

# BIBLIOTECHE DIGITALI

## ALCUNI STANDARD DI DESCRIZIONE CATALOGRAFICA E DI DESCRIZIONE DEI DOCUMENTI



# Definizione di biblioteca digitale

- La biblioteca digitale è una raccolta di testi in formato elettronico, resa disponibile attraverso Internet e quindi consultabile da chi utilizza un computer e un collegamento alla rete. Il Web è l'elemento che dà senso a una raccolta digitale, consentendo di superare le barriere spaziali e temporali e di accedere all'informazione ricercata in tempi veloci e da qualsiasi luogo. Il concetto di biblioteca digitale è diverso da quello di "biblioteca elettronica" e "biblioteca virtuale", che spesso vengono impropriamente utilizzate come sinonimi di biblioteca digitale.
- La biblioteca elettronica definisce quell'istituzione che ha automatizzato i propri servizi (ad esempio, con l'introduzione del catalogo elettronico o con la strutturazione delle informazioni secondo un database).
- La biblioteca virtuale è una biblioteca che si estende a tutto il patrimonio raggiungibile in modo tale da fornire l'accesso a qualunque tipo di informazione sia essa contenuta all'interno dell'istituzione sia all'esterno.

# Il problema della conservazione dei documenti [1]

- Il problema della conservazione a lungo termine del patrimonio documentale è molto importante sia per le biblioteche tradizionali che per quelle digitali ma, mentre per le prime il concetto di conservazione si lega alla deperibilità dei materiali (libri, periodici, manoscritti documenti d'archivio etc.), per le seconde esso si articola su tre livelli: livello hardware, livello software e livello dei sistemi di codifica dei documenti. L' invecchiamento delle tecnologie informatiche è molto rapido e impone un periodico aggiornamento di qualsiasi sistema informativo e tale aggiornamento rende, a volte, inaccessibili le risorse informative generate mediante gli strumenti divenuti obsoleti.

# Il problema della conservazione dei documenti [2]

- Questo può portare ad una situazione in apparenza paradossale: i libri a stampa possono tranquillamente superare i cinquecento anni di vita mantenendo pressoché inalterata la loro condizione fisica mentre un documento elettronico rischia di divenire inutilizzabile nel giro di pochissimi anni; per questo la conservazione a lungo termine dei documenti digitali richiede l'adozione di sistemi standardizzati.

# Come si dividono gli Standard

Esistono standard di descrizione catalografica che sono gli standard tradizionali come:

- MARC
- Regole di catalogazione (ecco due esempi):
  - 1) La CDD (Classificazione Decimale Dewey)
  - 2) La CDU (Classificazione Decimale Universale)
- ISBD (International Standard Bibliographic Description)

Altri standard invece si occupano della descrizione dei documenti, i più importanti sono:

- SGML
- XML



# MARC - Definizione

- Il formato MARC (Machine Readable Catalog) fu sviluppato nel 1965 dalla Library of Congress come mezzo per l'interscambio di informazioni catalografiche tra biblioteche con l'utilizzo della tecnologia informatica. Nonostante la sua diffusione a livello internazionale, si è assistito allo sviluppo di differenti versioni di questo formato: USMARC (United States MARC), UKMARC (United Kingdom MARC), INTERMARC (MARC francese), ANNAMARC (MARC italiano avviato nel 1973 dalla Biblioteca Nazionale di Firenze, ora non più in uso), etc. Per questo motivo, negli ultimi dieci anni, grazie ad una cooperazione internazionale, è stato sviluppato UNIMARC (Universal MARC), un formato standard che mira all'unificazione bibliotecaria attraverso specifiche utilizzabili in ogni paese.

# Come funziona MARC

- Un record MARC è sostanzialmente una sequenza di campi individuati da codici numerici compresi tra 0 e 999, ognuno dei quali contiene le informazioni descrittive disponibili su una determinata opera. All'interno di ogni campo, inoltre, l'informazione viene strutturata in sottocampi che permettono di manipolare con sicurezza e affidabilità la registrazione. Questo tipo di struttura molto elastica ha consentito l'inclusione di nuovi tipi di informazione senza problemi: ad esempio, è stato recentemente aggiunto il nuovo campo MARC 856 (Electronic location and access) destinato a contenere le informazioni per accedere alle risorse remote.

# Classificazione Decimale Dewey - Definizione

- Lo schema di Classificazione decimale Dewey venne pubblicato per la prima volta nel 1876 dal bibliotecario americano Melvil Louis Kossuth Dewey con il titolo “A classification and subject index for cataloguing and arranging the books and pamphlets of a library”. Il successo ottenuto, nonostante le vivaci polemiche suscitate, spinse Dewey a pubblicare nel 1885 una seconda edizione dell'opera con il titolo “Decimal classification and relative index”, le cui pagine passavano dalle 44 della prima edizione alle 486. Tappa finale della lunga evoluzione della Classificazione Dewey è la ventesima edizione, datata 1988 e pubblicata in versione integrale anche in italiano.



# Com' è strutturata la Classificazione Decimale Dewey [1]

- La CDD si fonda essenzialmente su due elementi principali:
    - 1) la classificazione di Bacone (in *The advancement of learning*, 1608) dello scibile umano
    - 2) il sistema decimale
  - Secondo Bacone lo scibile umano dipende da tre facoltà fondamentali dell'uomo:
    - MEMORIA - Storia
    - IMMAGINAZIONE - Poesia/Arte
    - RAGIONE- Filosofia
- Dewey giunse all'adozione di uno schema baconiano ma invertito:
- RAGIONE - Filosofia
  - IMMAGINAZIONE - Poesia
  - MEMORIA - Storia
- 
- A questa suddivisione si possono ricondurre le nove **classi** che nella CDD rappresentano ciascuna un determinato ramo del sapere:
    - RAGIONE - Classi 1-6
    - IMMAGINAZIONE - Classi 7-8
    - MEMORIA - Classe 9

# Com' è strutturata la classificazione Decimale Dewey [2]

■ La "struttura portante" della CDD si fonda dunque sui seguenti grandi raggruppamenti:

- 0 GENERALITÀ
- 1 FILOSOFIE E DISCIPLINE CONNESSE
- 2 RELIGIONE
- 3 SCIENZE SOCIALI
- 4 LINGUAGGIO
- 5 SCIENZE PURE
- 6 TECNOLOGIA (SCIENZE APPLICATE)
- 7 LE ARTI
- 8 LETTERATURA
- 9 GEOGRAFIA E STORIA GENERALI



# Com' è strutturata la classificazione Decimale Dewey [3]

■ A questi raggruppamenti fu abbinata la **notazione decimale**, questa notazione è espressa da una sequenza di numeri considerati sempre come decimali, come fossero cioè preceduti da uno zero (0,), sfruttando le proprietà del sistema decimale ogni indice può essere diviso in altri dieci teoricamente all'infinito. Poiché ogni indice della CDD deve essere costituito da almeno tre cifre otterremo: 000 – 999.

000 GENERALITÀ  
100 FILOSOFIE E DISCIPLINE CONNESSE  
200 RELIGIONE  
300 SCIENZE SOCIALI  
400 LINGUAGGIO  
500 SCIENZE PURE  
600 TECNOLOGIA (SCIENZE APPLICATE)  
700 LE ARTI  
800 LETTERATURA  
900 GEOGRAFIA E STORIA GENERALI

# Com' è strutturata la classificazione Decimale Dewey [4]

- Ognuna di queste classi può essere suddivisa in dieci **divisioni** 010 – 090, ognuna di queste divisioni a sua volta può essere suddivisa in dieci **sezioni** 010 – 019, ogni sezione potrà essere (teoricamente) suddivisa all'infinito aggiungendo alle prime tre cifre, dopo un punto decimale, altri numeri decimali

## Esempio Divisione classe 000

010 BIBLIOGRAFIA  
020 BIBLIOTECONOMIA E SCIENZE DELL'INFORMAZIONE  
030 ENCICLOPEDIA GENERALI  
040 NON ASSEGNATO  
050 PUBBLICAZIONI GENERALI IN SERIE  
060 ORGANIZZAZIONI GENERALI E MUSEOLOGIA  
070 GIORNALISMO, EDITORIA, GIORNALI  
080 RACCOLTE GENERALI  
090 MANOSCRITTI E LIBRI RARI

## Esempio sezioni

010 BIBLIOGRAFIA  
011 BIBILOGRAFIE GENERALI  
012 PERSONALI (BIBLIOGRAFIE)  
013 DI SPECIFICHE CATEGORIE DI AUTORI  
014 DI OPERE ANONIME E PSEUDONIME  
015 PER LUOGO DI PUBBLICAZIONE  
016 BIBILOGRAFIE  
017 CATALOGHI GENERALI PER SOGGETTO  
018 CATALOGHI PER AUTORE E  
CRONOLOGICI  
019 CATALOGHI DIZIONARI

# Classificazione Decimale Universale - Definizione

■ La classificazione decimale universale vede la sua origine nello schema di Dewey e si è estesa nel tempo sino a rappresentare il sistema di classificazione più usato in vari paesi. Si tratta in sostanza di uno schema in cui agli aspetti tipici delle classificazioni gerarchiche si sono aggiunti una pluralità di elementi volti a potenziarne le caratteristiche di sintesi e a trasformarlo in un sistema più dinamico ed efficace.

# Com' è nata la Classificazione Decimale Universale?

- La CDU è nata dal proposito di Paul Otlet e di Henri La Fontane di costruire un archivio in grado di raccogliere le registrazioni catalografiche di tutte le pubblicazioni prodotte dall'uomo - libri, articoli, rapporti tecnici, ecc. - , a sostegno di tale progetto, Otlet e La Fontaine fondarono a Bruxelles l'Institute internationale de bibliographie con lo scopo di unificare i metodi catalografici e procedere alla creazione del repertorio universale, questo sogno s'infranse. Tuttavia il loro sforzo non fu vano, e produsse come risultato l'elaborazione di un nuovo schema di classificazione.
- La prima edizione della CDU apparve in Francia nel 1905 con il titolo 'Manuel du Repertoire bibliographique universel' , cui fece seguito la 'Classification decimale universelle', pubblicata tra il 1927 e il 1933; la terza edizione integrale fu quella tedesca, iniziata nel 1933 ma completata soltanto nel 1952.

# Com' è strutturata la Classificazione Decimale Universale? [1]

- Basandosi sulla quinta edizione della Classificazione decimale Dewey, Otlet e La Fontaine cominciarono dunque ad apportarvi quei cambiamenti che ritenevano funzionali ai loro obiettivi. In particolare si focalizzarono sul principio del “dividi come”, tale principio permetteva di adottare, all'interno di una data classe, un certo numero di suddivisioni standard, in grado di specificare tutti quegli aspetti che potevano essere presenti nei documenti, ma che sarebbe stato impossibile prevedere ed elencare esaustivamente nelle tavole dello schema: si trattava della prima applicazione di un principio di sintesi all'interno di una classificazione gerarchica, principio riconosciuto da Dewey, anche se limitato ad ambiti piuttosto ristretti. Rendendo esplicita l'indicazione di Dewey, i due belgi enuclearono dunque una serie di “suddivisioni comuni” (cioè categorie d'ordine generale applicabili a qualsiasi contesto) nettamente distinte dalle vere e proprie suddivisioni elencate nelle tavole; rese riconoscibili da segni di interpunzione o simboli matematici, le suddivisioni comuni consentivano di esprimere sinteticamente i molteplici aspetti di un documento, dando vita a notazioni più precise ed espressive.

Tra le principali suddivisioni comuni bisogna ricordare le suddivisioni:

- **di luogo**, segnalate dalle parentesi:  
347(450) è il diritto privato italiano
- **di lingua**, precedute dal segno =  
59=50 è un testo di zoologia in italiano
- **di tempo**, indicate dalle virgolette  
820"17" è la letteratura francese del '700
- **di razza**, tra parentesi con il segno =  
17(=924) è l'etica degli ebrei
- **di forma**, segnalate da uno 0 fra parentesi  
159.9(03) è un dizionario di psicologia

# Com' è strutturata la Classificazione Decimale Universale? [2]

- Per esprimere due o più concetti appartenenti a discipline diverse fu introdotta la "relazione", espressa graficamente dal segno dei due punti; un documento sui rapporti della religione con la letteratura, ad esempio, aveva come simbolo notazionale 82:2; uno sull'arte in relazione con l'etica era rappresentato da 7:17.
- Alla fine degli anni Cinquanta, Désiré Kervégant, per eliminare l'ambiguità altamente dannosa nell'espressione dei concetti, propose di rendere esplicita la natura delle relazioni, precisandone la funzione, il significato o la finalità; per questo elaborò delle tavole di categorie (appartenenza, processi, dipendenza, orientamento, comparazione, ecc.) a cui si accompagnava un sistema di frecce diversamente orientate e seguite da un numero d'ordine.
- Esempio  
gli antibiotici originati da funghi saranno indicati da  
576.8.097<-32<-582.28  
in quanto il codice ->32-> esprime l'origine: antibiotici (576.097) originati da (<-32<-)  
funghi (582.28)  
gli antibiotici contro i funghi saranno indicati da  
576.8.097->->212->582.28  
in quanto ->212-> indica un processo sfavorevole ("contro" i funghi)



# ISBD – International Standard Bibliographic Description

- ISBD è un modello per la descrizione catalografica, ha la sua origine nell'incontro internazionale di esperti di catalogazione tenuto a Copenaghen nel 1969. L'ISBD consiste essenzialmente in una struttura normalizzata che individua gli elementi (articolati in aree) per descrivere e identificare qualsiasi tipo di documento, assegna loro un ordine e stabilisce per essi un sistema di interpunzione convenzionale.

# Le Funzioni di ISBD

- Le ISBD, specificando elementi della descrizione bibliografica e prescrivendo l'ordine in cui questi elementi vanno presentati e la punteggiatura con cui essi dovrebbero essere scanditi, propongono:
  - (1) di rendere interscambiabili registrazioni provenienti da fonti diverse, in modo che le registrazioni prodotte in un paese possano essere accolte in cataloghi di biblioteche o altri elenchi bibliografici in ogni altro paese.
  - (2) di facilitare l'interpretazione delle registrazioni stesse al di là delle barriere linguistiche, in modo che registrazioni prodotte per gli utenti di una lingua possano essere interpretate dagli utenti di altre lingue.
  - (3) di favorire la conversione delle registrazioni bibliografiche in forma leggibile dalla macchina.
- Bisogna infine sottolineare come la descrizione ISBD rappresenti solo una parte della registrazione bibliografica completa e come, normalmente, non sia usata da sola. Gli altri elementi di cui si compone una registrazione bibliografica completa, come intestazioni, voci di soggetto, titoli uniformi, meccanismi di ordinamento e tracciato, non sono contemplati dall'ISBD. Le regole per tali elementi normalmente vengono fornite nei codici di catalogazione.

# Standard ISBD

- Gli standard ISBD si dividono in varie sezioni specifiche, a seconda del materiale trattato:
  1. ISBD(G) lo standard generale di riferimento per tutti i tipi di materiale documentario
  2. ISBD(M) lo standard di riferimento per le monografie
  3. ISBD(S) lo standard di riferimento per i periodici
  4. ISBD(NBM) lo standard di riferimento per il materiale non librario
  5. ISBD(CM) lo standard di riferimento per il materiale cartografico
  6. ISBD(PM) lo standard di riferimento per le opere musicali a stampa
  7. ISBD(A) lo standard di riferimento per il libro antico
  8. ISBD(CF) lo standard di riferimento per computer files
  
- Ogni manuale ISBD è concepito in maniera da raccogliere un insieme coerente di indicazioni per la propria categoria di pubblicazioni, ma senza alcun intento di rendersi esclusiva. Tutte le ISBD si fondano sull'ISBD generale ISBD(G).

# Codifica internazionale dei libri

 Codice ISBN

 Codice ISSN

 Codice EAN



# Codice ISBN - Storia

- A causa della complessità e lo sviluppo dell'editoria è stato necessario adottare, su scala internazionale, un sistema unificato per la numerazione e la gestione automatizzata dei libri. Prima tappa di questo processo di sviluppo nei sistemi di standardizzazione libraria è stato lo Standard Book Numbering, ideato nel 1967 in Gran Bretagna e trasformatosi successivamente nell'attuale ISBN, il solo codice di identificazione internazionale per i libri e gli altri media, come definito dallo Standard ISO 2108.

# Codice ISBN – Definizione

- L'ISBN (International Standard Book Number) permette l'immediata e inequivocabile identificazione di un titolo o di un'edizione di un titolo di un determinato editore. La sua utilità si può facilmente riscontrare sia nell'ambito commerciale (rapidità e sicurezza nell'ordinazione e nell'evasione degli ordini, facilità nella rilevazione dei vari dati informativi e statistici) sia nel settore delle biblioteche (controllo, registrazione e varia utilizzazione bibliografica).

# Codice ISBN – Struttura [1]

■ L'ISBN è un numero formato da dieci cifre

(Esempio Il numero 88 222 0003 9), i requisiti previsti da tale sistema di numerazione in generale sono due:

- Il totale delle cifre di cui si compone il numero è sempre dieci
- Il numero risulta composto da quattro parti

# Codice ISBN – Struttura [2]

- **Prima parte:** numero di identificazione del gruppo nazionale, linguistico o geografico. Viene attribuito dall' Agenzia ISBN, esso può avere da 1 a 5 cifre, l' area di lingua italiana ha il numero 88.
- **Seconda parte:** numero di identificazione dell'editore. Viene attribuito dall'Agenzia ISBN per l'area di lingua italiana, esso può avere da 2 a 6 cifre.
- **Terza parte:** numero di identificazione del titolo. Tenuto conto che l' ISBN è composto di 10 cifre, il numero di identificazione del titolo potrà disporre di tante cifre quante saranno lasciate disponibili dal numero di identificazione della nazione e dell'editore che lo precede e dal numero di controllo che lo segue. In pratica, all'interno di un'area linguistica, restando fisso lo spazio occupato dal prefisso di identificazione del gruppo e dal numero di controllo, il numero dei codice disponibili per i singoli editori varierà in ragione inversa della lunghezza del loro prefisso.



# Codice ISBN – Struttura [3]

## ■ **Quarta parte:** numero di controllo

Il numero di controllo è una garanzia contro i possibili errori dovuti alla trascrizione manuale. Può variare da 0 a 10 (avendo a disposizione una sola posizione, il 10 viene convenzionalmente indicato in numero romano). Funziona in questo modo: quando un codice ISBN è inserito nel calcolatore, la prima operazione che effettua la macchina è di controllare se il numero è giusto. Il calcolatore moltiplica, con estrema rapidità, ogni cifra di cui si compone il numero per 10, 9, 8, 7, 6, 5, 4, 3, 2 (questi numeri sono detti "pesi"). Dopo aver fatto la moltiplicazione, i risultati vengono sommati e al totale viene aggiunto lo stesso numero di controllo. Il totale finale è poi diviso per quello che viene chiamato il modulo 11. Se, dopo che il totale è stato diviso per 11, non c'è resto, allora il codice è giusto ed è accettato dal computer. Se c'è il resto, il numero è respinto.

# Attribuzione del codice ISBN ai libri [1]

- Dopo esser stati elaborati correttamente e consegnati all' Editore, i numeri di codice vanno attribuiti ai volumi in modo che si possano identificare senza possibilità di errore, anche in questa fase si devono seguire le norme internazionali.

Il principio fondamentale è che ogni libro abbia un suo proprio ISBN, anche ogni edizione di uno stesso titolo ha un suo proprio ISBN.

Esempio:

Ril. Tela	Euro 10,00	ISBN 88-17-45345-5
Cartonato	Euro 7,00	ISBN 88-17-45346-3
Brossura	Euro 5,00	ISBN 88-17-45347-1

- Una volta assegnato, l'ISBN non può più essere utilizzato, anche se il libro al quale è stato inizialmente attribuito è esaurito da molto tempo.

# Attribuzione del codice ISBN ai libri [2]

- **Nuove edizioni** Ogni edizione che comporti modifiche sostanziali rispetto all'edizione precedente (pagine, illustrazioni, formato, ecc.) richiede l'attribuzione di un nuovo numero.
- **Ristampe** La pura e semplice ristampa non comporta l'assegnazione di un nuovo numero, anche nel caso in cui venisse variato il prezzo.
- **Coedizioni** I libri pubblicati in coedizione da due o più editori dovranno portare il codice ISBN dell'editore che li distribuisce.
- **Opere in più volumi** Un'opera in più volumi richiede un ISBN complessivo. Inoltre ogni volume, vendibile separatamente, di un'opera in più volumi, deve avere un proprio ISBN.

Un esempio

Opera completa: ISBN 88-222-0003-9

1- volume: ISBN 88-222-00004-7

2- volume: ISBN 88-222-0005-5

- **Opere distribuite** Ogni editore, oltre che alle proprie pubblicazioni, può assegnare un ISBN a ogni titolo di cui è distributore esclusivo, quando il libro sia stato pubblicato in un paese che non ha ancora adottato l'ISBN. Altrimenti è di regola usare l'ISBN che ad ogni singolo volume è stato attribuito dall'editore originale.

# Attribuzione del codice ISBN ai libri [3]

- **Cataloghi o listini editoriali** I cataloghi o i listini editoriali, oltre al numero ISBN stampato a fianco del libro, possono contenere anche un indice generale degli ISBN.
- **Pubblicazioni a fogli mobili** Se un'opera viene pubblicata in un'edizione a fogli mobili viene attribuito un ISBN per identificare quell'edizione, a quella data. Per successivi aggiornamenti o sostituzioni di pagine si dovranno assegnare altri ISBN.
- **Pubblicazioni elettroniche** Va attribuito l'ISBN esclusivamente a:
  - pubblicazioni in linea (ad esempio: un libro edito su Internet);
  - basi di dati bibliografiche o di documenti.
  - i documenti "linkati" fra loro possono avere un unico ISBN, solo se fanno parte di un'unica pubblicazione.

# Stampa del codice ISBN sui libri

- L'ISBN è solitamente stampato sul retro del frontespizio, deve apparire anche sulla quarta di copertina. Nel caso in cui ciò non fosse possibile, si dovrà collocare l'ISBN esternamente, in un punto ben visibile. Il corpo del carattere dell'ISBN non deve essere mai inferiore ai 9 punti. Le varie parti del numero sono separate da un trattino (-). Il numero va preceduto dalla menzione ISBN.

Esempio:

ISBN 88-XXX-XXXX-X

# Archivio del codice ISBN

- L'editore avrà cura di tenere debitamente archiviati i codici ISBN di volta in volta attribuiti ai propri titoli, in modo da poter effettuare, all'occorrenza, ogni necessario controllo.

# Codice ISSN - International Standard Serial Number

- Il codice ISSN è studiato per i periodici, identifica la testata e non i singoli fascicoli.  
In alcuni casi particolari (annuari, collane monografiche e altre pubblicazioni affini), si può attribuire un ISSN alla "serie" e un ISBN al singolo volume.

# Codice EAN - Introduzione


- Un ulteriore e decisivo passo in avanti nel cammino del sistema gestionale automatizzato che ha come base l'ISBN è ora costituito dal codice a barre EAN (European Article Numbering), destinato a facilitare ancora di più tutte le operazioni di controllo, distribuzione e vendita del libro. I vantaggi si possono così esemplificare:
  - sensibile riduzione dei tempi di fatturazione, grazie alla rapidità di lettura del codice a barre per mezzo di una penna ottica o di uno scanner collegati a un calcolatore;
  - diminuzione dei tempi di evasione degli ordini, grazie a un sistema autonomo di controllo e di gestione;
  - rapida ed efficiente gestione delle rese librerie;
  - notevole semplificazione nell' inventariazione dei volumi nei magazzini editoriali e in libreria;
  - la somma di tutti questi vantaggi comporta minori costi e grande risparmio di tempo.



# Codice EAN - Struttura

- il sistema EAN comporta che ad ogni prodotto venga attribuito un codice di identificazione composto di 13 cifre e di una corrispondente rappresentazione grafica consistente in un simbolo a barre verticali destinato alla lettura ottica. Nella sua utilizzazione, il codice EAN - collegato all'ISBN - ha la seguente struttura:
  1. Le prime tre cifre sono fisse e identificano internazionalmente il prodotto libro (978).
  2. Le cifre dalla 4a alla 12a corrispondono alle prime nove cifre dell'ISBN.
  3. L'ultima è la cifra di controllo (sempre numerica) che sostituisce quella (alfanumerica) dell'ISBN.
  
- Esempio: ISBN: 88-7075-083-3 EAN: 9788870750836

# Codice EAN – Stampa Libri e Prodotti non librari

 Il codice EAN va stampato orizzontalmente sulla quarta di copertina o sul retro della confezione in basso a destra, senza tagli o riduzioni.

# I linguaggi di markup

- Gli standard tradizionali trattano solo record bibliografici e non interi documenti composti di testo, immagini, suono, come quelli che caratterizzano una biblioteca digitale. Sono così nati, fuori dall'ambito bibliotecario, nuovi standard (linguaggi di markup) come SGML (Standard Generalized Markup Language) e XML (Extended Markup Language).
- I linguaggi di markup svolgono il compito di digitalizzare i testi, questi linguaggi informatici si basano su un insieme di istruzioni per la macchina, orientate alla descrizione dei fenomeni di strutturazione, composizione ed impaginazione del testo.

# SGML - Introduzione

- SGML è un metalinguaggio, cioè un linguaggio usato per la descrizione di altri linguaggi, il suo nome completo è Information Processing - Text and Office Systems - Standard Generalized Markup Language. Lo standard è nato come Gml (dal nome degli autori Godfarb, Mosher e Lorie) nel 1969 all'IBM, ed è stato usato per l'editoria prima di diventare standard Ansi Z39.59 e successivamente nel 1986 standard internazionale ISO 8879.

# SGML – Il markup

■ L'idea centrale dello standard è un tipo di marcatura generica chiamata “marcatura descrittiva” (descriptive markup) che definisce la struttura logica dei documenti. Il termine “marcare” (markup) è preso dall'editoria dove viene usato per la codifica sul manoscritto delle istruzioni date dall'editore al tipografo. Questa marcatura usata dai tipografi riguarda essenzialmente lo stile di presentazione del testo a stampa e si chiama più propriamente “marcatura procedurale”(procedural markup). Con la marcatura descrittiva è possibile utilizzare lo stesso documento per la stampa, per la ricezione in linea, per la pubblicazione su cd-rom oppure per il riuso del documento integrandovi altro testo. L'organizzazione di un documento non è espressa usando la codifica dei sistemi di scrittura, che è finalizzata alla presentazione grafica, ma sono evidenziate le parti in cui è strutturato il documento (ad esempio paragrafi, capitoli) insieme ad altre particolarità del testo (come note, tabelle, intestazioni). La marcatura descrive la struttura logica di un documento usando un modello gerarchico che riflette l'interrelazione nel testo di elementi ed entità. La struttura logica del documento è definita nel Document type definition (Dtd). Sgml può definire un numero infinito di tipi di documento, per ogni categoria di informazioni prodotta.

# Cos' è SGML?

- Lo standard SGML può essere considerato la grammatica di un documento in cui gli elementi che lo compongono sono il vocabolario utilizzato, mentre la loro relazione rappresenta la sintassi astratta del documento. Il Dtd definisce questa sintassi.  
Ogni documento in SGML è diviso in tre parti:
- la prima parte si chiama “Sgml declaration” ed appare come segue: `<!Sgml "Iso 8879:1986"`. Definisce il set di caratteri usato, caratteri a cui è attribuito un particolare significato ed altre particolarità che sono usate. Tutto ciò si omette se si usano i valori di default della sintassi del Dtd.
- la seconda parte specifica il Dtd cioè la struttura del tipo di documento: `<!DOCTYPE...` (seguito dal nome della classe di documenti)
- la terza parte è il contenuto del documento (occorrenza o “instance” di quella classe di documenti specificata nel Dtd) ed include ad esempio il titolo, i paragrafi, gli indici, le tabelle, i grafici ed eventuale audio collegato). Ogni elemento è definito da una etichetta (tag) che precede e conclude il contenuto.

Ad esempio `<TITOLO> Contenuto di questo elemento </TITOLO>`

# SGML - parser

- Per realizzare una collezione di documenti SGML è richiesta prima di tutto chiarezza di obiettivi e, una volta raggiunta questa chiarezza, un'approfondita analisi del Dtd del documento. Per questa analisi c'è un programma che si chiama Parser, il suo scopo è quello di guidare la marcatura dei documenti SGML e di controllare se i documenti sono o no conformi allo standard (questo processo si chiama “validazione”). Dopo la validazione ogni documento può essere riformattato nell'applicazione scelta, esattamente come se fosse stato creato direttamente con quel programma. Di solito il Parser è parte di un programma complesso che esegue l'editor, la conversione, registra i documenti ed eventualmente realizza anche i programmi di ricerca, navigazione e penetrazione.

# XML - Introduzione

- XML (eXtensible Markup Language) è stato sviluppato dal W3C (World Wide Web Consortium), è un linguaggio di markup realizzato per poter utilizzare in modo semplice i documenti strutturati e studiato per il Web. XML è una semplificazione di SGML, infatti, include le parti di SGML che sono state usate più frequentemente ma omette tutte quelle caratteristiche facoltative che lo hanno appesantito. E' un linguaggio estendibile e quindi permette agli utenti di definire le modifiche e gli attributi dei loro documenti; inoltre ha adottato Unicode: questo standard viene utilizzato per la codifica dei caratteri ed è in grado di supportare la maggior parte delle lingue. Poiché XML segue questo standard, è possibile utilizzare qualunque lingua per il markup e il software, XML è in grado di capirla senza costringere gli utenti ad usare l'inglese come codice per le istruzioni.



# XML - Caratteristiche

- consente di utilizzare documenti strutturati
- è estensibile, per cui permette di aggiungere sempre nuovi marcatori
- offre un ottimo formato di scambio di dati, è strutturato, non ambiguo e completamente leggibile (non binario) e sarà comunque riutilizzabile se si considera che anche i programmi ad ogni nuova versione cambiano formato
- la strutturazione e l'utilizzo di un linguaggio estensibile basato su tag consente una più semplice interazione con altri programmi, compresi i database, e quindi un trattamento dei dati più semplice ed efficace
- è indipendente dalla piattaforma e dal processore
- permette un semplice utilizzo di metadati, come Dublin Core, etc.
- consente ricerche più semplici e più efficaci: ad esempio, attraverso il controllo sui tag, sarà possibile effettuare una interrogazione tramite un motore di ricerca assicurandosi una risposta più precisa rispetto a ciò che realmente stiamo cercando
- offre un buon meccanismo di rappresentazione, una ottima capacità di rappresentare dati complessi (notazioni matematiche, interfacce grafiche)
- offre possibilità di presentazioni superiori a quelle di HTML che, per ottenere risultati simili, deve ricorrere all'utilizzo di Java o altri linguaggi

# FONTI

 <http://spazioinwind.libero.it>

 [www.alice.it](http://www.alice.it)

 [www.supereva.it](http://www.supereva.it)

