**, concepito come semplice formato strutturato ad utilizzo degli stessi creatori di risorse (autori o editori), mentre sull'altro versante si collocano i servizi per l'estrazione e raccolta automatica di metadati;
- **metadati ricchi a struttura complessa**, comprendenti appunto formati descrittivi ricchi, ossia strutture più complesse utilizzate per la selezione delle risorse e la presentazione adeguata agli scopi. La struttura di questi metadati comprende anche sottocampi con etichettature elaborate. Uno dei più importanti formati di metadati compresi in questa fascia è il **MARC**, corrispondente al formato più usato per i record bibliografici, essendo fortemente strutturato perché dotato di numerosi campi e differenti attributi. Una delle caratteristiche principali è data dal fatto che la creazione viene delegata a terze figure, vale a dire bibliotecari, documentalisti o figure preposte alla selezione dei documenti che devono essere descritti, solitamente in un database. La scelta delle risorse, dunque, non è casuale ed implica, in ogni caso, una creazione dei metadati quasi esclusivamente manuale. Gli standard utilizzati per questi metadati sono solitamente internazionali. Oltre che dettagliati nella loro presentazione, questi metadati sono sufficientemente espressivi da catturare una varietà di relazioni a diversi livelli. I metadati di questa fascia sono generalmente associati con attività di studio e ricerca, richiedono conoscenze specialistiche per la loro creazione e mantenimento e alimentano requisiti propri di domini specialistici.

Inoltre esistono tentativi di classificare i diversi schemi di metadati a seconda delle loro funzioni, cosicché è possibile distinguere tra:

- **metadati descrittivi**, ossia quelli che descrivono o catalogano un oggetto. Si tratta di informazioni che indicano il contesto, la gestione, i processi, la conservazione e l'uso della risorsa descritta. Appartengono a questa categoria, per esempio, registrazioni catalografiche o annotazioni (tag e commenti) di utenti;
- **metadati amministrativi e gestionali**, usati per scopi amministrativi o per la gestione delle risorse al fine di registrare informazioni sulla loro acquisizione e localizzazione, o sull'acquisizione dei diritti di proprietà intellettuale. Si tratta, ad esempio, di documentazione sull'hardware e il software, informazioni sulla digitalizzazione (formati etc.), rapporti di compressione, dati di autenticazione e sicurezza (chiavi crittografiche, password), o ancora documentazione della condizione fisica delle risorse e delle azioni intraprese per conservare le versioni fisiche e digitali delle risorse stesse, come accade per il ripristino (*refreshing*) e la migrazione dei dati.

### **Dublin Core**

Il *Dublin Core* (dal nome della città americana dell'Ohio) è un sistema di metadati costituito da un nucleo di elementi essenziali ai fini della descrizione di qualsiasi materiale digitale accessibile via rete informatica.

Il progetto del *Dublin Core* (nome completo: **Dublin Core Metadata Initiative - DCMI**) si sviluppò in ambito OCLC (*On line Computer Library Center*), la grande rete di servizi americana per le biblioteche. Nel marzo 1995 si tenne una conferenza nella città americana di Dublin, in Ohio, nel corso della quale i partecipanti (bibliotecari, archivisti, editori, ricercatori e sviluppatori di software, oltre ad alcuni membri dei gruppi di lavoro dell'IETF, l'*Internet Engineering Task Force*) convennero sulla necessità di creare un insieme di strumenti condivisi per l'accesso alle risorse digitali. Lo scopo era quello di stabilire un insieme base di elementi descrittivi che potessero essere forniti dall'autore o dall'editore dell'oggetto digitale, ed inclusi in esso, o da esso referenziati. Il consorzio di utenti così costituito cominciò poi a sviluppare un'architettura per i metadati che venisse incontro alle necessità dei venditori e dei produttori di informazioni.

#### • Elementi costitutivi

Il nucleo originario, proposto nel dicembre 1996, era costituito da quindici elementi di base, per poi estendersi anche a sottoelementi o qualificatori, pur mantenendo, nonostante gli sviluppi, una struttura stabile.

La traduzione italiana di riferimento della versione 1.1 del *Dublin Core Metadata Element Set* è curata dall'ICCU, l'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche.

#### • Titolo (*Title*)

Nome dato alla risorsa. In particolare, un Titolo sarà un termine con il quale la risorsa è formalmente conosciuta.

#### • Autore (*Creator*)

Entità che ha la responsabilità principale della produzione del contenuto della risorsa. Esempi di Autore possono essere una persona, un'organizzazione o un servizio responsabili del contenuto intellettuale della risorsa.

#### • Soggetto (*Subject*)

Argomento principale della risorsa. In particolare un Soggetto può essere espresso da parole o frasi chiave, oppure da codici di classificazione che descrivono l'argomento della risorsa. Solitamente questi termini vengono scelti tra i valori di un vocabolario controllato o di uno schema di classificazione formale.

#### • Descrizione (*Description*)

Spiegazione del contenuto della risorsa. Testo descrittivo libero che può includere un riassunto analitico, un indice o una rappresentazione grafica del contenuto.

#### • Editore (*Publisher*)

Entità responsabile della pubblicazione della risorsa. Esempi di Editore possono essere una persona, un'organizzazione o un servizio che si occupi di rendere disponibile la risorsa nella sua forma attuale.

• **Autore di contributo subordinato (*Contributor*)**

Entità responsabile della produzione di un contributo al contenuto della risorsa. Esempi di Autore secondario includono una persona, un'organizzazione o un servizio che contribuiscono alla produzione della risorsa.

• **Data (*Date*)**

Data associata ad un evento del ciclo di vita della risorsa. Normalmente la data è associata al momento di creazione o di disponibilità della risorsa e viene indicata attraverso una stringa di 8 caratteri nella forma YYYY-MM-DD, come definita nel profilo dello standard ISO 8601. In questo schema l'elemento data 1994-11-05 corrisponde al 5 novembre 1994.

• **Tipo (*Type*)**

Natura o genere del contenuto della risorsa. L'elemento Tipo include termini che descrivono categorie generali, funzioni, generi o livelli di aggregazione per contenuto presi generalmente da un vocabolario controllato.

• **Formato (*Format*)**

Manifestazione fisica o digitale della risorsa. Normalmente l'elemento Formato può includere il tipo di supporto o le dimensioni (grandezza e durata) della risorsa. *Format* può essere usato per determinare il software o l'hardware necessari alla visualizzazione o all'elaborazione della risorsa.

• **Identificatore (*Identifier*)**

Riferimento univoco alla risorsa nell'ambito di un determinato contesto. Solitamente le risorse vengono identificate per mezzo di una sequenza di caratteri alfabetici o numerici secondo un sistema di identificazione formalmente definito. Esempi di tali sistemi di identificazione includono l'URI - *Uniform Resource Identifier* (incluso l'URL - *Uniform Resource Locator*), il DOI (*Digital Object Identifier*) e l'ISBN (*International Standard Book Number*).

• **Fonte (*Source*)**

Riferimento a una risorsa dalla quale è derivata la risorsa in oggetto. La risorsa in questione potrebbe derivare, in tutto o in parte, da un'altra risorsa fonte.

• **Lingua (*Language*)**

Lingua del contenuto intellettuale della risorsa. Per i valori dell'elemento Lingua si utilizza un codice di linguaggio, seguito opzionalmente da un codice di paese, entrambi su due caratteri. Ad esempio: "it" per l'italiano o "en-uk" per l'inglese usato nel Regno Unito.

• **Relazione (*Relation*)**

Riferimento ad una risorsa correlata.

• **Copertura (*Coverage*)**

Estensione o scopo del contenuto della risorsa. Normalmente l'elemento Copertura include la localizzazione spaziale (il nome o le coordinate geografiche di un luogo), il periodo temporale (l'indicazione di un periodo, una data o una serie di date) o una giurisdizione (ad esempio il nome di un'entità amministrativa).

• **Gestione dei diritti (*Rights*)**

Informazione sui diritti esercitati sulla risorsa. Normalmente l'elemento Diritti contiene un'indicazione sulla gestione dei diritti sulla risorsa, o un riferimento al servizio che fornisce questa informazione. Questo campo comprende gli IPR (*Intellectual Property Rights*), il copyright e vari diritti di proprietà. Se l'elemento *Rights* è assente non si può fare alcuna ipotesi sui diritti della risorsa.

(Wikipedia)

## LINGUAGGI DI MARCATURA E STANDARD BIBLIOGRAFICI

La **teoria dell'informazione di Shannon** (1949) si basa sul principio che un segnale (messaggio) passa da un emittente (mittente), attraverso un trasmettitore, a un destinatario (ricevente), attraverso un recettore, lungo un canale fisico (supporto materiale). Il messaggio (composto di segni) deve essere codificato (costruito e combinato secondo certe regole, ovvero secondo un codice) da chi lo emette e decodificato da chi lo riceve. Il primo livello di comunicazione tra l'uomo è l'elaboratore deve dunque prevedere che esista un codice, noto ad entrambi, per lo scambio delle informazioni.

La macchina è in grado di interpretare solamente i due segni (0 e 1) del **codice binario**, ripetuti un numero indefinito di volte, cosa che tuttavia permette la rappresentazione di un numero virtualmente infinito di fenomeni. Qualunque oggetto digitale (sia esso un testo, un'immagine, un brano audio o video, un programma) è solo una sequenza di cifre 0 e 1. Si parla in questo caso di **bit** (contrazione di *Binary digIT*, «cifra binaria») come dell'unità di misura dell'informazione digitale, calcolata come quantità di 0/1 necessari a realizzare un oggetto digitale. Ogni segno (0 o 1) è un bit, di cui è stato definito un primo multiplo convenzionale, il **byte**, formato da otto cifre.

Per **codifica digitale** si intende dunque il processo indispensabile per tradurre in forma comprensibile all'elaboratore qualsiasi dato, evidentemente reso in codice binario. Per i caratteri alfanumerici esistono apposite tabelle di equivalenza che trasformano numeri e lettere in byte, mentre per codificare un'immagine è necessario definire e rappresentare in formato binario lo stato di ciascun punto (pixel) che la compone.

Questo primo livello di comunicazione permette, ad esempio, di digitalizzare un testo inserendo i caratteri attraverso la pressione dei diversi tasti della tastiera (periferica di input) e la loro visualizzazione sullo schermo dell'elaboratore, oppure di produrre una stampa tramite la stampante (periferica di output). Tuttavia, se ci si fermasse a questo primo livello, non ci sarebbe ancora la possibilità di realizzare una completa codifica dell'informazione contenuta in un documento.

Di fronte a qualsiasi informazione di natura testuale, infatti, occorre distinguere due componenti complementari: una, che possiamo definire **testo**, è strutturata in stringhe di caratteri e costituisce la rappresentazione astratta del contenuto, comprendente le segmentazioni logiche e le partizioni interne dei blocchi di scrittura; l'altra, il **documento**, è la realizzazione del testo in un manufatto, compresa l'organizzazione spaziale del supporto, la scrittura e il supporto stessi, nonché le eventuali immagini. Per questo motivo la codifica di un testo richiede indicazioni ulteriori rispetto alla sola codifica di caratteri, essendo necessarie altre informazioni che non necessariamente vengono visualizzate tramite caratteri visibili (come la strutturazione in capitoli, la presenza di titoli, note etc.).

È quindi utile distinguere una codifica dei dati elementari, che farà riferimento al testo, e una di più alto livello, definibile come **marcatatura** (*markup*) e consistente nell'aggiungere alle sequenze di caratteri che rappresentano il testo digitali ulteriori sequenze di caratteri, denominate *marcatori*, che consentiranno di descrivere determinati aspetti del documento, come quelli legati alla sua struttura logica o fisica. Siamo abituati a farlo, ad esempio, mediante i software della famiglia dei Word Processors, come MS Word, che consentono di aggiungere al testo informazioni di formattazione (grassetto, corsivo, paragrafi etc.) per mezzo di un'interfaccia grafica.

Esistono diversi **linguaggi di marcatatura**, inquadrabili sotto molteplici aspetti:

- **proprietario o aperto**: il markup può essere definito da un'azienda, che lega la possibilità di produrlo all'utilizzo di un determinato applicativo, oppure, al contrario, può consistere in una serie di istruzioni standardizzate e condivise in maniera che il codice sia liberamente disponibile;
- **leggibile o non leggibile**: il markup può essere realizzato in maniera tale da risultare leggibile senza che sia necessario ricorrere a particolari software per vederlo o interpretarlo; altrimenti può essere realizzato per mezzo di un applicativo che incorpora e nasconde i marcatori, in modo che il documento risultante possa essere correttamente interpretato solo dall'applicativo col quale è stato prodotto;
- **orientato al layout (presentazionale) o alla struttura (descrittivo)**: in tal caso ci si riferisce al fatto che i marcatori vengano utilizzati per definire e rappresentare le caratteristiche fisiche del documento, oppure per segnalarne la struttura logica, cioè il ruolo rivestito da ciascun blocco di testo (ad esempio, titolo, paragrafo, nota etc.);
- **procedurale o dichiarativo**: un markup può contenere una serie di istruzioni riguardanti le caratteristiche che dovrà avere l'output del processo, oppure contenere indicazioni sulle caratteristiche formali o di presentazione del documento stesso.

Da tutti questi elementi deriva, come vedremo in seguito, l'esistenza di diversi formati di dati, poiché ogni singolo sistema di marcatura consente la creazione di documenti con un certo formato, cosicché ciascun applicativo lega i documenti che produce al formato di riferimento: in pratica, ogni applicativo software dispone di una serie di possibili formati con cui creare e archiviare documenti. L'unico formato non marcato è il cosiddetto **plain text**, che veicola esclusivamente testo, senza alcuna informazione di contesto, né di tipo grafico (formattazione), né di tipo logico (struttura).

Nelle applicazioni di ambito bibliotecario, in particolare, assumono rilevanza specifica i linguaggi di markup dichiarativo e descrittivo, costituiti da un insieme di marcatori (spesso definiti **tag**) che, inseriti nel testo, contrassegnano i blocchi ai quali si riferiscono per la loro funzione logico-strutturale. Non essendo legati ad alcun applicativo e ad alcun genere di documento in particolare, tali linguaggi si definiscono anche **generic markup languages**, ai quali appartengono linguaggi molto noti, come SGML (il capostipite), HTML (utilizzato per la creazione delle pagine web) e XML (un sottoinsieme semplificato di SGML, concepito proprio per la codifica di testi elettronici di vario genere).

Rispetto alla classificazione prima descritta, è possibile definire questo gruppo di linguaggi come orientati alla descrizione della struttura (**descrittivi**), **non proprietari** e di **contenuto leggibile** dall'utente; sono inoltre indipendenti anche dal sistema operativo usato dall'elaboratore, perché generalmente si basano su istruzioni in formato *plain text*.

## EXTENSIBLE MARKUP LANGUAGE (XML)

Lo **Standard Generalized Markup Language (SGML)** è un metalinguaggio definito come standard ISO (ISO 8879:1986 SGML), nato con lo scopo di definire linguaggi da utilizzare per la stesura di testi destinati ad essere trasmessi e archiviati con strumenti informatici, ossia per la stesura di documenti in forma leggibile da computer (*machine readable form*).

Tuttavia, nonostante le grandi potenzialità, la sua enorme complessità lo rendeva poco utilizzato, cosicché nel 1998, dopo due anni di lavoro, il W3C (il consorzio sorto con l'obiettivo di migliorare i già esistenti protocolli e linguaggi per il World Wide Web e di aiutare il web stesso a sviluppare tutte le proprie potenzialità) pubblicò le specifiche di un nuovo linguaggio basato su SGML, ma semplificato: **XML**, acronimo di **eXtensible Markup Language**, così definito perché in grado di utilizzare marcatori per descrivere le caratteristiche logiche del documento e le eventuali relazioni semantiche tra le informazioni in esso contenute.

Attraverso XML è possibile creare documenti che si prestano ad essere condivisi e riutilizzati in diversi ambiti e da diversi programmi, un sottoinsieme di SGML semplificato e ottimizzato specificamente per applicazioni operanti nel World Wide Web. Il linguaggio è definito «estensibile» in quanto, più che essere un linguaggio, è un vero e proprio



metalinguaggio, utilizzabile per creare linguaggi di marcatura specifici. In pratica, a seconda del tipo di documento di interesse e delle necessità descrittive, è possibile definire una serie di marcatori specifici, cioè un'applicazione XML.

Le caratteristiche dei marcatori e le regole che ne contraddistinguono le relazioni sono codificate in una sorta di grammatica del linguaggio creato, definita **DTD (Document Type Definition)**. A fronte di ogni nuova esigenza di descrizione è possibile estendere la DTD, introducendo nuovi marcatori e specificandone le relazioni. Esistono diverse DTD già pronte per l'uso, ma è possibile anche crearne altre completamente nuove.

Come ogni linguaggio, XML ha una propria ortografia e una propria sintassi che vanno assolutamente rispettate. Anzitutto i tag, per essere identificati nel testo che marcano, sono compresi tra due parentesi uncinate (<tag>) e, secondo le specifiche della DTD utilizzata, possono contenere altri tag, contenere dell'informazione o essere vuoti.

Il contenuto di un tag è l'intero testo compreso tra i delimitatori di apertura e chiusura del tag stesso, rappresentati dal tag come sopra definito per l'apertura e dal tag con lo stesso nome, ma con una barra anteposta (ad esempio: <tag>testo marcato</tag>).

A parte il rispetto di poche regole formali, che permettono di definire un qualsiasi documento XML come ben formato e valido, è dunque possibile creare DTD di qualsiasi genere, per definire documenti di qualsiasi tipo.

Es.:

```
<?xml version="1.0" ?>
<libro>
  <titolo>Titolo del libro</titolo>
  <capitolo>
    <numerazione>Primo capitolo</numerazione>
    <titolo>Titolo del primo capitolo</titolo>
    <testo>
      Testo del primo capitolo
    </testo>
  </capitolo>
  <capitolo>
    <numerazione>Secondo capitolo</numerazione>
    <titolo>Titolo del secondo capitolo</titolo>
    <testo>
      Testo del secondo capitolo
    </testo>
  </capitolo>
</libro>
```

## MACHINE-READABLE CATALOGING (MARC)

Impiegato come linguaggio di marcatura, ma anche come schema di metadati, MARC è una famiglia di standard descrittivi per le risorse bibliografiche nati negli anni Sessanta e ancora largamente utilizzati dai sistemi integrati di gestione di biblioteca (ILS).

La necessità alla quale andava incontro il MARC era quella di trovare un mezzo di scambio delle informazioni bibliografiche che fosse contemporaneamente molto dettagliato nel descrivere i dati, assai sintetico dal punto di vista dello spazio occupato sul supporto di memorizzazione, comprensibile anche al di là delle differenze linguistiche.

Il formato obbliga a individuare i diversi elementi dell'informazione bibliografica attraverso etichette (o tag) costituite semplicemente da tre cifre. Per garantire che nelle fasi di registrazione e successiva lettura dell'informazione non si confonda, ad esempio, il titolo di una pubblicazione con il nome di un editore, fu necessario stabilire una corrispondenza fra etichette e campi della scheda bibliografica (stabilendo, ad esempio, che l'etichetta 100 corrispondesse all'autore principale e la 240 al titolo proprio).

Il formato consentiva di suddividere i campi in unità più piccole, dette *sottocampi*, individuate da una singola lettera o cifra. Così, ad esempio, l'autore principale era rappresentato propriamente dal sottocampo "a" del campo 100, oppure, come spesso si scrive tuttora, dal campo 100\$a o semplicemente 100a.

In sostanza, la specifica MARC definiva inizialmente sia la sintassi che la semantica della rappresentazione dell'informazione bibliografica. A lungo andare, però, la confusione tra questi due obiettivi ha avuto effetti negativi, cosicché si è arrivati a separare la sintassi dalla semantica e il formato è stato generalizzato e rigorosamente definito nello standard **ISO 2709**, mentre numerose diverse semantiche, grosso modo su base nazionale, sono state definite a partire da quel formato neutrale. Ad esempio, negli Stati Uniti è stato adottato USMARC, poi confluito nella specifica MARC21; in Europa sono nati FINMARC, DENMARC, UKMARC e numerosi altri, come ANNAMARC per l'Italia.

Un tentativo di mettere ordine in questa selva di specifiche più o meno incompatibili fra loro è stato avviato diversi anni fa dall'IFLA, la Federazione internazionale delle associazioni bibliotecarie, con la definizione della specifica **UNIMARC**. Si tratta di una semantica condizionata ancora da ISO 2709, ma definita con l'obiettivo di essere universale, cioè adatta alle esigenze dei diversi paesi e delle diverse comunità di utenti e professionisti. È anche molto più rigorosa di quelle nate in precedenza, perché sviluppata a tavolino da commissioni di studiosi di tutto il mondo, soggetta a revisioni periodiche e a continue evoluzioni.

Secondo lo standard ISO 2709, un **record** è costituito da **tre parti contigue**:

1. una **label** di ventiquattro caratteri (spesso chiamata anche **leader o guida**);
2. una **directory** formata da un numero variabile di voci (*entry*) e terminata dal carattere FT (*Field Terminator*, il carattere IS2 della tabella ISO 646);
3. un'area dati, costituita da una serie di campi separati fra loro dal carattere FT.

Ogni record termina con il carattere RT (*Record Terminator*, il carattere IS3 della tabella ISO 646).

Schematicamente, si ha dunque la seguente struttura:

LABEL	DIRECTORY	FT	CAMPO	FT	CAMPO	FT	...	CAMPO	FT	RT
-------	-----------	----	-------	----	-------	----	-----	-------	----	----

I campi contengono solo i dati veri e propri. Le informazioni necessarie a distinguere i campi fra loro e ad accedere a uno qualsiasi di essi sono contenute nella *directory*, nella quale si trovano appunto i nomi dei campi (*etichette*), sempre costituiti da tre caratteri con eventuali zeri iniziali di riempimento.

La struttura dei campi e delle voci della *directory* non è completamente definita dallo standard, in quanto alcune caratteristiche sono determinate dal contenuto della *label* e quindi dipendono dalle diverse implementazioni.

Come già anticipato, la **label**, contenente informazioni sul tipo di record, è costituita da **24 caratteri**, numerati a partire da 0 e così organizzati:

- **0-4**: lunghezza totale del record, comprensiva del terminatore di record. Essendo questo un numero di cinque cifre, non è possibile creare record più lunghi di 99.999 caratteri;
- **5**: stato del record, non definito dallo standard;

- **6-9:** codici applicativi, non definiti dallo standard;
- **10:** numero di indicatori;
- **11:** lunghezza dell'identificatore;
- **12-16:** inizio dell'area dati, a partire dall'inizio del record;
- **17-19:** altri codici applicativi;
- **20-23:** struttura delle voci della *directory*. In dettaglio: il carattere 20 indica quanti caratteri sono usati in una voce per rappresentare la lunghezza del relativo campo; il carattere 21 quanti caratteri sono usati in una voce per rappresentare la posizione del campo a partire dall'inizio dell'area dati; il carattere 22 quanti caratteri sono usati in una voce per rappresentare codici applicativi legati al relativo campo; il carattere 23 è riservato ad eventuali usi futuri.

A titolo di **esempio**, una *label* in UNIMARC può assumere il valore **00830Nam0#2200181###450#**, il cui significato è specificato nella tabella che segue.

Dati presenti nella <i>label</i>	Spiegazione
00830	lunghezza del record (830 caratteri)
N	nuova registrazione
am0	pubblicazione a stampa (a), monografica (m), senza legami con altre pubblicazioni (0)
#	
2	lunghezza degli indicatori
2	lunghezza dei codici di sottocampo
00181	posizione di partenza dei dati
#	livello di codifica: catalogazione con libro alla mano
#	conforme a ISBD
#	non definito
450	dati fissi
#	non definito

#### Esempio di record UNIMARC:

```
06269nam0M2201081      I450
0010020000000050009000201000041000291010008000701020007000782000062000852100044001472150018001912250062002093
0000280027141000930029970000460039270100480043880100230048689900780050989900490058789900670063689900620070389
9005900765899005500824899007000879899007600949899006201025899005801087899004901145899004701194899006201241899
0063013038990115013668990052014818990070015338990062016038990066016658990101017318990070018328990078019028990
0790198089900660205989900600212589900480218589900550223389900670228889900530235589900460240889900580245489900
7402512899004902586899005302635899006802688899004802756899006402804899005002868899007202918899005102990899007
6030418990069031178990058031868990062032448990063033068990077033698990045034468990060034918990050035518990067
0360189900650366889900970373389900540383089900850388489900710396989900490404089900780408989900550416789900520
4222899005404274899008904328899005604417899005804473899009704531899004904628899004804677899004904725899006004
774899005304834899005904887899006004946899007205006899005605078899005305134-IT\ICCU\ANA\0146351-20091020- -
a20040911d2004      |||||itac50      ba-| aita- aIT-1 aEnigma in luogo di marefCarlo Fruttero, Franco
Lucentini- aRomaLa Biblioteca di Repubblicaad2004- a378 p.d21 cm-0 a-HLa -IBiblioteca di Repubblica. Le
strade del giallov17- aSuppl. a: La Repubblica- 11001IT\ICCU\UBO\244782212001 a-HLa -IBiblioteca di
Repubblica. Le strade del giallov17- 1aFruttero, Carlo3IT\ICCU\CFIV\0073734070- 1aLucentini, Franco-
3IT\ICCU\CFIV\0073754070- 0aITbICCUc20091025-
```

Le specifiche dello standard UNIMARC sono disponibili in una versione ridotta messa a disposizione dall'IFLA all'indirizzo Internet <http://archive.ifla.org/VI/8/unimarc-concise-bibliographic-format-2008.pdf>, con disponibilità anche in lingua italiana all'indirizzo <http://unimarc-it.wikidot.com/>.

**Z39.50**, la cui prima versione risale al 1970, è un protocollo utilizzato in ambito bibliotecario per trasferire attraverso la rete informazioni disponibili su database.

Questo protocollo, basato su un'architettura *client/server*, specifica le strutture dei dati e le regole di scambio che permettono ad un sistema *client* (chiamato **origin**) di effettuare ricerche nelle basi dati di un sistema *server* (chiamato **target**) e di poter consultare i record risultanti.

La caratteristica più rilevante del protocollo Z39.50 è quella di offrire, a sistemi diversi per hardware, software e contenuto informativo, la possibilità di scambiarsi dati semanticamente riconoscibili anche al di fuori del contesto strutturale e linguistico in cui sono stati creati; il protocollo, infatti, standardizza la struttura e la semantica delle interrogazioni e dei record.

Tale interoperabilità semantica, unita alla capacità di offrire ricerche effettuate contemporaneamente su una molteplicità di basi di dati, si traduce per l'utente nella possibilità di:

- eseguire ricerche su basi di dati diverse attraverso un'unica interfaccia (quella del suo *client*);
- impostare la richiesta con il grado di precisione voluto (dei record ricercati, infatti, si possono definire sia il contenuto che la forma);
- negoziare sia la struttura dei record che si vogliono ricevere in risposta, sia la forma di rappresentazione;
- fondere e ordinare record provenienti da fonti diverse e presentarli o scaricarli nella forma desiderata.

Una richiesta di ricerca viene inviata a una base dati; i record trovati come risultato della richiesta verranno visualizzati sotto forma di un elenco (denominato **resultset**), da cui si potrà poi accedere ai record veri e propri. Il **resultset** può anche essere oggetto di un'ulteriore ricerca, che produrrà un altro **resultset**.

Gli attributi definiscono come il *server* e il *client* devono colloquiare, quali campi del record possono essere utilizzati come canali di ricerca e quali criteri devono essere specificati dall'utente per svolgere un'interrogazione. Il set di attributi usati dalle biblioteche è denominato **bib-1**.

Pur essendo ancora molto utilizzato, il protocollo Z39.50 è ormai piuttosto datato, poiché risalente a un'epoca di molto precedente alla nascita del WWW. Da qualche anno a questa parte, quindi, la *Library of Congress*, che mantiene le specifiche del protocollo, sta lavorando a una nuova versione, denominata **SRU** (*Search/Retrieval via URL*, <http://www.loc.gov/z3950/>). Le differenze principali di SRU rispetto a Z39.50 consistono soprattutto nella scelta del protocollo HTTP per l'invio delle *query* e nella codifica della risposta del *server* in formato XML.

## STRUMENTI

In ambito bibliotecario l'utilizzo delle tecnologie per automatizzare i processi interni e per supportare l'accesso degli utenti alle informazioni ha prodotto lo sviluppo di diverse applicazioni, alcune rivolte alla gestione e al contenuto, altre alla ricerca e alla disseminazione dell'informazione.

### SISTEMI INTEGRATI DI AUTOMAZIONE PER BIBLIOTECHE

Un sistema integrato di automazione per biblioteche (più noto con la sigla **ILS**, *Integrated Library System*, o anche **LMS**, *Library Management System*, o **SIGB**, *Système intégré de gestion de bibliothèque*) è un software in grado di gestire in maniera integrata tutti i processi che hanno luogo in biblioteca, registrando i dati necessari un'unica volta per poi riutilizzarli a seconda delle necessità.

Tipicamente, si compone dei seguenti **moduli**:

- **acquisizioni**;
- **catalogazione**;
- **circolazione**;
- **periodici**;
- **OPAC**.

La sua principale caratteristica, quindi, è la multifunzionalità, ottenuta grazie all'integrazione dei moduli che gestiscono questi processi in un unico sistema, che utilizza i dati in maniera correlata basandosi su un **database relazionale** (i più diffusi sono Oracle, proprietario, e MySQL, *open source*).

#### Database relazionali

Detti anche **RDBMS** (*Relational DataBase Management System*), si basano sul "modello relazionale" ideato nel 1970 da Edgar F. Codd della IBM.

Invece di utilizzare «righe» (*record*) collegate tra loro attraverso strutture "ad albero", Codd propose di dividere i dati in diverse tabelle. Ad esempio, per registrare informazioni sugli utenti (nome, informazioni di accesso, indirizzo e numeri di telefono) in un database navigazionale tutti questi dati sarebbero stati memorizzati in un unico record, mentre gli elementi non presenti (ad esempio, un utente di cui non sia noto l'indirizzo) sarebbero stati semplicemente omessi. In un database relazionale le informazioni vengono divise, ad esempio, nelle tabelle «utente», «indirizzi», «numeri di telefono» etc. e solo se i dati sono presenti viene creato, nella rispettiva tabella, un record.

Uno dei fattori più interessanti introdotti nei database relazionali sta nel collegamento delle tabelle: nel modello relazionale, per ogni record viene definita una «chiave», ovvero un identificatore univoco. Nella ricostruzione delle relazioni l'elemento di riferimento che distingue una riga dall'altra è costituito proprio da questa «chiave», richiamata nella definizione della relazione. La «chiave» può essere uno dei dati stessi che vengono memorizzati (ad esempio, per la tabella utenti, il codice fiscale della persona), o un campo che viene aggiunto specificatamente per questo scopo (spesso denominato **OID**, *Object Identifier*), o ancora una combinazione di più campi (chiave composta).

Questa operazione di "riunificazione" dei dati non è prevista nei linguaggi di programmazione tradizionali: mentre

l'approccio navigazionale richiede semplicemente di "ciclare" per raccogliere i diversi record, l'approccio relazionale richiede al programma di "ciclare" per raccogliere le informazioni riguardanti ciascun record. Codd propose come soluzione la creazione di un linguaggio dedicato a questo problema; linguaggio che poi si sviluppò nella codifica che oggi è universalmente adottata, costituendo il mattone fondamentale delle basi di dati: **SQL (Structured Query Language)**.

(Wikipedia)

I dati sono accessibili tramite il software, che li utilizzerà, a seconda delle procedure, come schematizzato sotto:

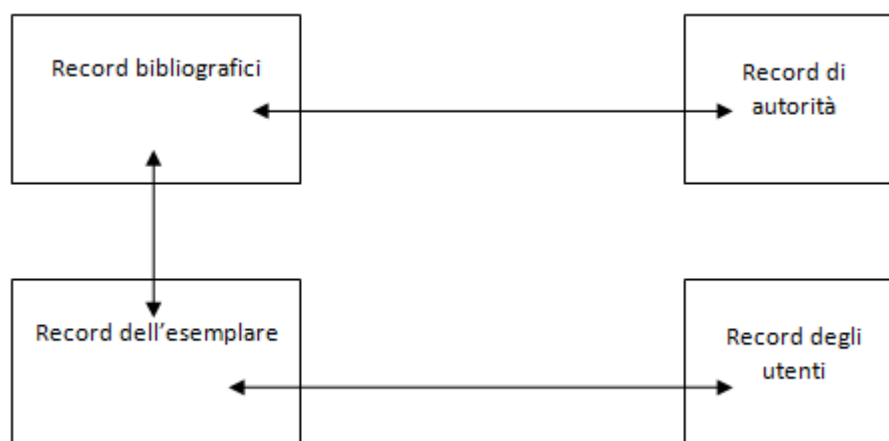


Figura 3: Schema di funzionamento di un ILS

Nel livello più alto vengono conservati i record bibliografici e di autorità. Solo un singolo record bibliografico viene immesso, anche se esistono diverse copie del titolo. Le informazioni sull'esemplare, contenenti dati sulla copia fisica posseduta (collocazione, disponibilità etc.), sono collegate al record bibliografico, con la precisazione che più record dell'esemplare possono essere collegati a un singolo record bibliografico. Quest'ultimo, a sua volta, è collegato anche ad uno o più record di autorità, frutto di pratiche di catalogazione volte al controllo delle forme accettate e uniformi di particolari dati del record bibliografico (normalmente: l'autore, il soggetto, il titolo uniforme). Infine l'archivio dei dati sugli utenti interagisce con i record dell'esemplare per la circolazione dei documenti.

I record, strutturati per campi di lunghezza variabile, sono normalmente esportabili in formato MARC (quello più diffuso in Europa è **UNIMARC**). Ciò consente alla macchina di riconoscere i dati in maniera automatica e univoca per la loro visualizzazione, nonché per il riutilizzo e lo scambio. La maggior parte degli ILS è compatibile con lo standard ISO 2709, che stabilisce le modalità con cui i dati di tipo bibliografico vengono immagazzinati e scambiati tra diversi sistemi di automazione, per garantirne l'interoperabilità; di recente alcuni hanno iniziato ad implementare il linguaggio XML per la codifica e lo scambio dei dati.

Oltre che del sistema di gestione del database, gli ILS sono dotati anche di un modulo per la ricerca, utilizzato sia a livello gestionale, per recuperare i record nell'ambito delle varie procedure, sia a livello utente, per consentire il recupero dell'informazione bibliografica necessaria (tramite l'OPAC, di cui si parlerà più avanti). La ricerca si basa sull'intermediazione dell'*inverted file*, un file di indice contenente l'elenco dei termini indicizzati a partire dai record bibliografici e la lista dei legami associati a ciascun termine. Collegato ad esso è il file *stopword*, che a sua volta contiene i termini da escludere durante la generazione dell'*inverted file* (ad esempio, articoli e preposizioni).

Attualmente la maggior parte degli **ILS**, per quanto concerne l'**architettura del sistema**, si basa su **due modelli principali**:

1. **client/server**, modello che prevede una "scissione" dell'applicazione informatica in due parti, di cui una risiede in un computer remoto (*server*) e normalmente archivia e gestisce i dati, l'altra risiede nel computer locale (*client*).

Quando il programma viene lanciato, per funzionare necessita dell'altra metà, sicché la va a cercare e interagisce con essa utilizzando un protocollo di rete comune. Questa modalità ha ovviamente consentito una migliore organizzazione del lavoro soprattutto nei sistemi bibliotecari: con un database condiviso (dati bibliografici per il catalogo, anagrafica utenti etc.), infatti, le operazioni sono risultate via via più semplici, rapide ed efficaci;

2. **web based**, modello corrispondente a un'evoluzione dell'architettura *client/server*. In tal caso l'applicazione è accessibile praticamente attraverso la Rete, utilizzando come *client* un qualsiasi browser web (Explorer, Firefox, Chrome etc.), senza che ci siano altre applicazioni installate nel PC dell'operatore. Questa possibilità consente ai produttori di aggiornare e far evolvere l'applicativo a costi ridotti, nonché di gestire meglio l'assistenza.

The screenshot displays the SBN Web interface for a specific bibliographic record. The header includes the title 'Polo BVE (Esercizio) - SBN Web' and navigation links like 'Guida | Logout'. A left sidebar contains a menu for 'Interrogazione' with options like 'Titolo', 'Autore', 'Marca', etc. The main area shows a form with various fields: 'Natura' (M), 'Tipo materiale' (M), 'Bid' (MOD13859E), 'Livello autorità' (95), 'Tipo record' (a), 'Paese' (IT), 'Lingua' (ITA), 'Genere', 'Tipo' (D), 'Data1' (2007), and 'Data2'. Below these are several text areas for 'Area del titolo', 'Area dell'edizione', 'Area della pubblicazione', 'Area della descrizione fisica', 'Area delle note', 'Nota di contenuto', 'Nota di abstract', 'Nota sul tipo di risorsa elettronica', and 'Nota sui requisiti'. The record details include the title 'Biblioteconomia : guida classificata / diretta da Mauro Guerrini ; condirettore Gianfranco Crupi ; a cura di Stefano Gambari, collaborazione di Vincenzo Fugaldi ; presentazione di Luigi Crocetti', the publisher 'Milano : Bibliografica, [2007]', and physical description 'XL, 1143 p. : ill. ; 25 cm.'.

Figura 4: Schermata del collegamento a SBN tramite SBN web

Di recente si è cominciato ad offrire la possibilità di utilizzare i software nella modalità **as a service (SAAS)**, in virtù della quale i clienti non acquistano il possesso del software, ma pagano per l'utilizzo dello stesso.

Esaminiamo ora i moduli di cui un ILS è normalmente composto:

- **catalogazione**: è il modulo che permette la creazione, l'aggiornamento e la gestione dei record bibliografici che costituiscono il catalogo della biblioteca. Normalmente è prevista un'interfaccia ai fini dell'inserimento strutturato per campi (titolo, edizione, pubblicazione, descrizione fisica etc.) secondo le aree delle ISBD. La punteggiatura standard utilizzata dalle ISBD per delimitare le aree (e talvolta anche gli elementi) in alcuni casi è inserita automaticamente dal software, in altri deve essere immessa manualmente dall'operatore. Determinati software prevedono la creazione del record bibliografico compilando direttamente i campi contrassegnati dalle etichette del MARC; normalmente tutti i software implementano un formato MARC per l'export dei dati e la loro condivisione.

Numerosi prodotti sono dotati di una maschera di immissione dati personalizzata a seconda del materiale da catalogare, che può presentare necessità di trattamento diverse (normalmente le opzioni disponibili riguardano monografie, periodici, libro antico, materiale non librario e, in alcuni casi, anche altri materiali speciali).

Al modulo di catalogazione sono collegate le liste presenti negli archivi di autorità. L'incremento di tali archivi può avvenire direttamente dalla maschera di catalogazione tramite l'inserimento nell'apposito campo collegato, oppure separatamente, nel qual caso dal modulo di catalogazione sarà possibile effettuare una ricerca e collegare il record di autorità necessario.

Spesso è possibile catturare record bibliografici sia da altri cataloghi on line (SBN, LOC etc.), grazie a un modulo Z39.50 integrato nel sistema, sia da risorse locali (CD, record precedentemente esportati e salvati). Se necessario, tali record possono essere anche modificati, in virtù appunto di questa modalità che viene definita **catalogazione "derivata"**. Per sistemi bibliotecari che adottano lo stesso software è invece possibile effettuare la cosiddetta **catalogazione "partecipata"**: in tal caso il catalogo bibliografico è unico per tutti, i record non vengono duplicati o scaricati in locale, mentre la biblioteca, dopo aver fatto una ricerca e individuato il record di interesse, si limita a localizzarlo, ovvero a inserire dati gestionali quali l'inventario, il posseduto, la collocazione etc., cosicché lo stesso record potrà essere condiviso da più biblioteche con un'unica catalogazione. Qualora il record non fosse presente, la biblioteca che per prima catalogherà provvederà ad inserire tutti i dati e renderà disponibile il record nel catalogo, dopodiché le altre biblioteche lo localizzeranno. La catalogazione partecipata è la principale funzionalità di SBN, che tramite procedure condivise permette alle biblioteche aderenti di lavorare in autonomia e con software diversi, integrandole in un sistema cooperativo basato su una rete nazionale;

- **circolazione:** è il modulo che gestisce il prestito del posseduto della biblioteca (e, se necessario, la consultazione in sede). È stato il primo processo automatizzato in biblioteca, cosicché può contare su oltre cinquant'anni di sviluppo, utilizzo e parametrizzazioni. In particolare consente di gestire il prestito, la restituzione e il rinnovo, il tutto con "settaggi" diversi da biblioteca a biblioteca, ma anche tra materiali diversi o tipologie di utenti differenti all'interno di una medesima biblioteca. Ad esempio, normalmente i DVD vengono prestati per una durata massima di 7 giorni e per un massimo di 3 esemplari per utente, mentre le monografie possono essere prestate per 30 giorni e fino a 5 per utente. Di norma il modulo produce una serie di statistiche fondamentali per la gestione e la valutazione del servizio di una biblioteca: numero degli utenti complessivi iscritti, numero degli utenti attivi nell'ultimo anno, numero di prestiti complessivi e per tipologia di materiale etc. Inoltre, incrociando i dati del prestito con le informazioni sull'utente è possibile fornire un servizio personale di disseminazione dell'informazione, con consigli specifici all'utente direttamente interessato e consigli ad altri utenti (secondo la formula lanciata da Amazon: «*chi ha letto questo... ha letto anche questo...*»). A parte o integrato direttamente viene reso disponibile un sottomodulo per la gestione del prestito interbibliotecario (ILL);
- **acquisti:** si tratta del modulo attraverso il quale vengono gestiti i dati amministrativi e contabili riguardanti le nuove acquisizioni, il *workflow* dell'impegno di spesa e della richiesta al fornitore, la gestione dell'*approval plan* (ovvero il processo di acquisizione del materiale librario di nuova pubblicazione in biblioteca in base ad accordi precisi con il fornitore e secondo un piano mirato di crescita delle collezioni) e degli ordini continui. Per ottimizzare i tempi e il lavoro viene normalmente immessa una precatalogazione del titolo ordinato; alcuni venditori forniscono come servizio il record MARC del libro acquistato, che sarà possibile importare nel catalogo. Tale modulo, spesso opzionale, viene utilizzato soprattutto da biblioteche medio-grandi o da sistemi bibliotecari con un centro acquisti unico;
- **periodici:** è il modulo, spesso opzionale, pensato per quotidiani, giornali e riviste a cui le biblioteche sono abbonate e che ricevono su base regolare. Consente di amministrare sia gli ordini e lo stato dell'abbonamento (segnalando, ad esempio, numeri mancanti), sia la gestione del posseduto e la sua eventuale circolazione.

## ONLINE PUBLIC ACCESS CATALOG (OPAC)



Il modulo che consente la pubblicazione on line del catalogo fu introdotto già alla fine degli anni Settanta. La possibilità di sostituire il vecchio catalogo cartaceo con un sistema automatizzato di ricerca consultabile direttamente dagli utenti anche da casa fu percepito fin dall'inizio come una fondamentale innovazione di grande potenzialità. Se inizialmente l'OPAC era concepito come una trasposizione del catalogo cartaceo, si è col tempo integrato con gli altri moduli dell'ILS, segnalando, ad esempio, anche la disponibilità mediante i dati del modulo di circolazione, sfruttando sempre più le possibilità offerte dal web e dall'interattività.

Per effettuare ricerche in un OPAC ci si avvale delle tecniche di *information retrieval* (utilizzo di operatori booleani, troncamento delle parole, ricerca per frase esatta etc.), cui si aggiungono modalità di navigazione quali lo scorrimento di liste e l'utilizzo dei link ipertestuali per ampliare o raffinare la ricerca.

Generalmente l'OPAC si presenta all'utente con una pagina introduttiva contenente alcune informazioni sul catalogo (consistenza, data di aggiornamento, funzionalità di ricerca etc.) e con un menu di scelta per la ricerca.

#### **Le possibilità disponibili sono:**

- **ricerca per campi:** i termini di cui si è in possesso per la ricerca vengono inseriti nei rispettivi campi (autore, titolo, soggetto etc.). Se, per esempio, si stanno cercando tutte le opere di Alessandro Manzoni presenti nell'OPAC, si digita la parola *Manzoni* nel campo "autore", mentre se si cerca l'opera *I promessi sposi* si potrà aggiungere, nel campo "titolo", la parola *sposi*. Spesso è possibile indicare, grazie all'utilizzo degli operatori booleani, la relazione esistente fra i campi, altrimenti, nel caso in cui questa opzione non fosse possibile, qualora venissero immessi valori in più di un campo, questi sarebbero normalmente immessi nella ricerca tramite AND;
- **ricerca per scorrimento liste:** l'utente accede alle liste di autorità (autori, soggetti etc.) e può navigare in esse ricercando e selezionando l'elemento desiderato; qualora il termine non venisse trovato, il sistema risponderebbe con le parole vicine nell'ordine alfabetico. Tramite questa ricerca è possibile evitare errori di spelling delle parole e immettere la stringa di soggetto esatta, o il nome dell'autore nella forma corretta;
- **ricerca libera:** tramite una singola maschera di inserimento dati possono essere digitati uno o più termini che verranno ricercati in tutti i campi del record.

Nell'effettuare la ricerca il sistema ignora i termini frequenti (articoli, preposizioni etc.), che vengono inseriti in una particolare lista di **stop words**; è inoltre indifferente immettere i termini in lettere maiuscole o minuscole. È invece necessario prestare attenzione allo spelling: un errore di battitura può produrre un risultato negativo, anche se le nuove generazioni di OPAC suggeriscono eventuali termini simili qualora l'esito della ricerca sia nullo o molto basso.

Nel caso in cui la ricerca abbia come risultato un unico record, la maggior parte degli OPAC risponde visualizzando direttamente il record completo con tutti i dettagli. Là dove venga recuperata più di una notizia bibliografica il sistema mostra una breve lista di record, all'interno dei quali l'utente può selezionare quello di interesse per visualizzarne i dettagli. Normalmente tale lista presenta gli elementi minimi necessari a identificare un'opera, mentre nel dettaglio del record vengono visualizzati tutti i dati immessi al momento della catalogazione, nonché le informazioni per il recupero della risorsa: collocazione (qualora si tratti di un oggetto posseduto dalla biblioteca) o URL (qualora si tratti di una risorsa remota disponibile on line).

La maggior parte degli OPAC consente altresì di raffinare la ricerca nel caso in cui venga recuperato più di un record: gli OPAC di nuova generazione, in particolare, visualizzano insieme alla lista breve i filtri che l'utente può selezionare per aumentare la precisione e ridurre il richiamo. Tali filtri vengono generati in base alle caratteristiche dei record recuperati (ad esempio, il sistema consentirà di selezionare la tipologia di materiale, qualora siano stati recuperati sia DVD che monografie; oppure la lingua dell'opera, nel caso in cui la biblioteca possieda l'opera in traduzione).

Grazie all'integrazione con il modulo "circolazione", nella maggior parte degli OPAC è possibile vedere in tempo reale la disponibilità del volume (DVD, periodico etc.) e l'utente può accedere alla propria area riservata per svolgere

comodamente da casa operazioni quali la prenotazione di un volume, il rinnovo del proprio prestito, l'invio di suggerimenti di acquisto, il salvataggio di ricerche effettuate etc.

Rispecchiando la struttura organizzativa e gestionale della biblioteca, l'OPAC può contenere i dati di un singolo istituto oppure può essere collettivo, qualora la biblioteca faccia parte di un sistema bibliotecario o di un consorzio, nel qual caso spesso l'utente può scegliere (selezionare) se effettuare la ricerca sul posseduto di tutte le biblioteche o di una in particolare.

In Italia l'OPAC collettivo più importante è quello di SBN (<http://opac.sbn.it/>), che attualmente ha una consistenza di oltre 11 milioni e mezzo di notizie bibliografiche.

Copyright © 2010 ICCU | Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche - Realizzato da Inera s.r.l.

Figura 5: La nuova maschera di ricerca di SBN

Strumento molto utile per la ricerca contemporanea su più OPAC è il **MAI - MetaOPAC Azalai Italiano** (<http://www.aib.it/aib/opac/mai2.htm>), gestito dall'AIB e dal CILEA (Consorzio interuniversitario lombardo per l'elaborazione automatica), che consente di interrogare con un'unica interfaccia web oltre 390 cataloghi connessi. Non si tratta, quindi, di un vero e proprio OPAC, essendo il MAI sprovvisto di un archivio di dati proprio, bensì di un OPAC di OPAC, che effettua la ricerca fra i record dei cataloghi connessi.

Strumento utilizzabile per recuperare notizie difficilmente reperibili, il MAI si basa su un repertorio, curato sempre dall'AIB e dal CILEA, che cataloga tutti gli OPAC di biblioteche italiane, consultabile all'indirizzo <http://www.aib.it/aib/opac/repertorio.htm> e attualmente costituito da circa 1.400 cataloghi, suddivisi per tipologia di biblioteca, area geografica e tipo di documenti; ne conseguono maschere di ricerca personalizzate, ovvero:

- “Ricerca selettiva” (<http://www.aib.it/aib/opac/mai.htm3>);
- “Ricerca per regione” (<http://www.aib.it/aib/opac/mai3.htm3>).

In ambito internazionale, invece, come MetaOPAC va ricordato il **KVK - Karlsruhe Virtual Catalog** ([http://www.ubka.uni-karlsruhe.de/kvk\\_en.html](http://www.ubka.uni-karlsruhe.de/kvk_en.html)), che consente la ricerca simultanea, tramite un'unica interfaccia, su cinquanta OPAC di tutto il mondo e una decina di fornitori librari commerciali.

Per i periodici è possibile ricercare in **ACNP - Archivio collettivo nazionale dei periodici** (<http://acnp.cib.unibo.it/cgi-ser/start/it/cnr/jfp.html>), catalogo dei periodici posseduti da 2.400 biblioteche in Italia. Per ogni testata vengono fornite le principali informazioni bibliografiche, gli opportuni legami a continuazioni, supplementi, titoli precedenti o varianti etc. Ciascuna biblioteca fornisce la propria consistenza e le informazioni sulla consultazione e le modalità di richiesta di fornitura di articoli. L'archivio contiene anche descrizioni di riviste elettroniche e un archivio di spogli.

A livello globale i più grandi OPAC del mondo sono il **catalogo della Library of Congress** (<http://catalog.loc.gov/>), con 14 milioni di record, e **Worldcat** dell'OCLC (<http://www.worldcat.org/>), con 240 milioni di record, frutto della partecipazione di 72mila biblioteche da 170 paesi di tutta la terra.

Gli **OPAC cosiddetti "di nuova generazione"** si sono concentrati su due aspetti principali:

- le modalità con cui la ricerca viene svolta e la presentazione dei risultati e dei record.** Numerosi OPAC presentano ora, come prima opzione di ricerca, la maschera "libera", con un unico campo in cui l'utente può digitare i termini, che verranno ricercati in tutti i campi disponibili dei record. In sostanza viene emulata la maschera del più famoso motore di ricerca nel web: *Google*. I risultati sono presentati in base a un algoritmo che calcola la rilevanza dei record rispetto ai termini immessi per la ricerca. I record bibliografici appaiono sempre più arricchiti, sia grazie a copertine, *abstract*, informazioni sull'autore e quant'altro il web renda disponibile gratuitamente, sia in virtù di appositi prodotti a pagamento;
- maggior interazione dell'utente con l'OPAC.** La famosissima libreria on line *Amazon* ha fatto scuola, sicché anche gli OPAC consentono oggi all'utenza di interagire con il sistema secondo i principi di quella che è stata definita la *Library 2.0*, ad esempio inserendo commenti e voti, oppure condividendo le informazioni su altri siti o *social network*. L'integrazione con il modulo "circolazione" e l'utilizzo di tecniche di marketing permettono all'OPAC di suggerire all'utente altre letture («*chi ha letto questo... ha letto anche questo...*»).



Figura 6: Una scheda bibliografica del portale Scoprirete della Rete bibliotecaria di Romagna e San Marino

Grazie all'evoluzione delle tecnologie per la ricerca e alla diffusione di prodotti digitali e di fonti di informazione on line sono stati introdotti nuovi strumenti per la ricerca in biblioteca. Lo scopo è connettere in maniera più efficace tutta l'informazione disponibile, per costruire un ambiente pienamente interconnesso: è quindi possibile collegare tra loro banche dati bibliografiche, di *abstract*, cataloghi on line delle biblioteche, archivi aperti con *full text* di articoli etc.

Uno degli approcci più significativi è stato il **linking**, ovvero la possibilità di creare collegamenti automatici direttamente dal record bibliografico verso altre fonti informative elettroniche (a pagamento o ad accesso libero). In pratica l'utente, una volta effettuata una ricerca su una risorsa e ottenuta una risposta (*origin*), può attivare i collegamenti esistenti e possibili (*source link*) ad altre risorse (*link target*), oppure accedere a servizi aggiuntivi, come la richiesta di prestito interbibliotecario o di *document delivery*, il collegamento a un sito di *e-commerce* per l'acquisto della risorsa etc.

**OpenURL**, il protocollo che consente lo scambio di metadati finalizzato alla gestione di servizi di *linking context-sensitive* (cioè contestualizzato), dal 2004 è uno standard aperto ANSI/NISO.<sup>1</sup>

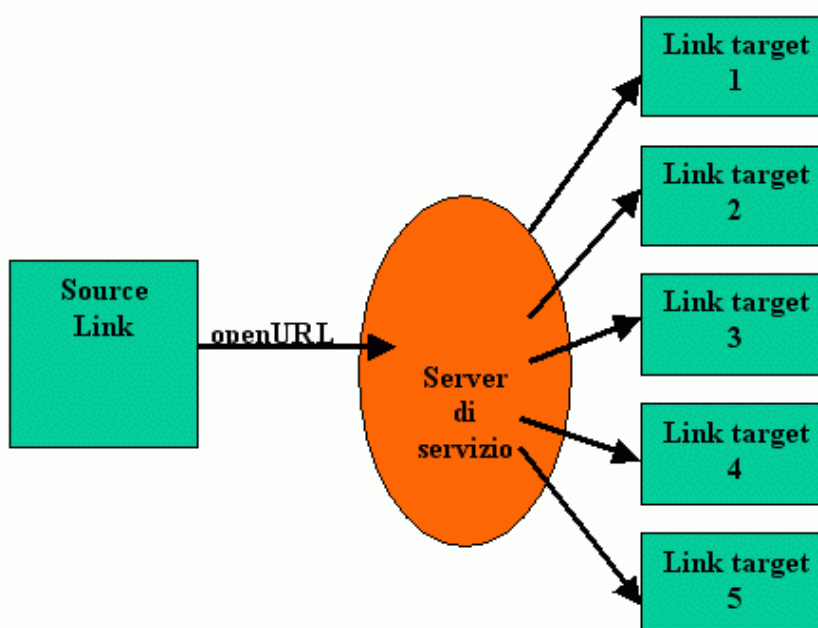


Figura 7: Rappresentazione grafica del funzionamento del protocollo OpenURL

Il prodotto considerato quasi per antonomasia "il" *link resolver* è **SFX**, distribuito da *Ex Libris* a partire dal 2001, attualmente in uso presso oltre 1.800 biblioteche di 50 paesi, ma sono comunque disponibili diversi altri prodotti, liberi o commerciali, che implementano questo standard.

Il passo successivo, sempre nel solco dell'interconnessione e con lo scopo di mettere a disposizione degli utenti il maggior numero possibile di risorse informative, consentendo una ricerca rapida ed efficace, è stato la diffusione di strumenti per la **ricerca federata**. Le biblioteche, infatti, arricchiscono sempre più le proprie collezioni fisiche con

---

<sup>1</sup> L'immagine utilizzata in Figura 7: Rappresentazione grafica del funzionamento del protocollo OpenURL è tratta da L'evoluzione dello standard OpenURL / Cinzia Bucchioni, «Bibliotime», a. VII, n. (novembre 2004), <http://didattica.spbo.unibo.it/bibliotime/num-vii-3/bucchion.htm>.

risorse elettroniche, come il *full text* dei periodici in abbonamento, anche se costringere l'utente ad effettuare più ricerche su differenti basi di dati non è efficiente.

Gli strumenti per la ricerca federata consentono, tramite un'unica interfaccia, di ricercare e presentare tutte le tipologie documentarie e le risorse della "biblioteca ibrida" provenienti dai molteplici fornitori. Oltre al software di metaricerca che normalmente viene collegato all'OPAC, le ditte che producono e distribuiscono questi strumenti forniscono le configurazioni per l'accesso a un numero esteso di risorse che permettono al motore di funzionare. Questa *knowledge base* preimpostata può arrivare a comprendere anche milioni di record ed è generalmente configurabile dalla biblioteca per consentire l'accesso solo alle risorse di interesse. Grazie a questi nuovi strumenti la biblioteca non solo permette all'utente di cercare le informazioni bibliografiche, come avveniva con l'OPAC tradizionale, ma fornisce direttamente l'informazione cercata proponendo tutte le risorse utili, anche quelle di cui l'utente può non essere a conoscenza.

## DIGITALIZZAZIONE

La *digitalizzazione* è il processo di creazione di un originale digitale che riproduce un originale analogico: ad esempio, la registrazione di un'esecuzione musicale su compact disc o la scansione elettronica di un manoscritto o di un quotidiano.

L'operazione consiste fondamentalmente nel passaggio dall'ambito dei valori continui a quello dei **valori discreti**, cioè numericamente discernibili, in maniera che il fenomeno fisico sia rappresentabile dall'elaboratore. Il passaggio dall'analogico al digitale comporta sempre una certa perdita di informazione e, in quanto tale, deve essere sempre frutto di attenta scelta.

Si noti che un brano musicale potrebbe essere registrato anche su un vinile, ma in tal caso non si parla di digitalizzazione, poiché il vinile resta un oggetto analogico: infatti, mentre la puntina del giradischi, muovendosi lungo il solco inciso, vibra riproducendo il suono registrato, nel caso del CD-ROM il brano registrato viene invece riprodotto attraverso una serie di 0 e di 1, cioè in linguaggio binario, cosicché la riproduzione avverrà attraverso l'elaborazione dei byte e non mediante un procedimento fisico, come ad esempio la vibrazione.

Ogni strumento o tecnologia di codifica comporta un'interpretazione dell'informazione originaria, dunque una sua trasformazione; a volte anche la rinuncia a particolari aspetti dell'informazione: per esempio, la digitalizzazione di un testo in formato *plain*, se da una parte consente di applicare le tecnologie di *text mining*, dall'altra esclude la possibilità di eseguire analisi paleografiche o grafologiche. Ciascun passaggio da un supporto a un altro comporta quindi alcune acquisizioni e diverse perdite, delle quali è importante essere consapevoli nel momento in cui si progetta una digitalizzazione.

### **Due sono gli aspetti da considerare nell'ambito dei progetti di digitalizzazione:**

1. la definizione e la struttura non soltanto dei metadati necessari per l'identificazione e la descrizione del documento digitale, ma anche dei metadati di tipo amministrativo e gestionale, indispensabili alla fruizione della risorsa;
2. la definizione di standard tecnici (come ad esempio quelli relativi al formato delle immagini digitali) quanto più possibile trasparenti e omogenei, anche in applicazione di standard internazionali.

Inoltre il processo di digitalizzazione deve tener conto in maniera particolare anche dei fenomeni legati all'invecchiamento della risorsa digitalizzata, riguardanti l'**obsolescenza dei formati e dei supporti**. Quando un software o una tecnologia per la decodifica di un'informazione digitale vengono abbandonati o quando l'hardware sul quale l'informazione era conservata non è più disponibile, allora le informazioni registrate possono essere facilmente perdute, semplicemente perché non sono più raggiungibili: tale problema è dovuto alla mancanza di standard, protocolli e metodi di conservazione digitale prestabiliti.

Molte versioni di programmi di word-processing, di hardware per lo *storage* di dati digitali, di standard di codifica per immagini o filmati risultano rapidamente sostituite con nuove versioni di software o con hardware completamente nuovo. I file nati per essere letti o modificati con un determinato programma (ad esempio, Microsoft Word) saranno illeggibili con altri programmi, così come, man mano che lo sviluppo di nuovi sistemi operativi e di nuovo hardware andrà avanti, anche le vecchie versioni di programmi sviluppati dalle stessa società diverranno inutilizzabili sulla nuova piattaforma (ad esempio, le versioni precedenti di Microsoft Works, prima di Works 4.5, non possono essere eseguite in Windows 2000 o in sistemi operativi successivi).

Le modalità attualmente più diffuse per la conservazione dell'informazione digitale sono:

- l'**emulazione**, cioè la riproduzione delle funzionalità di un sistema obsoleto all'interno di un sistema più recente. In quanto tale l'emulazione si è rivelata una strategia "popolare" per riportare in funzione videogiochi di vecchie piattaforme, sebbene non sia stata adottata in modo massiccio dalla comunità;
- il **refreshing**, cioè il trasferimento di dati tra due supporti digitali dello stesso tipo (ad esempio, da un compact disc a un altro). Questa strategia deve essere combinata con quella della migrazione quando il software del computer o l'hardware richiesto per leggere i dati non sono più disponibili, oppure risultano inadatti a leggere i formati dei dati stessi. Il *refreshing* può comunque diventare necessario in caso di deterioramento fisico del supporto;
- la **migrazione**, da intendersi come il trasferimento dei dati su un nuovo sistema, anche su supporti digitali diversi. Questa strategia può comportare la conversione delle risorse da un formato a un altro (ad esempio, la conversione di un documento Word in un PDF o in un OpenDocument), da un sistema operativo a un altro (ad esempio, da Solaris a Linux), o da un linguaggio di programmazione a un altro (ad esempio, dal linguaggio C a Java), in modo che la risorsa resti pienamente accessibile e funzionale. I dati che subiscono il processo di migrazione corrono il rischio di perdere taluni tipi di funzionalità con i nuovi formati (incapaci di lavorare o leggere), così come il convertitore o programma utilizzato per la conversione può essere incapace di interpretare tutte le particolarità del formato originale.

In aggiunta alle strategie sin qui descritte occorre menzionare, infine, la **duplicazione**, cioè la creazione di copie dei dati su più sistemi, non tanto come rimedio all'obsolescenza, quanto come rimedio a possibili guasti o incidenti che si possono verificare. Infatti, i dati disponibili in singola copia su un unico sistema sono soggetti al rischio di *crash* del software che li fa girare o dell'hardware, di alterazione intenzionale o accidentale, di catastrofi ambientali quali incendi, allagamenti etc., cosicché attraverso la duplicazione su più sistemi in posti differenti essi hanno molte più possibilità di mantenersi integri. La pratica del **backup** degli applicativi e dei dati rientra nell'ambito della duplicazione.

## OPEN ACCESS (ACCESSO APERTO)

*Open Access* (Accesso Aperto) significa «**accesso libero e senza barriere al sapere scientifico**». Si tratta di un movimento o, meglio, di una serie di strategie nate all'interno del mondo accademico, il cui scopo è riguadagnare possesso della comunicazione scientifica offrendo libero accesso ai risultati della ricerca.

Sfruttando le potenzialità offerte dalla Rete, gli articoli vengono resi accessibili gratuitamente, senza le restrizioni e le barriere previste dalle licenze tradizionali. La disseminazione dell'informazione garantisce un reale impatto: più un articolo è liberamente scaricabile, più è letto, più viene citato; ciò favorisce la condivisione del sapere e quindi un più rapido avanzamento della conoscenza, senza barriere, in tutto il mondo.

L'espressione *Open Access* fu utilizzata per la prima volta in un documento pubblico nel manifesto della *Budapest Open Access Initiative* (BOAI) del 14 febbraio 2002.

Nel complesso, **tre documenti** hanno accompagnato la nascita e lo sviluppo del movimento OA:

- il succitato manifesto di Budapest (BOAI) del febbraio 2002;

- il *Bethesda Statement*, reso pubblico nel giugno 2003 al termine di un incontro promosso dall'*Howard Hughes Medical Institute* (HHMI);
- la *Dichiarazione di Berlino* promossa dalla *Max Planck Gesellschaft* nell'ottobre 2003.

I tre documenti sono stati comunemente recepiti come unitari, ovvero come tappe di un progetto condiviso, tanto che **Peter Suber**, uno dei maggiori protagonisti del movimento, ha coniato la formulazione **BBB Definition** per evidenziare gli aspetti comuni delle definizioni di OA presenti nei loro dettati.

In **Italia** il movimento OA si affermò nel 2004 con un atto formale firmato a Messina: il 4 e 5 novembre di quell'anno, infatti, la Commissione CRUI per le Biblioteche di ateneo, in collaborazione con l'Università degli studi della città siciliana, organizzò un convegno con l'obiettivo di promuovere nel sistema universitario italiano i benefici derivanti dal ricorso a forme di editoria elettronica ad Accesso Aperto. Con cerimonia solenne, i rettori di trenta atenei sottoscrissero il «Documento italiano a sostegno della Dichiarazione di Berlino sull'Accesso Aperto alla letteratura accademica», intitolato *Gli atenei italiani per l'«Open Access»: verso l'Accesso Aperto alla letteratura di ricerca*, noto anche come **Dichiarazione di Messina** (al luglio 2011 la Dichiarazione risulta sottoscritta dalla quasi totalità delle università italiane).

Due sono le **strategie di pubblicazione** dell'*Open Access*:

1. la **Green Road**, cioè l'auto-archiviazione in archivi aperti. Si tratta di archivi digitali a carattere istituzionale o disciplinare, nei quali si deposita il *pre-print* o il *post-print* dell'articolo in accordo con le politiche di copyright dell'editore;
2. la **Gold Road**, vale a dire la pubblicazione su riviste ad Accesso Aperto che garantiscono la *peer review*, ma adottano un diverso modello economico, in virtù del quale non è richiesto alcun pagamento per accedere ai testi e i costi di pubblicazione sono coperti da una quota versata dall'autore o dalla sua istituzione (la tendenza è di ricomprendere i costi di pubblicazione nel budget iniziale stanziato per la ricerca).

Di recente, sulla scorta del crescente numero di enti di finanziamento che richiedono la pubblicazione *Open Access* per le ricerche finanziate con fondi pubblici, molti editori scientifici hanno cominciato a sperimentare modelli ibridi: se l'istituzione copre le spese di pubblicazione, l'articolo viene reso immediatamente disponibile in modalità *Open* anche sulle riviste tradizionali.

Strumento principale della *Green Road* sono gli archivi istituzionali, i quali, secondo una recente definizione di Mauro Guerrini, sono collezioni di risorse digitali *full-text* realizzate e gestite da una o più università (o da altri istituti) e contenenti i risultati di ricerche originali finanziate con denaro pubblico e privato, prodotte nell'ambito dell'attività istituzionale da docenti, ricercatori, dottori di ricerca e altri soggetti. La raccolta è liberamente accessibile in Rete dalla comunità degli studiosi e dei lettori e viene costituita tramite la volontaria auto-archiviazione (o autorizzazione al deposito) da parte dell'autore o mediante regolamenti che prevedono l'obbligo di deposito (*mandate*). In ogni caso l'auto-archiviazione rispetta sia la politica di deposito stabilita dall'ente, sia le norme che tutelano i diritti delle opere, degli autori e degli editori.

L'IR, realizzato secondo standard definiti nell'ambito dell'**Open Archive Initiative (OAI)**, comprende diverse tipologie di risorse, talvolta organizzate in un'unica sequenza nel medesimo archivio, talaltra in sequenze o liste distinte: libri, capitoli di libro, saggi e articoli già editi su riviste on line o cartacee, atti di convegni, relazioni presentate a convegni, *working papers*, rapporti tecnici, brevetti, materiale didattico, tesi di laurea, tesi di dottorato e, più in generale, testi, immagini, documenti multimediali. Ciascun contributo è corredato di metadati descrittivi, semantici e gestionali che ne consentono il reperimento attraverso motori di ricerca e altri strumenti di recupero dell'informazione.

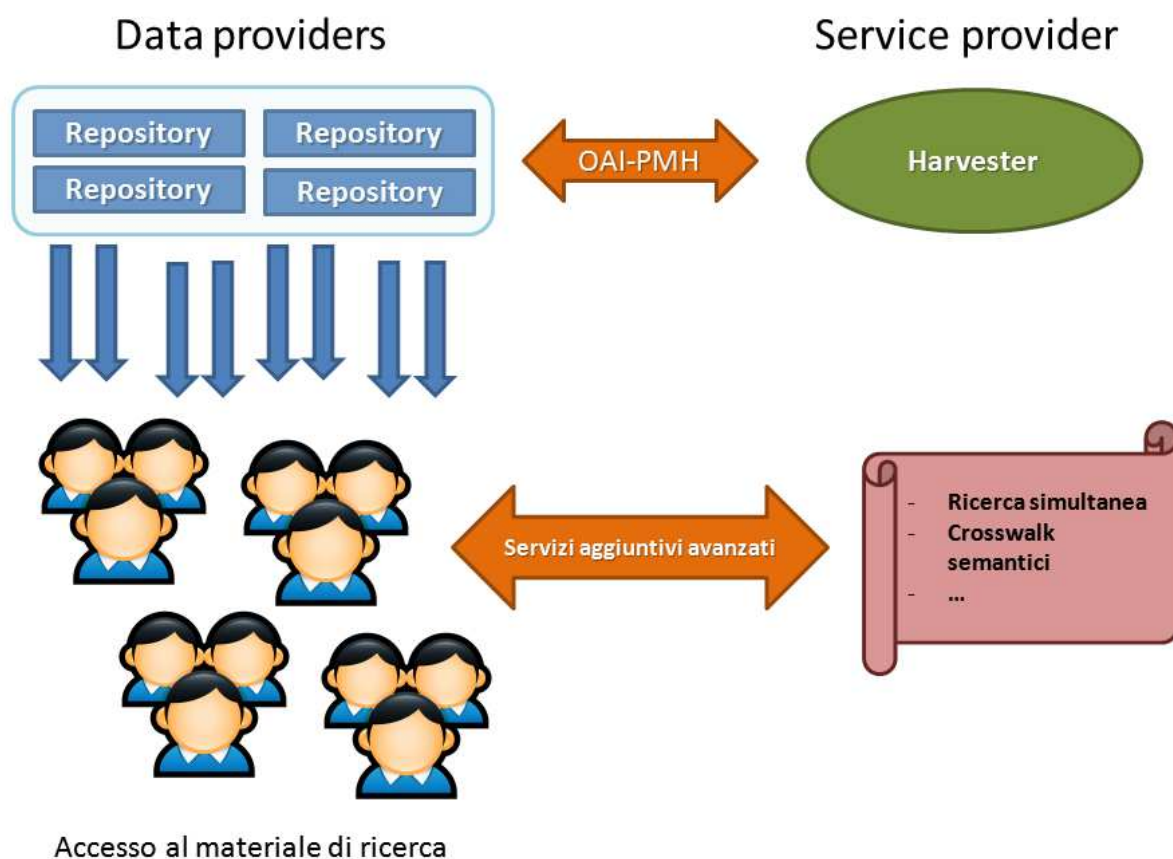
In pratica l'IR costituisce una vetrina della produzione scientifica dell'ente, un modello di comunicazione e disseminazione della conoscenza nell'era di Internet, ovvero un servizio d'informazione e documentazione rivolto alla collettività. I vantaggi per gli autori che depositano i propri contributi in un IR sono molteplici; in primis, una rapida e



ampia diffusione dei risultati della loro ricerca: rapida, perché non dovrà aspettare i tempi di pubblicazione tipici della stampa; ampia, perché il contributo sarà accessibile in un ambiente digitale diffuso a livello internazionale, con conseguente possibilità di aumentare la visibilità e il grado d'impatto della ricerca stessa (infatti, più un articolo è liberamente scaricabile, più facilmente potrà essere letto e quindi potrà essere, presumibilmente, più citato).

L'**architettura informativa dell'OAI** prevede l'esistenza di due diverse tipologie di servizi:

- i **data provider**, contenitori di *full-text* e relativi metadati;
- i **service provider**, che raccolgono tramite *harvesting* i dati da diversi *data provider* e permettono agli utenti di utilizzare interfacce semplici e a valore aggiunto.



Andrea Marchitelli, *Architettura OAI*, 2011

Figura 8: Rappresentazione grafica dell'architettura OAI

La prevista interoperabilità viene raggiunta attraverso l'uso generalizzato del protocollo **OAI-PMH** (*OAI-Protocol for Metadata Harvesting*), attualmente alla versione 2.0 (<http://www.openarchives.org/OAI/openarchivesprotocol.html>). OAI-PMH veicola, attraverso HTTP (il protocollo di scambio dei dati caratteristico del web), dati in XML e gli argomenti delle richieste vengono trasformati in parametri GET o POST.

Il protocollo ammette **sei tipi di richieste (verbs)**:

- **Identify:** richiede all'archivio di identificarsi e presentare una propria descrizione, secondo elementi standard (URL di esempio: <http://export.arxiv.org/oai2?verb=Identify>);
- **ListMetadataFormats:** richiede la lista di tutti i formati di metadati supportati dall'archivio (URL di esempio: <http://export.arxiv.org/oai2?verb=ListMetadataFormats>);



- **ListSets:** richiede all'archivio di mostrare la struttura dei set (URL di esempio: <http://export.arxiv.org/oai2?verb=ListSets>), intendendo per *set* gli elementi opzionali di raggruppamento che permettono di costruire insiemi tipologici di risorse contenute in un *data provider* (ad esempio per tipologia);
- **ListIdentifiers**, forma abbreviata di *ListRecords*: richiede solamente gli *header* degli item (URL di esempio: [http://export.arxiv.org/oai2?verb=ListIdentifiers&metadataPrefix=oai\\_dc](http://export.arxiv.org/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc), dove si noti la compresenza di due parametri, il *ListIdentifiers* e il *metadataPrefix*, che richiede il filtro dei dati solo agli identificatori che hanno dati in *oai\_dc*);
- **ListRecords:** raccoglie l'elenco dei record da un archivio (URL di esempio: [http://export.arxiv.org/oai2?verb=ListRecords&metadataPrefix=oai\\_dc](http://export.arxiv.org/oai2?verb=ListRecords&metadataPrefix=oai_dc));
- **GetRecord:** richiama un record individuato (URL di esempio: [http://export.arxiv.org/oai2?verb=GetRecord&&metadataPrefix=oai\\_dc&identifier=oai:arXiv.org:1108.3824](http://export.arxiv.org/oai2?verb=GetRecord&&metadataPrefix=oai_dc&identifier=oai:arXiv.org:1108.3824), dove si notino i parametri aggiuntivi *identifier* e *metadataPrefix*, i quali specificano che la richiesta concerne i metadati di uno specifico record individuato con il suo identificatore, cioè una chiave di riconoscimento univoca del record all'interno dell'archivio, in formato *oai\_dc*).

Il set di metadati obbligatorio definito dal protocollo è il cosiddetto ***oai\_dc***, sostanzialmente la versione semplificata del già citato *Dublin Core*, anche se ogni *data provider* può esporre i metadati dei record che contiene anche in altri formati. Questa scelta, se da una parte garantisce una forte interoperabilità, dall'altra conduce a una perdita di informazione, poiché il set *oai\_dc* rischia di semplificare e appiattire eccessivamente il contenuto informativo.

In riferimento ai *data provider*, tra i software maggiormente utilizzati vanno ricordati:

- per la realizzazione di archivi aperti: **DSpace** (<http://www.dspace.org/>, realizzazione di esempio, E-LIS - eprints in Library and Information Science: <http://eprints.rclis.org>), **Eprints** (<http://www.eprints.org/>, realizzazione di esempio: archivio istituzionale dell'Università di Trento, <http://eprints.biblio.unitn.it/>);
- per la realizzazione e gestione di riviste OA: **OJS - Open Journal Systems** (<http://pkp.sfu.ca/?q=ojs>, realizzazione di esempio: <http://jilis.it>).

Diverse, invece, le realizzazioni di *service provider*. Il primo e più importante al mondo fu **OAister**, ora acquisito da OCLC e divenuto parte di WorldCat (<http://oaister.worldcat.org/>). Per l'Italia va senz'altro citato **PLEIADI**, il *service provider* nazionale per la letteratura scientifica (<http://www.openarchives.it/pleiadi/>). Imprescindibili, altresì, i grandi repertori mondiali che elencano archivi e riviste OA, come ad esempio **DOAJ** (<http://www.doaj.org>), la Directory delle riviste ad Accesso Aperto, e **OpenDOAR** (<http://www.opendoar.org/>), la Directory dei repository ad Accesso Aperto.

### Biblioteca digitale

La biblioteca digitale, di cui oltre verranno definiti alcuni dei principali componenti, è un sistema informativo digitale distribuito che nasce dall'esigenza di mettere a disposizione dell'utenza risorse digitali complesse, cioè oggetti digitali accessibili in ambiente strutturato, in maniera tale che esse siano conservate in maniera persistente e durevole, efficacemente ritrovabili, univocamente identificate e dotate di contenuto informativo e metadati.

Fra le molte definizioni esistenti, presentiamo in traduzione italiana quella della *Digital Library Federation* (<http://old.diglib.org/about/dldefinition.htm>), estremamente sintetica, ma assai profonda:

*«Le biblioteche digitali sono organizzazioni che mettono a disposizione risorse, incluso uno staff specializzato, per selezionare, strutturare, offrire accesso intellettuale, interpretare, distribuire, conservare l'integrità e assicurare la persistenza nel tempo di collezioni di opere digitali in maniera tale che esse siano prontamente ed economicamente disponibili per l'uso da parte di una comunità o di un raggruppamento di comunità».*

Da tale definizione emergono alcuni aspetti importanti, il primo dei quali è costituito dal fatto che ci troviamo di fronte a «collezioni» di risorse, nel senso che i singoli documenti non sono considerati nella loro unitarietà e atomicità, ma piuttosto vengono organizzati attraverso i loro collegamenti, che ne permettono la navigazione ipertestuale; inoltre, tali collezioni sono strutturate, cioè gestite attraverso una qualche riconoscibile architettura, come quella di un database.

Le relazioni esistenti tra i documenti conservati non sono necessariamente limitate alla sola collezione, ma possono allargarsi anche ad altri documenti e collezioni. In tal senso, spesso il World Wide Web è considerato esso stesso un paradigma di biblioteche digitali, pur mancando di alcuni elementi fondamentali, come la struttura organica e consapevole dell'informazione che contiene.

Dunque l'utente dispone, oltre che di contenuti, anche di una serie di servizi e strumenti che la biblioteca digitale gli mette a disposizione per la fruizione dei contenuti stessi. In linea di massima, poiché tali entità, secondo quanto dichiara anche il Manifesto per le biblioteche digitali (<http://www.aib.it/aib/cg/qbdigd05a.htm3>), sono anzitutto delle biblioteche, esse mettono a disposizione degli utenti servizi di tipo bibliotecario. In particolare, gestiscono una selezione e acquisizione delle risorse, la loro conservazione nel tempo, l'organizzazione all'interno della collezione e la loro catalogazione descrittiva e semantica. Inoltre, dal punto di vista della distribuzione, forniscono servizi di supporto al reperimento e alla consultazione dei materiali.

Ecco alcuni esempi di biblioteche digitali italiane ed estere:

- **Internetculturale**: si tratta del progetto, avviato dalla Direzione generale per i beni librari del Ministero per i beni e le attività culturali e gestito dall'ICCU, che rende disponibili i documenti tramite il portale <http://www.internetculturale.it/>, dove sono visibili i documenti digitalizzati a bassa risoluzione dalle biblioteche e da altre istituzioni culturali italiane aderenti al progetto;
- **Manuzio** (<http://www.liberliber.it/>): progetto nato nel 1993 e promosso dall'associazione *Liber liber*, che diffonde gratuitamente testi letterari italiani (classici, o comunque opere non protette dalla legge sul diritto d'autore);
- **Gallica** (<http://gallica.bnf.fr>): biblioteca digitale nata nel 1997 dalle collezioni della Biblioteca nazionale di Francia. Ad oggi offre più di un milione e mezzo di documenti digitalizzati, costituiti da immagini, testi, documenti sonori etc.;
- **Europeana** (<http://www.europeana.eu/>): è la Biblioteca Digitale Europea, contenente circa quindici milioni di risorse;

- **The European Library** (<http://www.theeuropeanlibrary.org>): progetto promosso dalla Commissione Europea, dedicato alla ricerca del posseduto delle biblioteche nazionali europee, che consente anche di ricercare testi digitalizzati provenienti da una parte di esse;
- **Google books** (<http://books.google.com>): lanciato nel 2004 da *Google* con la digitalizzazione delle raccolte di alcune biblioteche prima solo statunitensi e poi anche europee, mostra informazioni relative ai libri e a parte di essi (se il libro non è protetto da copyright viene visualizzato per intero). Sono inoltre presenti link verso librerie on line per acquistare un libro, per cercare recensioni nel web, per trovare un libro in una biblioteca.

## Periodici elettronici (*E-JOURNAL*)

I periodici elettronici, o **e-journal**, sono oggi una componente fondamentale delle collezioni della biblioteca digitale. A diversi anni dall'avvento del web, parlando di *e-journal* ci si riferisce a periodici digitali disponibili, liberamente o mediante la sottoscrizione di abbonamenti, attraverso la rete.

Una suddivisione utile di questo tipo di risorse identifica:

- **periodici digitali nativi**, disponibili sin dalla loro origine in formato elettronico;
- **periodici digitalizzati**, cioè realizzati attraverso un procedimento di trasformazione dell'originale cartaceo (analogico) in uno digitale.

Dal punto di vista commerciale entrambe le categorie ospitano, sebbene in proporzioni diverse, sia periodici ad accesso gratuito, sia periodici ad accesso tariffato. Sotto il profilo dell'autorevolezza, riviste dotate di comitato scientifico, *peer review* e ISSN sono anch'esse presenti trasversalmente.

I periodici elettronici ad accesso tariffato sono generalmente disponibili, dietro corresponsione di una quota di abbonamento, sul sito degli editori o di particolari mediatori (*aggregatori*), i quali possono svolgere anche funzione editoriale in proprio.

Attraverso tali servizi è possibile non solo ricercare il titolo della testata e scorrere il contenuto dei diversi fascicoli, ma anche ritrovare, tramite i canali di autore e titolo, singoli articoli pubblicati, poiché a differenza della tradizionale pubblicazione cartacea quella digitale offre una maggiore **granularità**, ossia un maggiore livello di dettaglio. In tal senso risultano fondamentali i servizi di *reference linking*, legati al protocollo OpenURL, che facilitano l'integrazione fra le diverse risorse elettroniche: database, periodici elettronici, cataloghi ed altre risorse informative. Infatti consentono il collegamento diretto tra le citazioni (contenute in articoli, libri, *e-prints* etc. e rintracciate in una determinata banca dati), le risorse disponibili e gli appositi servizi erogati per gli utenti di uno specifico ente, come un'università o un centro di ricerca.

Grazie alla diffusione delle politiche e dei modelli economici per l'editoria ad Accesso Aperto, le biblioteche si stanno affermando, altresì, come protagoniste attive della creazione, oltre che della selezione, catalogazione e conservazione, di questo tipo di risorse. Basti pensare che il software più diffuso al mondo per la creazione e la gestione di periodici elettronici, **OJS - Open Journal Systems** (<http://pkp.sfu.ca/?q=ojs>), venne inizialmente sviluppato per conto della biblioteca della canadese Simon Fraser University.

## E-book

Un *e-book* è un libro in formato digitale. Si tratta di file consultabili su computer, telefonini di ultima generazione (*smartphone*) e appositi lettori digitali. Il termine deriva dalla contrazione delle parole inglesi **electronic book** e viene utilizzato per indicare la **versione digitale di una qualsiasi pubblicazione**.

Nel caso ci si voglia riferire al dispositivo di lettura è più corretto indicarlo come lettore di *e-book*, in inglese ***e-book reader***, termine con il quale si intende sia il dispositivo hardware su cui esso viene letto, sia il software che permette la lettura del file.

Per la lettura di un *e-book* sono necessari diversi componenti:

- il documento elettronico di partenza;
- un formato elettronico con cui realizzare la pubblicazione (ad esempio, l'*ePub*);
- un software di lettura compatibile con il suddetto formato;
- un dispositivo hardware di lettura (il più appropriato per la lettura di testi è un *e-book reader* con tecnologia *e-ink*).

È sbagliato confondere un qualunque documento in formato digitale con un *e-book*, il quale non si limita a presentare la sostanza del documento cartaceo, ma cerca anche di replicarne la forma, in modo da rendere la lettura il più possibile simile a quella che si avrebbe sfogliando le pagine di un libro; da ciò deriva che tutte le azioni normalmente svolte su un libro cartaceo, come lo scorrimento delle pagine o l'inserimento di un segnalibro, possono essere emulate dal software del dispositivo di lettura.

Il libro elettronico, nell'imitare quello cartaceo, approfitta ovviamente dei vantaggi offerti dalla propria natura digitale, che risiedono principalmente nella possibilità di:

- essere un ipertesto, con link che conducono da un punto all'altro del documento, o a documenti diversi (anche disponibili sul web e non solo residenti sul lettore);
- inglobare elementi multimediali;
- utilizzare dizionari o vocabolari contestuali;
- eseguire ricerche sul testo dell'opera in lettura.



Figura 9: Un'immagine del lettore di e-book "Kindle" della Amazon

Gli *e-book* vengono realizzati e pubblicati nei più svariati formati, molti dei quali, però, come il PDF o altri formati testuali, non sono stati originariamente concepiti per essere dei veri e propri formati di *e-book*.

I nuovi formati, invece, studiati appositamente per i libri elettronici, se da un lato hanno consentito di superare i limiti dei formati più vecchi, dall'altro hanno tuttavia creato altri problemi, in quanto il mercato ha indirizzato l'industria editoriale verso formati "proprietary", ossia legati a specifiche piattaforme hardware e software, che condizionano fortemente le scelte dell'utente finale. Questo fa sì che all'utente possa capitare, ad esempio, di doversi confrontare con *e-book* (file) che possono essere letti solo con determinati tipi di software, oppure con un hardware che non supporta il formato di *e-book* che si possiede. Per tentare di risolvere questi inconvenienti sono nati progetti come l'*Open eBook Forum*, che, basandosi su codice aperto, si propone di realizzare un formato di *e-book* per molte piattaforme diverse e aperto ad implementazioni audio-video, nonché ad animazione e forme narrative non lineari.

I formati appositamente ideati per gli *e-book* prevedono generalmente la compressione del testo, con una riduzione dell'ingombro del file in memoria, e si adattano meglio ad essere visualizzati sui piccoli schermi dei dispositivi portatili. Altra caratteristica che spesso possiedono è quella di prevedere sistemi di protezione del documento (DRMS - *Digital Rights Management Systems*) in grado di impedire la copia di libri protetti da copyright.

Tra i formati più usati per gli *e-book reader* vanno certamente menzionati i seguenti:

- ePub, formato aperto, evoluzione di OeB, che si sta affermando come lo standard più diffuso;
- Mobipocket (mobi), utilizzato da Amazon Kindle;

- **LIT**, formato nativo Microsoft, ormai obsoleto.

Nella giungla dei diversi formati svolgono un ruolo fondamentale alcuni convertitori, fra i quali spicca **Calibre** (<http://calibre-ebook.com/>), che si distingue per le notevoli doti di funzionalità e flessibilità. Si tratta di un software *open source* disponibile per sistemi operativi Microsoft, Mac e Linux, che consente di conservare e organizzare i propri *e-book* e convertirli nel formato leggibile dal proprio lettore (ovviamente in assenza di sistemi crittografici per la protezione). Inoltre permette anche di creare *e-book* in numerosi formati (ePub compreso), come pure rende possibile la creazione di *e-book* a partire da documenti esportati dal web.

## Banche dati bibliografiche e indici citazionali

Una *banca dati* è una collezione di informazioni registrate in formato leggibile dall'elaboratore, relative ad un preciso dominio di conoscenze, organizzata allo scopo di poter essere consultata dai propri utilizzatori.

Le **banche dati** possono essere anzitutto divise in:

- **primarie o di fonte**, contenenti informazioni immediatamente utilizzabili, ad esempio con *full text* dei record;
- **secondarie o di riferimento**, contenenti riferimenti a informazioni da recuperarsi in un secondo momento.

Generalmente le banche dati bibliografiche sono a pagamento, anche se alcune sono disponibili gratuitamente on line. Fra le più importanti vanno annoverate le bibliografie nazionali, che registrano tutte le pubblicazioni (a stampa o su altri supporti) prodotte in un determinato paese, e i cataloghi dei libri in commercio, i quali includono tutto ciò che si può reperire sul mercato librario nazionale ad una determinata data. Per entrambe le tipologie, nel caso dell'Italia si segnalano la BNI (Bibliografia nazionale italiana), curata dalla Biblioteca nazionale centrale di Firenze, e *Alice*, catalogo italiani dei libri in commercio.

La nascita degli **indici di citazione** è legata al cosiddetto *impact factor*, sistema utilizzato per valutare il valore di un'opera in ambito scientifico, calcolando il suo impatto sulla letteratura e sul pensiero del tempo in base al numero di citazioni ricevute da altri autori e opere. Gli indici di citazione sono pubblicati dall'**ISI (Institute for Scientific Information)**, che mantiene una banca dati grazie alla quale, oltre a passare dal citante al citato, si può cercare chi cita un determinato documento.

In seguito alla nascita del movimento *Open Access* è stato creato un indice sperimentale di citazioni per la letteratura scientifica OA denominato **Citebase** (<http://www.citebase.org/>), contenente riferimenti a *pre-* e *post-print* da archivi aperti di fisica, matematica, scienze dell'informazione e medicina. **Google Scholar** (<http://scholar.google.com/>) è invece la denominazione del prodotto creato dal più famoso motore di ricerca sulla letteratura accademica. Consente un accesso facile e immediato a diverse fonti (riviste, libri, monografie e articoli provenienti da editori e da società professionali, ma anche archivi aperti e siti web), con lo scopo di elencare gli articoli in base all'*impact factor*, dunque valutando il testo completo dell'articolo, l'autore, la pubblicazione in cui è riportato, nonché il numero di volte in cui viene citato in altri documenti accademici. I risultati più pertinenti vengono sempre visualizzati nella prima pagina.

## Il web

**Internet**, soprattutto negli ultimi vent'anni, ha avuto una diffusione e una crescita esponenziali, al punto che si calcola che attualmente esistano oltre 450 milioni di siti web (<http://news.netcraft.com/archives/2011/08/05/august-2011-web-server-survey-3.html>). La relativa facilità del linguaggio HTML, la possibilità di comprare spazi sul web o di averne anche gratis e la mancanza di una "selezione editoriale" consentono sostanzialmente a chiunque di mettere on line qualsiasi tipo di contenuti.

La ricerca in ambiente WWW si avvale di strumenti assai potenti di ricerca per parole, quali sono i motori di ricerca (Google in primis). Per trovare una soluzione all'*information overload* sono stati creati dei **virtual reference desk**, il cui nome rimanda alla sala di consultazione di una biblioteca e che non a caso risultano utili nelle ricerche di *quick*

*reference* negli istituti bibliotecari. Esistono VRD specializzati, che organizzano risorse attinenti a una determinata disciplina o argomento, e spesso sono realizzati da esperti del settore. Nella maggior parte dei casi i link vengono valutati e commentati.

A tale proposito, in ambito italiano si possono ricordare:

- Segnaweb (<http://www.segnaweb.it>), a cura di AIB e CILEA;
- VRD per le biblioteche della Regione Toscana (<http://www.bibliotecari.it/SPT/VRD/SPT--Home.php>), organizzato secondo il sistema di classificazione Dewey;
- Eureka (<http://biblioteche.provincia.vicenza.it/eureka/>), a cura di Renzo Brun per il Servizio bibliotecario provinciale vicentino;
- Risorse web (<http://www.biblioteca.colognomonzese.mi.it/index2.php?consez=risorse>), a cura della Biblioteca civica di Cologno Monzese.

A livello internazionale, invece, occorre menzionare:

- INTUTE (<http://www.intute.ac.uk/>);
- ipl2: information you can trust (<http://www.ipl.org>);
- The WWW virtual library (<http://www.vlib.org>), corrispondente al catalogo più datato, iniziato da Tim Berners-Lee, l'ideatore del web.

Sono accessibili on line anche alcuni strumenti come **dizionari, enciclopedie e risorse similari di prima consultazione**, fra cui:

- <http://www.sapere.it> della De Agostini, che presenta on line numerose informazioni visualizzabili gratuitamente, mentre altre sono solo per gli abbonati;
- <http://www.britannica.com/>, versione on line della prestigiosa enciclopedia omonima, anche in questo caso in parte gratuita e in parte a pagamento;
- <http://education.yahoo.com/reference/encyclopedia/>, *Columbia Encyclopedia* in versione ridotta, consultabile gratuitamente su Yahoo!;
- <http://www.encyclopedia.com>, contenente informazioni provenienti da varie fonti (*The Columbia Encyclopedia*, *Oxford's World Encyclopedia*, *the Encyclopedia of World Biography*);
- [http://it.wikipedia.org/wiki/Pagina\\_principale](http://it.wikipedia.org/wiki/Pagina_principale), enciclopedia multimediale e multilingue al cui sviluppo può contribuire chiunque, gestita da utenti volontari e autofinanziata tramite sottoscrizioni e donazioni (il progetto in italiano, avviato nella primavera del 2001, conta ad oggi oltre 830.000 voci).

Un grosso limite riguardo alle fonti informative on line è la **valutazione dei contenuti**, poiché il web è un contenitore immenso di informazioni di qualsiasi tipo provenienti da tutto il mondo. Mancando sostanzialmente una politica di controllo editoriale su quanto pubblicato, chiunque può veicolare qualunque tipo di informazione in forma anonima per l'utente finale.

La proposta di **Elena Boretti** (<http://www.aib.it/aib/contr/boretti1.htm>) è di adattare la famosa griglia di **Whittaker**, ideata per le fonti cartacee, ai documenti della Rete; le pagine web verrebbero così valutate considerando i seguenti aspetti:

- *autorevolezza*: va verificato se il nome dell'autore è specificato e se il suo iter di studi o quello professionale danno garanzia riguardo al trattamento dell'informazione;
- *progetto*: l'optimum è che scopo della pubblicazione, destinatari e tipologia di trattamento del soggetto siano dichiarati nel sito;
- *contenuto*: occorre analizzare la chiarezza con cui l'argomento è trattato, la profondità, l'accuratezza, l'affidabilità, ma anche l'eventuale adozione di un particolare punto di vista che potrebbe rendere parziale la trattazione. Anche lo stile utilizzato gioca un ruolo importante, come il livello di esposizione. Altri elementi importanti sono la frequenza di revisione e la presenza di una bibliografia e di link esterni;
- *struttura*: se per le fonti cartacee si valuta la presenza di indici, suddivisioni e rinvii, per le pagine web si può verificare se il documento è lineare o ipertestuale, se ci sono degli indici, se i link interni sono coerenti e utili;
- *impaginazione e manifattura*: certamente la qualità della stampa non trova un corrispettivo in ambiente elettronico, ma è importante, ad esempio, che il sito sia visualizzabile con qualsiasi browser, che sia leggibile e graficamente equilibrato. Ad esempio, costituisce una garanzia l'URL stabile ed esplicitato nelle pagine;
- *posizionamento*: anche le pagine web acquisiscono credito se citate o indicizzate in altre fonti, o se superano positivamente il confronto con altre simili.



### BREVE STORIA DELL'AUTOMAZIONE NELLE BIBLIOTECHE ITALIANE

Tecnologia informatica e automazione delle procedure sono concetti che iniziarono a circolare nelle biblioteche italiane intorno alla fine degli anni Sessanta del secolo scorso, quando videro la luce le prime realizzazioni in tal senso. Il punto di svolta del processo si può fissare in seguito a un evento tragico della storia italiana: l'alluvione di Firenze del 1966, con i gravissimi danni occorsi al patrimonio della Biblioteca nazionale centrale. Nell'ottobre 1968, infatti, si svolse a Firenze un seminario nel corso del quale vennero comunicati i risultati di due studi che erano stati condotti, nel quadro degli aiuti dell'*American Library Association* alle biblioteche italiane, da John Finzi della *Library of Congress* e Joseph Becker dell'EDUCOM, circa la possibile riorganizzazione dei servizi della Biblioteca nazionale centrale di Firenze. Il convegno si concluse con una serie di risoluzioni, la prima delle quali prevedeva l'avvio di uno studio sull'informatizzazione della Bibliografia Nazionale Italiana, con l'adozione del **MARC** e l'avvio del programma **ANNA** (Automazione Nella NAzionale). A stretto giro di tempo venne avviato anche il progetto *Biblio*, con lo scopo di automatizzare le procedure della Biblioteca nazionale centrale di Roma.

In quello stesso periodo nacquero anche i consorzi universitari, in particolare per il supporto al supercalcolo, che negli anni successivi avrebbero poi avuto una parte rilevante nell'automazione delle biblioteche, specialmente di quelle universitarie.

Se l'automazione della BNI, per un insieme di motivi connaturati al sistema bibliotecario italiano nel suo complesso (come la mancanza di una vera cooperazione catalografica e la tendenza a considerare la BNI come una fonte autorevole per il raffronto, ma non per una vera e propria catalogazione derivata, anche se effettuata manualmente), rappresentò un momento importante nella storia dell'automazione, seppure con scarse conseguenze pratiche, gli anni Settanta furono caratterizzati dal fiorire di sperimentazioni in ambienti quali i centri di documentazione di importanti istituti di ricerca (CNR, CNEN o ISS).

In tutte le più importanti e significative esperienze (fatte salve alcune sperimentazioni in Lombardia e Piemonte, volte a favorire la cooperazione territoriale) l'attenzione fu posta soprattutto sulla razionalizzazione delle procedure gestionali interne, come l'acquisizione e la catalogazione, e in misura decisamente inferiore sull'erogazione di servizi all'utenza, come ad esempio la circolazione dei documenti. Questa scelta di principio ebbe come conseguenza, anche a causa di precise opzioni progettuali e di finanziamenti sempre troppo scarsi, il costante rinvio dell'automazione dei servizi al pubblico.

Bisognò attendere almeno fino alla fine degli anni Settanta perché venisse alla luce un progetto di cooperazione a livello nazionale mirato alla costruzione di un servizio di prestito interbibliotecario e denominato **SNADOC** (Servizio Nazionale di Accesso ai DOcumenti), che fu alla base della nascita del **Servizio bibliotecario nazionale (SBN)**. Il progetto, imperniato programmaticamente sull'idea di creare un "servizio" cooperativo, non si poneva più l'obiettivo di razionalizzare le procedure interne di gestione, bensì lo scopo, assai più ambizioso, di creare «un nuovo disegno complessivo della realtà bibliotecaria italiana», come ebbe a sottolineare Luigi Crocetti.

Il disegno della rete nazionale fu elaborato tra il 1980 e il 1985 con lo scopo di creare **due servizi**: uno di accesso al documento, attraverso la realizzazione di un catalogo comune costituito dalle risorse bibliografiche delle maggiori biblioteche italiane; l'altro inteso come servizio nazionale di circolazione dei documenti. Entrambi i servizi furono resi possibili grazie alla presenza di una macchina denominata **Indice**, contenente le informazioni minime atte a identificare una notizia bibliografica, unitamente a informazioni circa la localizzazione e il posseduto. Nel disegno originale l'Indice aveva il compito di indirizzare le richieste di servizi (prestito o prelievo del record bibliografico) verso la base dati locale contenente la descrizione completa del documento, o verso la biblioteca in possesso del documento richiesto. Le diverse biblioteche, inoltre, erano parte di sistemi locali denominati **Poli**, dotati degli opportuni strumenti per essere collegati al sistema centrale. Tale collegamento era garantito dal fatto che i vari sistemi locali utilizzavano le stesse procedure, anche se sviluppate su diverse piattaforme hardware.

Nel 1985, a seguito di uno studio di fattibilità, furono prese decisioni che rovesciarono l'impostazione precedente: l'Indice, da smistatore di richieste, divenne un vero e proprio catalogo costituito da record bibliografici completi, creando una serie di problematiche come la duplicazione di record e la necessità di procedure di allineamento tra la base dati centrale e quelle periferiche.

Il sistema così delineato vide la luce nel 1992, quando le due Biblioteche nazionali centrali di Roma e Firenze attivarono il collegamento con l'elaboratore Indice. Dal momento dell'ideazione teorica a quello della messa in servizio passarono però dodici anni, nel corso dei quali i cambiamenti tecnologici furono talmente radicali che fin dal suo primo apparire l'architettura informatica si rivelò già datata.

Le scelte tecnologiche adottate determinarono una sostanziale chiusura del sistema a livello di struttura del dato bibliografico, di hardware e protocolli di comunicazione e di colloquio Indice-Poli; inoltre, il fatto che i primi e più attivi Poli, grazie ad una convenzione con lo Stato, potessero cedere, ad altre regioni che ne avessero fatto richiesta, gli applicativi che erano stati realizzati *ad hoc*, ebbe come conseguenza:

- l'assenza di un vero regime di concorrenza, che impedì il corretto sviluppo del mercato nazionale dei software per biblioteche;
- l'impossibilità di discernere tra le politiche istituzionali dell'ente pubblico e quelle commerciali dell'azienda collegata, impegnata a sviluppare applicativi SBN su incarico di Stato e Regioni;
- l'assenza di ricerca e sviluppo nel settore, limitando il ruolo di molte aziende italiane a quello di semplice integrazione di lacune o difetti degli applicativi SBN.

In pratica SBN condizionò fortemente il mercato dell'automazione in Italia, creando una dicotomia tra le biblioteche SBN e quelle non aderenti. Fino agli anni Novanta sono coesistiti (e in certe realtà minori italiane coesistono ancora oggi) prodotti diversi per l'automazione dei vari processi di gestione delle biblioteche. Software diffusi in ambito internazionale o sviluppati in Italia hanno trovato una nicchia di mercato nell'offrire alle biblioteche determinati strumenti che consentissero di fornire anche servizi diversi rispetto a quelli SBN, nel cui ambiente tradizionalmente è stata prestata molta attenzione soprattutto alle funzioni di catalogazione. Inoltre il fattore economico ha indubbiamente condizionato questo settore: a causa delle poche risorse economiche a disposizione delle biblioteche gli sviluppi sono sempre stati limitati, il mercato è rimasto statico e la diffusione di prodotti avanzati (quindi costosi) si è ridotta a poche realtà.

Vista la natura centralizzata di SBN, che prevede la gestione dell'architettura informatica su grandi elaboratori (*mainframe*), in contrapposizione alla diffusione sempre maggiore di personal computer, si è stabilita nel corso degli anni una dicotomia tra le grandi biblioteche, che hanno scelto di aderire a SBN, e le piccole biblioteche, che invece hanno scelto uno dei pochi software per microelaboratore più diffusi in Italia negli anni Ottanta e Novanta (Tinlib, CDS/Isis, Dobis/libis, Aleph, Sebina).

Mentre nel resto d'Europa e, soprattutto, negli Stati Uniti già negli anni Ottanta si diffondevano diversi applicativi in grado di gestire in maniera integrata le varie procedure bibliotecarie (ILS - *Integrated Library System*), in Italia furono solo le biblioteche universitarie ad adottare questo tipo di prodotto (Aleph in particolare), prima di tutto perché avevano a disposizione maggiori risorse, in secondo luogo perché erano entrate in SBN più tardi rispetto agli enti locali, così da essere meno vincolate a quella realtà (il protocollo firmato anche dal Ministero dell'Università, oltre che dagli enti fondatori, ossia Beni culturali e Coordinamento delle Regioni, risale al 1994).

Le altre tipologie di biblioteche optarono per l'automazione di singole procedure (catalogazione, se non facevano parte di SBN, e prestito; negli anni Novanta anche l'OPAC). La spinta alla cooperazione e la conseguente creazione di sistemi bibliotecari, insieme alla diffusione di Internet, furono i fattori che indubbiamente favorirono la condivisione dei dati e delle procedure, oltre a consentire una contrattazione migliore nei confronti dei produttori/distributori di software in termini sia economici, sia di personalizzazione e assistenza sui prodotti.

Negli ultimi trent'anni l'evoluzione dei software per biblioteche è andata di pari passo con le nuove potenzialità offerte dall'informatica; in particolare si possono individuare tre stadi di evoluzione:

- a) efficienza dei processi interni;
- b) accesso alla collezione locale;
- c) accesso a risorse selezionate, anche al di fuori del posseduto della biblioteca.

Considerando tale schematizzazione nella sua prospettiva funzionale, si può dire che a questi tre stadi sia corrisposta man mano una maggiore attenzione all'utente finale e all'accesso, più che al possesso (e relativa gestione).

Per quanto riguarda l'architettura, all'inizio i programmi furono concepiti per un uso sostanzialmente locale, ovvero in modalità **stand alone**, in virtù della quale tutto (database e applicativo) era contenuto nel computer utilizzato dall'operatore, dopodiché si è passati ad una modalità **client/server**, con notevole miglioramento delle procedure e maggiori possibilità di condivisione. Alcuni applicativi sono stati resi disponibili grazie ad emulatori di terminale (ad esempio, *Putty*), programmi che consentono di lanciare applicazioni residenti sul server e lavorare direttamente «come se» si stesse operando su di esso. In quella fase i programmi sviluppati avevano interfacce esclusivamente testuali.

UTENTE: IB001	FUNZIONE: BIBA01	DATA: 20/07/2007	ORA: 14:54
-----			
S B N - Biblioteca dell'Associazione Italiana Biblioteche			
*****	COLLOCAZIONE DOCUMENTO	-	TITOLO LEGATO ALL'INVENTARIO *****
Numero d'inventario : 000015326			
MOD1385952 (M) *Biblioteconomia : guida classificata / diretta da Mauro Gu errini ; condirettore Gianfranco Crupi ; a cura di Stefano Gambari ; collabo razione di Vincenzo Fugaldi ; presentazione di Luigi Crocetti. - Milano : Bi bliografica, [2007?]. - XL, 11			
Precisazione del documento : <input type="text"/>			
USCITA : <input type="checkbox"/>			
per continuare premere ENTER			

Figura 10: Schermata del collegamento a SBN tramite Putty

Infine, a seguito dell'avvento di Internet, del World Wide Web e delle interfacce grafiche, nonché dei sistemi di database imperniati sul modello relazionale, si sono diffusi i **software web based**, nei quali l'applicazione è accessibile via web attraverso una connessione di rete.

## SBN

Il **Servizio bibliotecario nazionale** (SBN) è la rete delle biblioteche italiane promossa dal Ministero per i beni e le attività culturali con la cooperazione delle Regioni e dell'Università, coordinata dall'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU). Attualmente vi aderiscono **oltre 4.600 biblioteche** statali, di enti locali, universitarie, scolastiche, di accademie e istituzioni pubbliche e private operanti in diversi settori disciplinari. Le biblioteche partecipanti sono raggruppate in **74 Poli locali**, costituiti da un insieme più o meno numeroso di biblioteche che gestiscono tutti i propri servizi con procedure automatizzate. I Poli sono a loro volta collegati all'**Indice SBN**, che contiene il catalogo collettivo delle pubblicazioni acquisite dalle biblioteche aderenti al Servizio bibliotecario nazionale.

Come illustrato anche in precedenza, il sistema si basa sulla **catalogazione partecipata**: il documento viene catalogato solo dalla prima biblioteca che lo acquisisce, mentre le altre si limitano a catturare la descrizione bibliografica già presente nell'Indice, aggiungendo la propria localizzazione. La base dati dell'Indice SBN comprende attualmente materiale antico, moderno, musica, grafica e cartografia.

### **Come aderire al Servizio bibliotecario nazionale**

Possono aderire al Servizio bibliotecario nazionale tutte le biblioteche italiane, sia pubbliche che private, che seguono gli standard nazionali e internazionali per la catalogazione e fornitura dei servizi. La biblioteca che intende aderire alla rete SBN può scegliere tra due diverse modalità: collegarsi ad un Polo già esistente oppure creare un nuovo Polo. I costi complessivi dell'entrata in rete da parte della biblioteca sono ovviamente in relazione alla modalità che intende adottare.

#### **• Collegarsi ad un Polo già esistente**

La biblioteca o le biblioteche interessate all'adesione devono contattare direttamente i responsabili del Polo prescelto, al fine di concordare le modalità da seguire e stipulare una convenzione che regoli gli impegni reciproci. All'ICCU viene inviata comunicazione delle nuove adesioni.

#### **• Creare un nuovo Polo**

La biblioteca o le biblioteche interessate devono prendere anzitutto contatti con la Regione, che esprime un parere in merito alla costituzione del nuovo Polo. L'ICCU, che è delegato dal Ministero per i beni e le attività culturali (MiBAC) al coordinamento della rete SBN, può fornire supporto tecnico-scientifico al costituendo Polo.

Il soggetto che propone la costituzione del nuovo Polo predispone un'apposita convenzione con il MiBAC, che sarà firmata dal soggetto proponente e dal Direttore generale per le biblioteche, gli istituti culturali e il diritto d'autore, dopo che il Comitato tecnico-scientifico SBN avrà approvato la costituzione del nuovo Polo. La convenzione è accompagnata da un piano di progetto, nel quale sono precisate le modalità di adesione e di cooperazione che il nuovo Polo intende attuare nell'ambito del Servizio bibliotecario nazionale.

Il Polo costituendo deve comunicare il nome del referente di Polo, che si impegna a diffondere a tutte le biblioteche di Polo le informazioni ricevute dall'ICCU circa le attività dell'Indice e della rete.

#### **• Impegni cooperativi e normative SBN**

Il Ministero per i beni e le attività culturali promuove la diffusione di SBN d'intesa con le Regioni e con il Ministero dell'università e della ricerca scientifica e tecnologica.

I Poli che aderiscono al SBN sono tenuti a sottoscrivere con il MiBAC un'apposita convenzione, impegnandosi a cooperare sia per incrementare il catalogo collettivo SBN, sia per rendere disponibili i documenti posseduti. A tal fine ogni partner s'impegna ad osservare norme catalografiche comuni e a rendere disponibili i propri documenti, in originale o in copia, agli altri partner che ne facciano richiesta; si impegna, inoltre, a sostenere i costi di installazione del software ed eventualmente di acquisizione dell'hardware necessario, nonché i costi di collegamento con l'Indice.

Il MiBAC, a sua volta, si impegna a provvedere ai costi di manutenzione e di sviluppo del Sistema Indice sul quale è ospitato il catalogo collettivo della cooperazione.

Attualmente, in seguito all'evoluzione dell'Indice e all'apertura a sistemi di gestione non SBN, l'Istituto sta provvedendo alla certificazione di applicativi commerciali che intendono sperimentare il colloquio con l'Indice tramite il protocollo SBNMARC. In tal modo sarà possibile potenziare la rete SBN allargando l'adesione ad istituzioni

che utilizzano software differenti, così da consentire una più ampia ed efficace distribuzione dei servizi offerti dal sistema centrale.

Alle nuove biblioteche che entrano in rete l'ICCU e i partner (Regioni, Poli locali, università etc.) garantiscono un programma di formazione che può essere articolato su vari livelli, in relazione alle necessità delle singole biblioteche.

L'adesione al SBN è disciplinata da accordi, convenzioni e protocolli scaricabili dal sito dell'ICCU a partire dalla pagina web <http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/aderire/>.

L'OPAC consente l'accesso a 11.663.696 notizie bibliografiche, corredate da 57.563.318 localizzazioni (aggiornato al 4 settembre 2011). Ad esse sono collegati l'Anagrafe delle biblioteche (che consente di visualizzare informazioni sui recapiti e sui servizi della singola biblioteca) e l'OPAC locale, tramite il quale l'utente può recuperare maggiori informazioni sul documento (collocazione, disponibilità etc.). Sempre tramite l'OPAC SBN è possibile accedere a servizi quali la richiesta di fornitura di riproduzioni o il prestito interbibliotecario nell'ambito del servizio nazionale ILL SBN.

A partire dal 2002 è stato avviato il progetto **Evoluzione dell'Indice SBN**, con lo scopo di colmare il ritardo tecnologico su cui si basa l'architettura della rete nazionale, tramite il perseguimento dei seguenti obiettivi:

- rinnovamento tecnologico dell'hardware e del software con passaggio su piattaforma aperta e utilizzo di protocolli standard;
- razionalizzazione, integrazione e ristrutturazione delle basi dati (diventate nel frattempo tre e separate tra loro: *Libro moderno, Libro antico, Musica*);
- apertura dell'Indice SBN a sistemi di gestione della biblioteca non SBN che utilizzino i più diffusi formati bibliografici, mediante la realizzazione di un'interfaccia standard che consenta a sviluppatori di software di dotare altri applicativi del colloquio con l'Indice, grazie all'adozione del nuovo protocollo appositamente sviluppato e denominato **SBN-MARC**;
- gestione di livelli di partecipazione al SBN diversificati (solo cattura dei dati, localizzazione del proprio posseduto, inserimento di nuove catalogazioni, correzioni);
- sviluppo di nuove funzionalità, come la *catalogazione derivata* (possibilità di importare dati bibliografici da altre basi dati) e la *catalogazione a blocchi* (riversamento in Indice di archivi bibliografici precedentemente costituiti, con confronto dei dati).

La prima parte del progetto si è conclusa nel febbraio 2004, con la messa in produzione del nuovo **Indice2**. Da quel momento è partito il processo di **certificazione della conformità** al protocollo SBN MARC dei prodotti di diverse *software house*, certificazione resasi necessaria affinché tali applicativi possano dialogare con l'Indice. La certificazione attesta il corretto scambio di messaggi tra l'applicativo certificato e l'Indice SBN, senza costituire certificazione di qualità del prodotto.

I **livelli di dialogo** disponibili sono i seguenti:

- **livello 2**, comprendente le sole funzioni di ricerca e localizzazione per possesso;
- **livello 3**, che comprende, oltre alle precedenti, anche le funzioni di catalogazione e correzione dei record non condivisi;
- **livello 4**, comprendente le funzioni di ricerca, localizzazione per possesso e per gestione, catalogazione, correzione anche dei record condivisi e allineamento.

Ad agosto 2011 gli applicativi che hanno ottenuto la certificazione di conformità al Protocollo SBN-MARC risultavano essere quindici.

L'elenco dei software certificati è disponibile sul sito dell'ICCU alla pagina [http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/certif\\_conf\\_protoc\\_sbnmarc/pagina\\_345.html](http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/certif_conf_protoc_sbnmarc/pagina_345.html).

## L'AUTOMAZIONE DELLE BIBLIOTECHE ITALIANE OGGI

Pur non esistendo dati numerici ufficiali, si può empiricamente stimare che la maggior parte delle strutture bibliotecarie esistenti e attive sia oggi dotata di computer, connessione a Internet e procedure automatizzate; nella maggior parte dei casi aderisce anche a una rete o ad un sistema bibliotecario.

Da alcuni anni a questa parte nel *Rapporto sulle biblioteche italiane* edito dall'AIB (Associazione italiana biblioteche) viene pubblicato un capitolo dedicato alla diffusione dei sistemi di automazione, a cura di Vanni Bertini. Secondo l'edizione 2009-2010 sarebbero oltre 10.400 le biblioteche in cui risultano installati i software di automazione delle procedure specificati nella tabella che segue.

TOTALE INSTALLAZIONI - ITALIA						
	2009		2007		Variazione 2007-2009	
		biblioteche		biblioteche		biblioteche
Aleph	62	1.011	62	1.011	0	0
Amicus/Opensuite	24	276	19	226	5	50
Aquabrowser	13	59	8	28	5	31
Biblionauta	75	760	67	900	8	-140
Bibliowin4.0	603	824	344	552	259	272
Bibliowin5.0	109	145	43	64	66	81
Bibliowin OPAC	235	669	197	514	38	155
EOS.Web	18	107	19	48	-1	59
ErasmNet	14	249	9	138	5	111
Kentika/Alexandrie	37	37	46	50	-9	-13
Libero	7	140	7	63	0	77
Q-Series	4	230	4	213	0	17
Sebina	470	4.982	419	4.089	51	893
So.Se.Bi. TLM 3	122	297	110	260	20	143
So.Se.Bi. TLM 4	8	106				
So.Se.Bi. TLM OPAC	126	448	104	159	-80	26
Tinlib	25	25	41	40	-16	-15
UnibiblioX	47	76	15	44	32	32
Voyager	0	0	0	0	0	0
Vubis Smart	3	n.d.	3	n.d.	0	n.d.

A livello tecnologico attualmente la maggior parte degli applicativi esistenti è di tipo integrato: con un programma unico si gestiscono tutte le procedure occorrenti (acquisizione, catalogazione, prestito, OPAC), condividendo i dati necessari a livello centrale. Grazie a questa struttura è possibile ottimizzare il lavoro di *back-office* e offrire servizi migliori agli utenti: l'OPAC consente di visualizzare se un documento è in prestito o meno, l'utente può accedere alla propria area riservata gestendo rinnovi e richieste, il bibliotecario può usufruire della catalogazione derivata e condivisa etc.

Probabilmente lo sviluppo più importante degli anni Duemila che ha riguardato gli ILS si riferisce alla declinazione del concetto di apertura tramite il rispetto di standard (i più importanti dei quali sono XML, DOI, OpenURL, MARC, Dublincore) che consentono l'interoperabilità e lo scambio di dati, così come l'implementazione di vari protocolli (Z39.50, ILL, OAI-PMH) ha migliorato i servizi offerti agli utenti quanto ad accesso all'informazione. Le evoluzioni recenti più avanzate concernono l'integrazione dei sistemi di automazione e dei dati con altre fonti (metadati di oggetti digitali, dati archivistici e museali etc.) e con altri applicativi nati in ambienti diversi da quelli più strettamente bibliotecari (ad esempio, quelli per la gestione di archivi aperti). Indubbiamente la maggior parte del lavoro di ricerca e

sviluppo circa l'automazione delle biblioteche ha riguardato l'OPAC, puntando sulla ricerca estesa ad altre informazioni (banche dati, periodici elettronici, informazioni editoriali che arricchiscono il record bibliografico etc.) e sulle modalità con cui tale ricerca viene svolta e presentata (emulando siti commerciali di successo come Google o Amazon, fornendo la possibilità all'utente di interagire il più possibile commentando e condividendo le informazioni secondo i principi di quella che è stata definita la *Library 2.0*).

Da notare, infine, che permangono nel nostro Paese diverse installazioni di software di vecchia generazione, soprattutto in piccole biblioteche dove gli investimenti economici sono ridotti, strutture in cui solo alcuni processi sono stati automatizzati o in cui ancora non si è passati ad un sistema integrato, utilizzando quindi applicativi diversi per gestire i vari processi.

Di seguito si presentano le caratteristiche salienti e alcuni dati informativi dei principali software diffusi in Italia.

#### • **ALEPH (*Automated Library Expandable Program*)**

Sviluppato alla fine degli anni Settanta presso la *Hebrew University* di Gerusalemme, costituisce ancora oggi uno dei prodotti più diffusi nel mondo. Distribuito da *Ex Libris*, in Italia è il prodotto leader nel campo dei grandi sistemi esterni al mondo SBN, soprattutto universitari. Sempre in Italia è attiva sin dal 1993 l'Associazione utenti ALEPH (ITALE).

Per la gestione dei moduli di cui si compone (Catalogazione/Copie, Acquisizione/Periodici, Circolazione, ILL) ha un'interfaccia grafica (GUI - *Graphical User Interface*) in modalità *client/server*, mentre l'OPAC è gestito con un'interfaccia web. Basato sui principali standard di settore, ALEPH consente integrazione di risorse, ricerche e derivazioni in remoto, estesa connettività con sistemi e database esterni. I contenuti e le interfacce possono essere gestiti in 20 alfabeti diversi grazie alla completa implementazione di Unicode.

*Ex Libris* distribuisce anche altri prodotti per la ricerca, tra cui spicca **Primo**, un *discovery tool* che permette agli utenti di effettuare ricerche su un'ampia gamma di contenitori diversi di dati tramite un'unica interfaccia, in quanto il prodotto include programmi per l'*harvesting* di tutte le risorse della biblioteca. È interoperabile con i sistemi bibliotecari già esistenti e, nel contempo, indipendente da essi.

Fra gli strumenti di gestione delle informazioni si distribuiscono **SFX**, un *link resolver* OpenURL che offre agli utenti link contestuali al *full text* degli articoli e ad altre risorse definite dalla biblioteca, e **MetaLib**, un *discovery tool* per la ricerca di risorse informative remote da molteplici fornitori.

#### • **Biblionauta**

Sistema integrato prodotto da Nexus e composto da più moduli (acquistabili anche separatamente), sviluppato utilizzando il software di *information retrieval* CDS/ISIS dell'Unesco. È stato uno dei primi software impiegati dalle biblioteche pubbliche a produrre moduli *web based* per la catalogazione. La prima installazione di EasyWeb venne realizzata in collaborazione con il Sistema bibliotecario di ateneo dell'Istituto universitario di architettura di Venezia nel giugno del 1995. Easycat è in uso dal 2000 circa.

Si compone di: *Opac* (a sua volta composto dai moduli Easyweb e Metaesy, metacatalogo che permette di accedere in maniera integrata a database sia bibliografici che descrittivi di fondi di beni documentali, storico-artistici etc.); *Cat* (catalogazione partecipata e derivata, composto dal modulo Easycat e da Zefiro, applicazione specifica per la catalogazione derivata, utilizzabile in modo indipendente dal software applicativo di catalogazione in uso all'utente); *Fluxus* (per la gestione del prestito locale e a distanza); *Dialogo* (per la catalogazione in Indice SBN2); *Acquista* (per la gestione degli acquisti e periodici).

#### • **Bibliowin**

Attualmente disponibile in due versioni: una su software Windows, che necessita del modulo OPAC WEB per la ricerca on line; l'altra denominata Bibliowin 5.0, un sistema integrato che viene invece proposto come portale di biblioteche e archivi interamente basato su web e disponibile in versione ASP/SAAS.

Bibliowin 4 è stato sviluppato utilizzando Microsoft Access, il più diffuso strumento di database relazionale disponibile per Windows. L'interfaccia è di semplice e immediato utilizzo, grazie anche alla familiarità della presentazione. Il programma, studiato per rispondere alle esigenze di medie e piccole biblioteche civiche, funziona sia in modalità *stand alone*, sia in rete locale di PC (LAN), consentendo di accedere agli archivi da più postazioni contemporaneamente. Tra i moduli opzionali: acquisizioni, prestito interbibliotecario, acquisizione documenti elettronici (OCR), OPAC web, sistema di notifiche automatiche all'utente.

Bibliowin 5, che invece è interamente *web based*, si propone per la gestione di sistemi bibliotecari in catalogazione partecipata ed è integrabile con un CMS appositamente studiato per la gestione del sito web della biblioteca. L'OPAC può essere gestito in multilingua e consente l'implementazione di strumenti di metaricerca, nonché la personalizzazione per bambini e ragazzi.

#### • **Sebina**

Sviluppato nella seconda metà degli anni Ottanta su commissione della Soprintendenza ai Beni librari della Regione Emilia Romagna, è attualmente il sistema più diffuso in Italia, sia in ambito SBN (il 65% delle biblioteche SBN fa parte di Poli Sebina) che fuori. Caratteristica dei sistemi che adottano Sebina è quella di essere sistemi misti, pur con un predominio complessivo (circa il 60%) di biblioteche pubbliche. Prodotto e distribuito da *Data Management PA Solutions*, è arrivato alla sua quinta generazione con Sebina OpenLibrary, interamente *web based*.

Il progetto innovativo più recente è SebinaYOU, una piattaforma costruita per realizzare vari livelli di OPAC avanzato e di portale di servizi. In particolare: offre un sistema di ricerca per rilevanza e di ricerca federata; consente la partecipazione attiva dell'utenza (commenti, *rating*, liste di lettura condivise, *tagging* etc.); propone servizi simili a quelli di Amazon («*chi ha letto... ha letto anche...*», «*top ten*» etc.); permette l'interoperabilità con altri sistemi non strettamente bibliotecari. È dotato, altresì, di una suite di servizi dedicati ai ragazzi che comprende un Opac dall'interfaccia amichevole, con percorsi di ricerca guidati (realizzati grazie a una rete semantica) e una serie di servizi pensati *ad hoc*. Di recente sono state lanciate anche alcune applicazioni per *smartphone*, le prime in Italia in ambito bibliotecario.

#### • **So.Se.Bi.**

È uno dei produttori storici in Italia, diffuso soprattutto in Sardegna. Oltre ad una serie di moduli aggiuntivi per alcune attività specifiche (controllo postazioni, analisi del patrimonio), l'offerta di So.Se.Bi. si articola intorno ai due prodotti gestionali TLM3 e TLM4, nonché intorno ai moduli OPAC Tilipirke e WebOPAC.

TLM3 è pensato in particolare per le biblioteche di media o piccola dimensione; compatibile con Windows, grazie al modulo aggiuntivo Tilipirke permette la pubblicazione del catalogo su internet.

TLM 4, fornito sia in modalità tradizionale che SaaS, è stato uno dei primi software ad ottenere la certificazione SBN e per questo si propone come la piattaforma software di un Polo bibliotecario che voglia lavorare all'interno del Servizio nazionale.

Entrambe le versioni integrano il nuovo servizio «Catalogazione Rapida», studiato per l'inserimento del pregresso tramite catalogazione derivata di centinaia o migliaia di titoli in un'unica operazione.

#### • **Sbnweb**

È il nuovo applicativo realizzato dal Ministero per i beni e le attività culturali nell'ambito del progetto di evoluzione di SBN UNIX in architettura *client/server*. Sviluppato utilizzando software *open source*, è fornito in versione *web based* in uso gratuito alle biblioteche che intendano adottarlo (la manutenzione generale dell'applicativo è a carico dell'ICCU, mentre sono a carico del Polo le spese di gestione locale). Oltre ai moduli acquisti, catalogazione, gestione del patrimonio e OPAC, con questo nuovo applicativo vengono gestiti anche i servizi all'utenza, compresa l'integrazione con il nuovo servizio di prestito interbibliotecario ILL SBN.



### Software *open source*

Per poter essere definito «libero» un software deve garantire quattro libertà fondamentali:

- libertà di eseguire il programma per qualsiasi scopo (libertà 0);
- libertà di studiare il programma e modificarlo (libertà 1);
- libertà di copiare il programma in modo da aiutare il prossimo (libertà 2);
- libertà di migliorare il programma e di distribuirne pubblicamente i miglioramenti, in modo tale che tutta la comunità ne tragga beneficio (libertà 3).

Attualmente i prodotti *open source* per la gestione di biblioteche in Italia sono diffusi solo in pochissime realtà, anche perché non si è mai costituito un movimento forte a favore della promozione di questa filosofia. Tuttavia, è possibile individuare, nella storia dell'automazione bibliotecaria italiana, prodotti diffusi e utilizzati senza avere alle spalle distributori commerciali, nonché tentativi di elaborare prodotti *open source* "nativi" o di rilasciare i sorgenti di software inizialmente proprietari.

Tra questi va citato in particolare **CDS/ISIS**, sviluppato e distribuito gratuitamente dall'UNESCO (con la collaborazione di molti altri enti) a partire dal 1985 e distribuito in Italia dall'Associazione DBA. Nel nostro Paese CDS/ISIS è stato utilizzato da un consistente numero di biblioteche, soprattutto pubbliche, grazie anche alla distribuzione di applicazioni come ISISMarc (un'interfaccia di inserimento dati in MARC), Teca (per la catalogazione di materiale bibliografico e non, secondo le norme ISBD), EDAN [per il materiale bibliografico antico secondo ISBD(A)] e BIBLO (per la gestione bibliografica, amministrativo-contabile e del prestito, sviluppata e utilizzata dal Sistema bibliotecario urbano di Venezia). Anche Winiride, software per la gestione delle biblioteche scolastiche, si basa su CDS/ISIS.

Da un'indagine condotta nel 2009 è emerso che in Italia sono circa una ventina gli istituti bibliotecari che si sono dotati di gestionali *open source* predisposti appunto per le biblioteche (in particolare *PMB - Pour Ma Bibliothèque* e *Koha*). Si tratta per la maggior parte di biblioteche scolastiche o di enti privati, che nel contesto italiano costituiscono certamente l'anello più debole di un sistema bibliotecario sempre più in difficoltà a causa dei tagli ai bilanci. Un caso interessante (successivo all'indagine) è quello della Biblioteca nazionale Braidense di Milano, che ha adottato un software *open source* (VuFind) per l'OPAC.

Tradizionalmente nei software *open source* i bibliotecari e gli amministratori di sistema vedono svantaggi come costi maggiori per consulenze tecniche a livello locale, assenza di garanzie su supporto e *upgrade*, rapporti diretti con informatici, sviluppatori, e dipendenza da essi. In realtà anche per i software *open source* è ormai diffuso un *business model* che garantisce l'assistenza nell'installazione e nello sviluppo, pur mantenendo il prodotto libero.

## BIBLIOGRAFIA

- BERTINI Vanni (2010). *I sistemi di automazione*. In *Rapporto sulle biblioteche italiane 2009-2010* / a cura di Vittorio Ponzani; direzione scientifica di Giovanni Solimine. Roma: Associazione italiana biblioteche, p.158-171.
- BREEDING Marshall (2007). *Next-Generation Library Catalogs*. "Library Technology Reports" v. 43 n. 4.
- CASTANO, S., MONTANELLI, S. FERRARA, A. (2009). *Informazione, conoscenza e Web per le scienze umanistiche*. [Milano]: Pearson addison Wesley.
- CIOTTI, Fabio, RONCAGLIA, Gino. (2010). *Il mondo digitale : introduzione ai nuovi media*. Roma: Laterza.
- GUERRINI, Mauro. (2010). *Gli archivi istituzionali: open access, valutazione della ricerca e diritto d'autore*. Milano: Bibliografica.
- KAPLAN Michael (2009). *Library automation*. In *Springer handbook of automation* / Nof Shimon Y. New York: Springer, p. 1285-1298.
- LEOMBRONI Claudio (2003). *Una vicenda controversa: l'automazione delle biblioteche in Italia*. In *La storia delle biblioteche: temi, esperienze di ricerca, problemi storiografici*. Roma: Associazione italiana biblioteche, p. 167-196.
- LEOMBRONI Claudio (2008a). *Servizio Bibliotecario nazionale (SBN)*. In *Biblioteconomia. Guida classificata* / diretta da Mauro Guerrini; condirettore Gianfranco Crupi; a cura di Stefano Gambari; collaborazione di Vincenzo Fugaldi; presentazione di Luigi Crocetti. Milano: Editrice Bibliografica, p. 224-238.
- LEOMBRONI Claudio (2008b). *Automazione delle biblioteche*. In *Biblioteconomia. Guida classificata* / diretta da Mauro Guerrini; condirettore Gianfranco Crupi; a cura di Stefano Gambari; collaborazione di Vincenzo Fugaldi; presentazione di Luigi Crocetti. Milano: Editrice Bibliografica, p. 248-268.
- METITIERI, Fabio. RIDI, Riccardo. (2006). *Biblioteche in rete : istruzioni per l'uso*. Roma: Laterza.
- RIDI Riccardo (2007). *La biblioteca come ipertesto*. Milano: Editrice bibliografica.
- RIDI Riccardo (2010). *Strumenti e strategie per la ricerca di informazioni WWW* (Versione 3.11). <<http://www.burioni.it/forum/ridi-mot.htm>>
- RONCAGLIA, Gino (2010). *La quarta rivoluzione: sei lezioni sul futuro del libro*. Roma: Laterza.
- SCOLARI, Antonio. (2000). *UNIMARC*. Roma: Associazione italiana biblioteche.
- TOMASI, Francesca. (2009). *Metodologie informatiche e discipline umanistiche*. Roma: Carocci.
- WESTON Paul Gabriele (2007). *La gestione elettronica delle biblioteche*. In *Biblioteconomia: principi e questioni* / a cura di Giovanni Solimine e Paul Gabriele Weston. Roma: Carocci, p. 221-256.

## INDICE DELLE IMMAGINI

Figura 1: Rappresentazione schematica di entità, attributi e relazioni	5
Figura 2: Rapporto intenzionalità/completezza delle risorse informative	7
Figura 3: Schema di funzionamento di un ILS	22
Figura 4: Schermata del collegamento a SBN tramite SBN web	23
Figura 5: La nuova maschera di ricerca di SBN	26
Figura 6: Una scheda bibliografica del portale Scoprirete della Rete bibliotecaria di Romagna e San Marino	27
Figura 7: Rappresentazione grafica del funzionamento del protocollo OpenURL	28
Figura 8: Rappresentazione grafica dell'architettura OAI	32
Figura 9: Un'immagine del lettore di e-book "Kindle" della Amazon	37
Figura 10: Schermata del collegamento a SBN tramite Putty	43

## INDICE ANALITICO

<b>A</b>	DTD .....16
AIB ..... <i>Vedi</i> Associazione Italiana Biblioteche	Dublin Core .....11; 12
ALEPH .....47	<b>E</b>
AND ..... <i>Vedi</i> Operatori Booleani	e-book .....36; 37; 38
ANNA	E-BOOK.....35; 36
MARC .....41	encoding .....5
Associazione italiana biblioteche.....9; 46	entità
attributo	database .....4
database.....4	ePub .....37
<b>B</b>	<b>estensione</b> ..... <i>Vedi</i> file
Biblionauta .....47	Eureka .....39
Bibliowin .....48	<b>F</b>
<b>C</b>	file .....4
Citebase .....38	FRBR <i>Vedi</i> Functional Requirements for Bibliographic Records
conoscenza .....3	Functional Requirements for Bibliographic Records.....4
<b>D</b>	<b>G</b>
database .....3	Google Scholar .....38
dato .....3	<b>I</b>
DBMS ..... <i>Vedi</i> database	ILS.....17; 20; 21; 22; 24; 42; 47
Digitalizzazione .....12; 28; 29; 35	

<i>Information Retrieval</i> .....	6
informazione .....	3
Internet.....	38
ISO 2709 .....	17

## K

<i>Koha</i> .....	49
-------------------	----

## L

Linguaggi di marcatura .....	<i>Vedi Markup</i>
------------------------------	--------------------

## M

<b>MARC</b> .....	11; 17; 41
markup .....	15
<b>MIME</b> .....	<i>Vedi Multipurpose Internet Mail Extensions</i>
<i>Multipurpose Internet Mail Extensions</i> .....	5

## N

NOT.....	<i>Vedi Operatori Booleani</i>
----------	--------------------------------

## O

OPAC.....	20; 21; 23; 24; 25; 26; 28; 42; 45; 46; 47; 48; 49
OpenURL.....	27; 28; 35; 47
Operatori Booleani .....	8
OR.....	<i>Vedi Operatori Booleani</i>

## Q

<i>query</i> .....	6
--------------------	---

## R

relazione	
database.....	4
rumore.....	<i>Vedi Information retrieval</i>

## S

SBN.....	22; 23; 25; 41; 42; 43; 44; 45; 46; 47; 48; 49
Sbnweb .....	49
Sebina .....	48
Segnaweb.....	39
Servizio bibliotecario nazionale .....	<i>Vedi SBN</i>
SFX .....	47
<b>SGML</b> .....	16
silenzio .....	<i>Vedi Information retrieval</i>
SNADOC .....	41
So.Se.Bi. ....	48
SRU.....	19

## T

Teoria dell'informazione .....	14
--------------------------------	----

## U

UNIMARC .....	17
---------------	----

## V

VuFind .....	49
--------------	----

## W

WEB. 3; 5; 9; 11; 15; 16; 22; 24; 25; 26; 32; 35; 36; 38; 39; 40; 43; 44; 47; 48; 49	
Whittaker .....	39
Winiride .....	49
Worldcat .....	26
WWW.....	<i>Vedi Web</i>

## X

XML .....	15
XOR .....	<i>Vedi Operatori Booleani</i>

## Z

<b>Z39.50</b> .....	19
---------------------	----