

## 5 L'accesso alle risorse digitali e i metadati

### 5.1. Introduzione

La varietà dei contenuti e dei supporti degli oggetti digitali, la diversa provenienza geografica ed istituzionale degli attori che nella rete sono direttamente coinvolti nella creazione, distribuzione e recupero delle risorse elettroniche, i differenti modelli economici che stanno alla base delle attività di produzione e accesso sono le componenti fondamentali che caratterizzano l'ambiente digitale.

A fronte di questa diversificazione vi è una duplice esigenza: da una parte l'integrazione delle risorse per garantire servizi di accesso esteso e combinato, dall'altra la salvaguardia, per singole comunità di utenti, della specificità della descrizione, gestione e recupero dei diversi materiali.

Per fare ordine nel caos del web e quindi gestire adeguatamente gli oggetti digitali dal momento della loro creazione fino al loro accesso è necessario disporre di informazioni che ne descrivono il contenuto, la struttura, le caratteristiche tecniche, le condizioni di uso e le modalità di conservazione per la disponibilità futura. Tali informazioni sono denominate metadati.

E' unanimemente riconosciuto oggi che l'interoperabilità e l'interazione necessarie fra le biblioteche digitali dipende in realtà da una efficace condivisione degli elementi informativi che descrivono e garantiscono la gestione delle risorse trattate: i metadati.

Al mondo ben delimitato, sia pure variegato, delle biblioteche, dei centri di documentazione e dei servizi di indicizzazione negli ultimi dieci anni si è affiancato e in parte sovrapposto un mondo disomogeneo, quello della biblioteca digitale distribuita, che inevitabilmente non può conformarsi in pieno al modello biblioteconomico che ha visto per molti anni operare efficacemente tecniche consolidate per l'identificazione e la ricerca dei documenti. Come è noto questi strumenti sono costituiti dalle regole di

catalogazione che incorporano principi definiti a livello internazionale, dagli standard per la descrizione (le varie ISBD per i diversi tipi di materiale), dai formati bibliografici MARC (Machine Readable Catalog) per la codifica dei dati catalografici in modo analitico e normalizzato.

Tali strumenti sono ancora validi per una parte del materiale digitale, quello su cui la biblioteca è in grado di garantire un accesso puntuale e mantenere un controllo sulla sua evoluzione, sia per quanto riguarda il contenuto delle risorse che la loro localizzazione.

Due importanti caratteristiche del mondo digitale stanno all'origine dell'esigenza di nuovi strumenti per la descrizione e la gestione delle risorse. Una è costituita dalla quantità e varietà sempre crescente di risorse presenti in rete: questa continua proliferazione non consente ai produttori di risorse elettroniche e ai fornitori di servizi, per motivi economici ed organizzativi, di utilizzare in modo esteso gli strumenti complessi ma anche in certo modo settoriali usati dalle biblioteche. L'altra componente è legata alla natura stessa delle risorse digitali, che si prestano ad un diverso livello di analisi delle singole unità di informazione che le compongono, secondo i servizi da produrre. Il riferimento in questo caso è al fenomeno della granularità: la descrizione e la gestione (e quindi i relativi metadati per l'accesso e il controllo) possono riguardare il livello della collezione nel suo insieme oppure un oggetto in essa contenuto, fino ad una unità elementare, come una singola pagina o un'immagine, che deve essere opportunamente correlata all'oggetto di cui fa parte.

Resta comunque l'esigenza di gestire e recuperare le risorse in modo puntuale e preciso, che va oltre i risultati offerti dai pur potenti motori di ricerca commerciali che operano con tecniche particolari per la raccolta e l'accesso alle risorse. Molto raramente essi sono in grado di contestualizzare l'informazione e quindi di garantire una ricerca puntuale, orientata su determinati elementi, dal momento che non utilizzano i metadati per il recupero, ma il più delle volte il testo pieno ed alcune parti delle risorse da cui estraggono automaticamente i termini da indicizzare. I limiti sono ben noti: mancata conoscenza, per l'utente, della copertura (e

quindi della consapevolezza di quali siano le risorse in grado di essere catturate a seguito di una ricerca), scarsa precisione e quindi fenomeni di 'rumore' nel recupero, presentazione dei risultati della ricerca secondo criteri di priorità non sempre chiari e spesso devianti perché legati a motivi commerciali e pubblicitari, infine volatilità e inattendibilità degli indirizzi su cui le risorse dovrebbero essere accessibili.

Recupero puntuale e integrazione di risorse eterogenee in un servizio unificato a favore dell'utente finale sono due sfide in certo modo contrastanti, ma da superare e raggiungere auspicabilmente in modo congiunto, a cui sono dedicati oggi notevoli sforzi da parte di una molteplicità di esperti: creatori di schemi (insiemi di metadati omogenei per un determinato campo disciplinare o tipo di risorse) e di linguaggi di marcatura, catalogatori, tecnici informatici, implementatori di servizi, di programmi applicativi anche di pubblico dominio e di protocolli specifici per la raccolta di metadati.

## **5.2 La proliferazione degli schemi di metadati**

Nell'ambiente del web la biblioteca digitale è ormai distribuita: sotto un unico tetto confluiscono risorse eterogenee, in parte destrutturate e prive di metadati, in parte descritte da schemi di metadati differenti, tanto negli attributi assegnati alle risorse, quanto nel livello di analisi.

Gli elementi e quindi i metadati che contraddistinguono le risorse rispecchiano non solo la loro natura e il loro contenuto, ma anche le modalità di produzione, le caratteristiche tecniche, la protezione dei diritti e le condizioni di accesso, le specifiche per il loro uso nel tempo.

I metadati possono essere suddivisi in tre categorie: metadati descrittivi, amministrativi gestionali, strutturali.

Più precisamente le funzioni che i metadati svolgono possono essere schematicamente riassunte nel modo seguente:

- a) Trovare, identificare, selezionare risorse (metadati descrittivi)
- b) Gestire gli oggetti digitali di una collezione e quindi garantirne l'acquisizione, archiviazione e fruizione, l'utilizzo sulla base di eventuali

diritti e licenze, la conservazione e l'uso futuro, la certificazione dell'autenticità e integrità (metadati amministrativi, gestionali e tecnici)  
c) Collegare le varie componenti delle risorse per un'adeguata e completa fruizione (metadati strutturali).

Il presente studio prende in considerazione le tre tipologie di metadati. Maggiore enfasi è data alla prima categoria di metadati, in quanto si tratta di un'area di forte sviluppo dal punto di vista dell'elaborazione di standard e nella quale sono state raggiunte una maturazione concettuale e una ricchezza di sperimentazioni anche ai fini della condivisione di risorse, che è un aspetto fondamentale della biblioteca digitale.

Le altre tipologie di metadati sono altrettanto essenziali; al momento, con il diffondersi di progetti digitali, la problematica di definire i metadati che documentano le caratteristiche tecniche, il ciclo di vita e le condizioni di uso delle risorse è salita alla ribalta e numerose sono le raccomandazioni e iniziative che si stanno concretizzando nell'elaborazione di modelli e schemi di metadati orientati a scopi gestionali. In questo settore la standardizzazione è di livello più basso rispetto all'area dei metadati descrittivi, ma è importante seguirne gli sviluppi al fine di prendere in attenta considerazione, in progetti di biblioteca digitale, le diverse funzioni a cui i metadati sono collegati e operare quindi le opportune scelte su quali elementi adottare sulla base degli schemi disponibili.

### **5.2.1 I metadati descrittivi e il processo di creazione degli schemi**

Fin dall'inizio della comparsa del web diverse comunità disciplinari e istituzionali si sono dedicate alla definizione dei loro schemi di metadati, identificando gli elementi che caratterizzano il contenuto e la forma delle risorse da loro prodotte o comunque trattate. Tale attività si è realizzata in cooperazione con settori ed organismi affini in modo da creare, nei diversi campi, schemi omogenei, capaci di descrivere e gestire risorse che condividono caratteristiche comuni. Alla base della predisposizione di un insieme definito di elementi per i vari tipi di oggetti digitali vi è un duplice intento: fornire uno strumento consolidato a chi crea i metadati che

identificano questi oggetti e allo stesso tempo consentire un recupero coerente delle risorse descritte.

L'elaborazione degli schemi si svolge dunque attraverso il lavoro congiunto di esperti del settore provenienti da istituzioni affini. Si tratta qui di un livello di aggregazione piuttosto forte, che si realizza nella definizione in comune di elementi, garantendo comunque un grado di flessibilità ed estensibilità da parte di singole applicazioni.

Fra gli schemi più diffusi che si sono venuti sviluppando in seguito alla produzione di risorse digitali e all'esigenza di accesso sul web figurano quelli relativi ai settori dei musei, degli archivi, dei materiali didattici, della documentazione ufficiale governativa e dei servizi che rientrano nel cosiddetto 'governo elettronico', dei dati geospaziali, delle immagini, degli oggetti multimediali, della musica.

Il MARC (Machine Readable Catalog), nelle due versioni oggi più diffuse, MARC21 e UNIMARC, rappresenta ancora il formato per eccellenza utilizzato dalle biblioteche per la codifica dei record bibliografici, anche se per la descrizione delle risorse digitali gestite dalla biblioteca si sta affermando l'esigenza di modelli che si affianchino al MARC: da una parte Functional Requirements for Bibliographic Record (FRBR), dall'altra un formato elementare di metadati, da utilizzare per un'ampia gamma di risorse. Questo ultimo schema di base, denominato Dublin Core, viene più avanti ampiamente illustrato per il suo significato particolare: non si tratta dello schema di metadati per eccellenza, ma di un minimo comune denominatore da utilizzare per integrare risorse e applicazioni eterogenee in servizi di accesso unificato.

E' utile sottolineare che FRBR non è in realtà un vero e proprio schema codificato, comprensivo di una serie di elementi determinati; esso è piuttosto un modello concettuale, una categorizzazione di entità in cui può essere scomposto il tradizionale monolitico record bibliografico. Le entità di cui si compone FRBR sono costituite dalla creazione intellettuale o artistica, dalla responsabilità intellettuale, dai concetti espressi nelle opere che sono oggetto di analisi. Le creazioni intellettuali rappresentate in

documenti o oggetti possono scomporsi a loro volta in quattro entità: l'opera, l'espressione, la manifestazione e l'esemplare. In tale quadro le entità sono fra loro collegabili per esprimere la paternità intellettuale, l'attività di pubblicazione, di possesso, di custodia, la pertinenza ad una determinata categoria concettuale. Tale modello si presta molto bene alla gestione della biblioteca ibrida, in cui sono trattate pubblicazioni tradizionali ed elettroniche e dove è essenziale mettere in relazione versioni multiple di una risorsa, rifacimenti, edizioni su supporti e formati diversi. Le applicazioni di questo modello sono all'inizio<sup>59</sup> ed esso suscita attualmente molto interesse a fini didattici e per le potenzialità offerte in fase di ricerca e navigazione dei dati.

Encoded Archival Description (EAD)<sup>60</sup> è lo standard in formato SGML/XML sviluppato per la descrizione di materiale archivistico, ma anche di collezioni bibliografiche, oggetti museali e fondi di manoscritti. È essenzialmente uno strumento nato per la creazione di inventari, sotto forma di basi dati spesso altamente strutturate, così come lo schema descrittivo DTD-EAD (Document Type Definition), disegnato per rispondere alle esigenze di analisi a livello di fondo, sottofondo, serie e singolo documento.

Metadata Object Description Schema (MODS)<sup>61</sup>, messo a punto dalla Library of Congress di Washington per la descrizione delle risorse elettroniche, è un formato essenzialmente bibliografico che include un sottoinsieme dei campi MARC, utilizzando però etichette mnemoniche. La sua messa a punto e applicazione in diversi progetti della Library of Congress trova la sua ragione nel fatto che è più ricco di Dublin Core, configurandosi quindi come un modello di metadati di livello medio.

Per la descrizione e gestione dei materiali didattici diverse sono le iniziative di standardizzazione sui metadati al fine di aiutare insegnanti, venditori di risorse digitali, studenti e ricercatori nella elaborazione e utilizzo di tutto

---

59 Australian Literature Gateway-AUSTLIT. <http://www.austlit.edu.au/>

60 Encoded Archival Description (EAD). <http://www.loc.gov/ead>

61 MODS. <http://www.loc.gov/standards/mods>

quello che viene prodotto in materia su supporto digitale, come corsi on line, dispense, prodotti multimediali, ma anche documentazione varia a supporto della formazione, come ad esempio schede di valutazione di corsi e di formatori.

Instructional Management System (IMS)<sup>62</sup> è insieme una iniziativa di cooperazione e una serie di standard sviluppati per costruire un sistema aperto nel campo dello studio e della didattica. IMS include diverse specifiche che si riferiscono alla produzione di materiale didattico fino all'elaborazione di uno standard di metadati che è stato creato in collaborazione tra istituzioni fra le quali figura l'Institute of Electrical and Electronic Engineers (IEEE). Denominato Learning Object Metadata Standard (LOMS)<sup>63</sup>, l'insieme di elementi è suddiviso in varie categorie, che contengono i loro specifici metadati: questi si riferiscono non solo alle caratteristiche dei singoli oggetti, ma anche all'identificazione in modo standard tanto dei destinatari delle risorse quanto delle modalità di utilizzo dei vari materiali didattici. Un aspetto molto importante degli oggetti trattati in questo settore, che si riflette nei metadati previsti, è la molteplicità dei punti di vista con cui uno stesso contenuto può essere valutato e analizzato da parte di insegnanti, studenti, produttori.

Government Information Locator Service (GILS)<sup>64</sup> è lo standard di metadati più utilizzato negli Stati Uniti per le informazioni fornite dal governo federale, ma è adottato anche in altri paesi<sup>65</sup>. Gli elementi che compongono questo schema rappresentano una serie estesa dell'insieme degli attributi BIB-1 del noto protocollo ANSI/NISO Z39.50, assai diffuso nel mondo degli OPAC bibliotecari e utilizzato per interfacciare cataloghi distribuiti. Gli elementi definiti sono solo 28, ma è possibile precisare il loro significato attraverso l'uso di qualificatori. Molte sono le 'mappature' o corrispondenze che sono state elaborate fra GILS e altri schemi di metadati, come ad esempio MARC21, EAD, Dublin Core.

---

62 IMS. <http://www.imsproject.org>

63 LOMS. <http://ltsc.ieee.org/wg12>

64 GILS. <http://www.gils.net>

65 GILS implementers. <http://www.gils.net/examples.html>;  
<http://www.gils.net/implement.html>

Le applicazioni nel campo dei dati geospaziali si basano sul lavoro svolto dal Federal Geographic Data Committee (FGDC)<sup>66</sup> che ha elaborato uno standard per la descrizione di risorse in questa area, chiamato Content Standard for Digital Geospatial Metadata (CSDGM, ma spesso denominato con l'acronimo dell'organismo che lo ha elaborato: FGDC). Gli oltre 300 elementi che caratterizzano le risorse digitali sono divisi in sezioni relative ad esempio alle informazioni di identificazione, qualità dei dati, organizzazione dei dati spaziali, date, autori dei metadati. Diversi sono gli adattamenti del formato (chiamati 'profili di applicazione') e numerosi i progetti che fanno uso di questo standard.

Gli standard di metadati per gli oggetti artistici sono numerosi, orientati a descrivere singole opere e immagini. Esiste uno standard di base a cui i vari schemi si uniformano, denominato Categories for the Description of Works of Art (CDWA)<sup>67</sup>, messo a punto con il coordinamento dell'Istituto Getty. Le applicazioni che trattano oggetti artistici dovrebbero essere modellate sugli schemi messi a punto in questo settore, con la consapevolezza che i progetti finora realizzati non hanno utilizzato un singolo schema come tale, ma hanno operato adattamenti e combinazioni di elementi.

Vengono qui elencati alcune iniziative, come Computer Interchange Museum Information (CIMI)<sup>68</sup> che ha dato origine allo schema Conceptual Reference Model (CIDOC-CRM)<sup>69</sup>, diffusamente utilizzato in varie applicazioni, anche italiane.

Multimedia Content Description Interface e Moving Pictures Expert Group (MPEG) rappresentano un sistema che fornisce un insieme di strumenti standardizzati per descrivere contenuti multimediali. Lo standard MPEG-7<sup>70</sup> è formato da diverse parti, relative alla codifica dei record descrittivi, al linguaggio per la definizione della sintassi delle descrizioni MPEG (chiamato MPEG-7) Description Definition Language-DDL), alle linee guida e procedure di test di conformità nell'applicazione dello standard.

---

66 FGDC. <http://www.fgdc.gov/metadata/metadata.html>

67 CDWA. <http://www.getty.edu/research/institute/standards/cdwa/index.html>

68 CIMI. <http://www.cimi.org/now.html>

69 CIDOC conceptual reference model. <http://cidoc.ics.forth.gr/>

70 MPEG. <http://mpeg.telecomitalia.com/standards/mpeg-7/mpeg-7.htm>

Visual Resources Association (VRA) Core Categories<sup>71</sup> è uno standard per la creazione di record che descrivono opere visive e immagini, utile per grafici e fornitori di applicazioni e servizi che trattano collezioni di immagini digitali. Nella definizione dei 17 elementi l'intento è stato quello di essere in linea con lo schema Dublin Core. È previsto l'uso di qualificatori per rendere più precisi gli elementi stessi, i quali possono riferirsi sia al livello dell'opera che a quello di una singola immagine.

Da ricordare è anche un importante strumento messo a punto dal Comitato NISO (l'ente di standardizzazione americano) per la descrizione e la gestione di immagini fisse<sup>72</sup>: si tratta di un dizionario di dati che riporta i metadati tecnici caratteristici di questo tipo di risorse.

Online Information Exchange (ONIX)<sup>73</sup> è un formato di metadati mantenuto da diverse associazioni che operano nel settore della produzione e del commercio librario, fra cui European Group for Electronic Commerce in the Book and Serials Sector (EDItEUR). Messo a punto per lo scambio fra editori e librai di dati bibliografici descrittivi, amministrativi e strutturali relativi a libri a stampa ed elettronici (è in corso il trattamento esteso anche a video e periodici), ONIX può essere impiegato in versione semplificata o integrale.

Utilizzato da importanti librai come Amazon, Barnes & Nobles, da distributori quali Baker & Taylor, Follett e Ingram e da editori quali Cambridge University Press, McGraw-Hill e Yale University Press, lo standard è di interesse anche per le biblioteche nei loro scambi con il mondo del commercio librario e a questo scopo è già stata resa operativa la conversione di ONIX verso il MARC.

Text Encoding Initiative (TEI)<sup>74</sup> è uno standard in formato SGML per la codifica e lo scambio di testi nell'area umanistica e letteraria. Nato alla fine degli anni '80 negli Stati Uniti per importanti progetti di conversione digitale, su quella base è stato costituito un consorzio (TEI-C) a cui

---

71 VRA-Core. <http://www.vraweb.org/vracore3.htm>

72 Technical metadata for digital still images. <http://www.niso.org/pdfs/DataDict.pdf>

73 ONIX. <http://www.editeur.org/onix.html>

74 Text Encoding Initiative : the TEI guidelines. <http://www.tei-c.org/Guidelines2/index.html>

partecipano istituti prestigiosi, come l'università di Bergen in Norvegia, quelle di Oxford e della Virginia. Compiti fondamentali del consorzio sono la manutenzione dello standard e attività di sviluppo di progetti, di formazione e consulenza.

Nello standard l'intestazione TEI (TEI Header), sviluppata in collaborazione con bibliotecari, rappresenta la descrizione bibliografica di un documento. Le quattro parti di cui si compone l'intestazione comprendono: a) la descrizione del file elettronico e quindi il testo in forma digitale; b) la descrizione della codifica, dove viene indicata la relazione fra il testo in formato elettronico e le fonti da cui è tratto, come ad esempio le normalizzazioni operate, il tipo di codifica, etc.; c) il profilo del testo, che contiene informazioni sul testo stesso, come il soggetto o la classificazione, sulle persone descritte o coinvolte nella produzione del testo; d) la storia della revisione, dove viene registrata, da chi esegue la codifica del testo, la situazione dei cambiamenti apportati nella produzione del testo elettronico.

TEI è uno standard consolidato; avendo al suo attivo l'applicazione in una novantina di progetti in tutto il mondo, esso costituisce un punto di riferimento fondamentale in progetti di digitalizzazione e codifica di collezioni testuali.

Gli standard adottati per la descrizione di collezioni digitali musicali sono vari, dal MARC a EAD e Dublin Core, ma diverse sono le iniziative dirette alla definizione di schemi specifici per il materiale musicale, come ad esempio Standard Music Description Language (SMDL)<sup>75</sup>, utilizzato in vari progetti fra cui Cantate<sup>76</sup>, Music XML<sup>77</sup> in cui sono definiti metadati per la pubblicazione di dischi, MusiCat<sup>78</sup> che rappresenta un formato XML per la catalogazione di materiale musicale.<sup>79</sup>

---

75 SMDL : a brief discussion of SMDL. ISO/IEC draft international /Stephen R. Mounce.  
<http://www.techno.com/smdl.htm>

76 CANTATE project. <http://projects.fnb.nl/cantate/default.htm>

77 Music XML. <http://www.recordare.com>

78 MusiCat DTD. <http://www.people.virginia.edu/~pdr4h/MusiCat>

<sup>79</sup> Una lista di schemi, i cui siti sono mantenuti dalla Music Library Association, è visibile al sito: <http://www.lib.ox.ac.uk/immpwg/>.

E' utile sottolineare che questa breve rassegna comprende a titolo esemplificativo solo una selezione di schemi: si tratta dunque di insiemi di elementi che sono già collaudati in una serie di applicazioni e che vengono continuamente verificati e valutati alla luce delle varie sperimentazioni e dei nuovi tipi di risorse prodotte. Tutti hanno un punto in comune: la preoccupazione di rapportarsi e quindi di poter essere convertiti verso uno schema di base, rappresentato da Dublin Core, per quegli elementi base che sono comuni alla maggior parte delle risorse, come ad esempio titolo, creatore, data, soggetto, etc.

### **5.2.2 La corrispondenza fra gli schemi per l'accesso integrato**

Rispetto a solo pochi anni lo scenario degli schemi di metadati, riportato al paragrafo 1.6.1.2 dell'edizione precedente dello Studio di fattibilità, è cambiato notevolmente e la proliferazione degli schemi di metadati ha reso e rende tuttora problematica l'integrazione di risorse. Mentre questi si sono venuti sviluppando per i vari settori applicativi e disciplinari, si sono ben presto avvertite le implicazioni ai fini della costruzione di servizi di accesso unificato ed è così iniziato un lungo lavoro, tuttora in corso, per garantire un certo livello di corrispondenza fra i differenti schemi. Sono state così preparate molte 'mappature' fra i vari metadati, che si realizzano con tabelle di comparazione e conversione fra schemi diversi che definiscono differentemente i propri elementi. Le tavole di corrispondenza costituiscono lo strumento grazie al quale opportuni programmi sono in grado di interpretare un elemento e quindi un attributo di una risorsa e metterlo in relazione con lo stesso (o simile) attributo, anche se definito in modo diverso da un altro schema di metadati.

Il caso del MARC nel mondo delle biblioteche può essere illuminante come esempio di standard che comporta comunque il suo raccordo con altri sistemi di identificazione e codifica dei dati, ed ora anche con schemi di metadati utilizzati per la descrizione di oggetti digitali. Dal MARC come formato standard di codifica per lo scambio bibliografico messo a punto nella metà degli anni '60 dalla Library of Congress si è passati alle varie versioni nazionali del MARC che hanno avuto bisogno di numerose

mappature fra di loro per facilitare la comunicazione e il trasferimento dei dati fra biblioteche e paesi diversi. Oggi si assiste ad una successiva normalizzazione del MARC, sotto la spinta di considerazioni pragmatiche e quindi anche economiche: la convergenza su due versioni ampiamente adottate, MARC21 e UNIMARC. Questi ultimi sono gli standard che vengono oggi confrontati e mappati frequentemente con gli schemi di metadati per le esigenze di gestione della biblioteca ibrida, che si trova a gestire insieme materiale di tipo tradizionale e digitale. Dalle mappature fra i vari MARC nazionali dunque si è passati alla necessità di definire la corrispondenza fra MARC e schemi di metadati: questo è il chiaro segno della funzione della biblioteca che spazia su un più vasto mondo: quello delle risorse digitali e della loro integrazione in nuovi servizi a beneficio di un'utenza molto più ampia e sempre più sofisticata nelle sue esigenze di ricerca.

Esempi di mappature sono quelle fornite dal Centro UKOLN, presso l'Università di Bath in Inghilterra, di cui viene riportata, a titolo di esempio, una selezione dalla lista messa a disposizione.<sup>80</sup>

- MARC 21 to Dublin Core: MARC to Dublin Core Crosswalk – Library of Congress, Network Development and MARC Standards Office, February 2001. <URL:<http://www.loc.gov/marc/marc2dc.html>>
- Dublin Core to USMARC: Dublin Core/MARC/GILS Crosswalk – Library of Congress, Network Development and MARC Standards Office, November 1999. <URL:<http://lcweb.loc.gov/marc/dccross.html>>
- Dublin Core to EAD. In: Tony Gill, Anne Gilliland-Swetland and Murtha Baca, Introduction to Metadata. Los Angeles, Calif.: Getty Information Institute, 2000. <URL:[http://www.getty.edu/gri/standard/intrometadata/3\\_crosswalks/index.htm](http://www.getty.edu/gri/standard/intrometadata/3_crosswalks/index.htm)>

---

<sup>80</sup> L'elenco esteso è disponibile al sito: <http://www.ukoln.ac.uk/metadata/interoperability>

- TEI header to USMARC/Dublin Core: Recommended Mappings OTA Header/USMARC/Dublin Core Elements - by Richard Giordano (University of Manchester, Department of Computer Science), 8 November 1996. [The full version can be downloaded as a TEI Lite document from the Oxford Text Archive publications page, but a summary in HTML is also available]. <URL:<http://ota.ahds.ac.uk/> >
- FGDC to USMARC: Crosswalk: FGDC Content Standards for Digital Geospatial Metadata to USMARC - by Elizabeth Mangan (Geography and Map Division, Library of Congress).  
<URL:<http://www.alexandria.ucsb.edu/public-documents/metadata/fgdc2marc.html>>

### 5.2.3 Modalità di integrazione dei metadati

La predisposizione di mappature fra metadati rispecchia dunque l'esigenza di creare delle corrispondenze fra gli elementi di schemi diversi utilizzati per la descrizione di risorse, al fine di integrarle in servizi di accesso unificato. Questa tecnica si dimostra efficace quando si tratta di mettere a confronto due o comunque un numero limitato di schemi, ma per realizzare servizi di biblioteca digitale le applicazioni devono trattare molto spesso risorse dei tipi più diversi, differentemente descritte.

In linea generale si distinguono oggi due ipotesi di massima per la gestione della diversità dei metadati. Esse non sono alternative, ma possono coesistere nelle applicazioni che trattano e integrano oggetti digitali eterogenei:

a) Soluzione del 'web semantico'. Essa prevede che l'accesso alle risorse e alle loro reciproche relazioni può essere realizzato grazie ad una modalità standard con cui sono espresse e quindi esplicitamente dichiarate le proprietà delle risorse. Le proprietà sono gli attributi (o i metadati) degli oggetti digitali, assegnati secondo quanto stabilito dai vari schemi di metadati

b) Soluzione di un linguaggio comune, che significa il ricorso a termini univoci (metadati standard), con cui vengono identificate le caratteristiche essenziali delle risorse e quindi le loro proprietà.

Di seguito vengono brevemente analizzate le due soluzioni che vedono oggi diverse implementazioni in progetti di biblioteca digitale.

**a) Soluzione del 'web semantico'.** Con questa espressione si intende una collezione eterogenea di dati, e quindi di risorse di vario tipo, residenti nell'ambiente del web in continuo divenire, che possono essere interpretate e rese accessibili in modo efficace nonostante la diversità dei linguaggi e quindi degli schemi di metadati con cui sono descritte.

L'obiettivo del web semantico è l'accesso alla varietà di risorse informative in modo articolato, rendendo esplicite le relazioni con cui le risorse sono associate in rete. Questa rete di relazioni si costituisce grazie all'attribuzione di proprietà alle risorse secondo schemi di metadati che utilizzano termini diversi per esprimere gli attributi, come ad esempio: 'creatore', 'autore', 'titolo', 'affiliazione' riferita all'ente di appartenenza di una persona, 'indirizzo EM', 'ha sede in', etc.

Nella filosofia del web semantico dunque coesistono diversi schemi di metadati e vari sono gli attributi assegnati alle risorse, ma l'elemento di congiunzione è dato da un modo uniforme con cui è possibile dichiarare il rapporto fra gli oggetti digitali e le loro proprietà: si tratta di un'architettura realizzata grazie a un linguaggio standard: Resource Description Framework (RDF)<sup>81</sup>, messo a punto dal Consorzio per il World Wide Web (W3C). Lo standard prescrive un modo normalizzato di dichiarare le proprietà, secondo il modello: Risorsa-Proprietà-Valore e quindi Soggetto-Predicato-Oggetto.

In questo modo, anche se i predicati assegnati alle risorse sono dei più vari, le relazioni che intercorrono fra le risorse e le relative proprietà, grazie allo standard RDF, sono interpretabili e comprensibili dalle applicazioni preposte alla navigazione e all'accesso.

---

<sup>81</sup> Resource Description Framework. <http://www.w3c.org/RDF/>

La funzionalità di RDF viene analizzata ulteriormente nella sezione dedicata alla codifica dei record di metadati.

**b) Soluzione di un linguaggio comune** e quindi di uno schema uniforme per la descrizione del materiale, che identifica alcuni elementi essenziali delle risorse, al di là delle loro particolarità, che sono definibili da schemi di metadati specifici.

E' utile sottolineare che tale soluzione non significa la rinuncia da parte dei vari settori a dettagliare gli elementi distintivi delle risorse da descrivere, ma semplicemente la fornitura (sia in forma 'originale' che tramite conversione) di record descrittivi con un numero delimitato di elementi, espressi da termini univoci, chiaramente definiti e adatti per la descrizione e il recupero di una vasta tipologia di risorse.

Uno schema base di questo tipo è stato messo a punto per questo specifico scopo: esso è denominato Dublin Core Metadata Element Set (DCMES), di seguito definito semplicemente Dublin Core. La soluzione è semplice e lineare in un contesto di aggregazione di risorse: garantisce un recupero certamente meno preciso rispetto ad applicazioni specifiche su metadati ricchi e dettagliati, ma offre la possibilità di recuperare risorse eterogenee grazie al riconoscimento di elementi comuni, offrendo allo stesso tempo a chi crea i metadati uno strumento facile da apprendere e agile per descrivere le proprie risorse, eventualmente integrando con ulteriori elementi quando è richiesta un'analisi più dettagliata.

Le applicazioni di questo schema elementare sono ormai diffuse in molti settori, non solo da parte degli organismi tradizionalmente preposti ai servizi nel settore della ricerca scientifica e umanistica, come ad esempio le biblioteche, le università, i musei, gli archivi, i servizi di indicizzazione bibliografica, ma anche in aree documentarie di diverso tipo, come ad esempio quello aziendale e del commercio elettronico per lo scambio di informazioni e transazioni, del governo elettronico, della didattica, etc.

#### 5.2.4 Caratteristiche e funzioni dello schema Dublin Core

L'esigenza di fondo a cui lo schema elementare Dublin Core<sup>82</sup> risponde è quella di consentire la coesistenza efficace (interoperabilità) di più risorse nel mondo della produzione digitale, mediante la loro identificazione attraverso un numero limitato di attributi: è questa la soluzione studiata per far convergere attori diversi che operano sotto lo stesso tetto (produttori, distributori, utenti finali) che hanno sicuramente funzioni ed esigenze proprie, ma che sono accomunati dalla necessità di gestire e accedere all'informazione in modo unificato.

Molti progetti e applicazioni operative hanno oggi condiviso questa convinzione: uno schema semplice, costituito da alcuni elementi che sono presenti nella maggior parte delle risorse, è in grado di mettere in comunicazione le varie comunità per la documentazione del proprio materiale, anche convertendo i metadati da loro prodotti, che possono essere, nelle specifiche applicazioni da loro implementate, molto ricchi e dettagliati. Si tratta dunque di un linguaggio comune, definito, nel mondo internazionale del web, con il termine inglese 'pidgin', per mettere in evidenza che si tratta di un vocabolario limitato ed essenziale, adatto ad una comunicazione di base, ma che consente un recupero con un discreto livello di precisione a fronte della disparità di oggetti trattati.

Lo schema è stato definito originariamente nel 1995 a Dublin, nell'Ohio, ad opera di un gruppo di esperti provenienti dal mondo del trattamento dell'informazione, da tecnici informatici, bibliotecari, archivisti. Dublin Core è un insieme di 15 elementi, concepito come base di condivisione fra diverse applicazioni, essenzialmente al fine di trovare e recuperare il materiale su web. La sua funzione essenziale è comunemente definita infatti a livello internazionale con l'espressione 'resource discovery'.

Gli elementi sono: *Titolo, Soggetto, Descrizione* (o abstract), *Copertura, Fonte, Relazione, Lingua, Creatore, Autore secondario, Editore, Gestione dei diritti, Data, Formato, Tipo di risorsa, Identificatore*.

---

<sup>82</sup> Dublin Core Metadata Element Set : version 1.1 : reference description.  
<http://dublincore.org/documents/dces/>

Al fine di garantire funzioni essenziali come l'adozione più ampia possibile dello schema, l'adattabilità a risorse e contesti più vari ed anche la possibilità di raffinamenti ulteriori, nelle specifiche di applicazione dello schema è prevista la massima flessibilità di uso. Tutti gli elementi infatti sono opzionali, ripetibili e presentabili in qualsiasi ordine. Tradotto in più di 50 lingue (l'elenco multilingue degli elementi è visibile al sito della Dublin Core Metadata Initiative<sup>83</sup>), il formato è estensibile mediante la definizione di ulteriori elementi opportunamente identificati da un prefisso che ne indica lo schema di appartenenza. Ciò vale per aggiungere, in specifiche applicazioni che utilizzano Dublin Core, ulteriori metadati tipici di singoli settori disciplinari non coperti dallo schema elementare, ed anche metadati tecnici e amministrativi, utili per la gestione delle risorse. Questa tecnica si realizza con i profili di applicazione, trattati più avanti.

La necessità di esprimere in modo più preciso certi valori identificati dagli elementi ha spinto alla definizione dei 'qualificatori'. Ciò è apparso utile specialmente (ma non solo) nei casi in cui un record Dublin Core è prodotto mediante conversione a partire da record altamente strutturati e creati secondo schemi molto analitici (un caso tipico è costituito dai record MARC trasformati in Dublin Core per condividere informazioni comuni in progetti di integrazione di risorse digitali). Lo schema di base formato dai 15 elementi è stato arricchito dalla possibilità di esprimere due tipi di qualificatori per rendere più precisi, per raffinare e contestualizzare gli elementi stessi: a) i qualificatori di schema; b) i qualificatori di raffinamento.

E' utile sottolineare l'intento che sta alla base della predisposizione dei qualificatori: rendere più precisi i valori espressi negli elementi da chi crea i metadati, ma non di estenderne il significato. Infatti se questo principio non fosse rispettato, l'interoperabilità ne sarebbe compromessa.

Si distinguono così due modelli: Dublin Core semplice e Dublin Core qualificato. Al momento i qualificatori Dublin Core approvati ufficialmente sono 51.

---

<sup>83</sup> DCMI. <http://dublincore.org>

I qualificatori di schema specificano gli schemi o i vocabolari controllati utilizzati per esprimere il valore assegnato ad un elemento. Un esempio è rappresentato da: "Subject.classification.DDC. 330.945". Il qualificatore DDC (riferito alla Classificazione Decimale Dewey) nel campo *Subject* consente di interpretare al meglio il simbolo numerico, di personalizzare le interfacce di ricerca, di selezionare opportunamente gruppi di risorse sulla base del sistema di indicizzazione.

I qualificatori di raffinamento forniscono precisazioni sul tipo e la natura di un valore contenuto in un elemento. Nell'esempio immediatamente sopra "Classification" è un qualificatore di raffinamento dell'elemento *Subject*. Riferito all'elemento *Title* può specificare se si tratta di un titolo alternativo (Es: Title.alternative). Riferito all'elemento *Date* può precisare che si tratta della data di creazione, oppure della data di disponibilità o di modifica (Es.: Date.created. 2002-10-25).

Ai fini delle scelte da operare sugli schemi da utilizzare in progetti di biblioteca digitale occorre quindi comprendere chiaramente la natura e il significato di Dublin Core, a volte fraintesi. Questo schema elementare è in realtà qualche cosa di più di un insieme di 15 elementi per descrivere le risorse. E' un vero e proprio linguaggio, anche se semplice, che consente di fare delle dichiarazioni sugli oggetti digitali. Come sopra evidenziato il vocabolario di Dublin Core contiene due classi: gli elementi e i qualificatori. Nelle dichiarazioni fatte per descrivere le risorse questi hanno la funzione che nel linguaggio naturale hanno rispettivamente i nomi e gli aggettivi. Gli aggettivi possono essere ignorati da un'applicazione che vuole limitarsi al trattamento e alla comparazione fra elementi, oppure possono essere opportunamente trattati sfruttandone il valore semantico. Il significato di un vocabolario ristretto quale è quello rappresentato da Dublin Core è di essere facile da apprendere e utilizzare, nella convinzione che la limitatezza e la genericità degli elementi faciliti l'interoperabilità e che una maggiore standardizzazione sarebbe irrealistica nel mondo diversificato del web.

### 5.2.5 La produzione dei metadati

La produzione dei metadati è un'attività che ha inevitabilmente un costo. Qualsiasi progetto di biblioteca digitale si trova di fronte al problema della creazione, derivazione, conversione dei metadati utili per garantire l'accesso alle risorse trattate. Oltre a questioni di natura tecnica intervengono ovviamente anche considerazioni di ordine economico, di disponibilità di mezzi e risorse professionali e le soluzioni sono naturalmente diverse secondo il tipo degli oggetti trattati, i servizi da fornire, il livello di utenza a cui questi sono rivolti e ovviamente i mezzi a disposizione.

Schemi ricchi e dettagliati, specifici di determinate risorse e settori disciplinari come ad esempio CIDOC-CRM per gli oggetti museali, EAD per il materiale archivistico, il MARC, utilizzato per la preparazione e scambio di record bibliografici e adottabile per la descrizione delle risorse digitali scelte dalla biblioteca sulle quali essa si impegna a garantire una descrizione analitica, un controllo e un accesso efficace, sono complessi da preparare e richiedono un'elevata specializzazione professionale.

Una prassi comune è quella di ricorrere alla conversione di record preparati con schemi analitici verso record Dublin Core, ai fini dell'integrazione di risorse che altrimenti sarebbe compromessa. Molti progetti, di cui una selezione viene illustrata più avanti, utilizzano questa tecnica.

Un altro sistema, utilizzato da diversi progetti che necessitano di rendere disponibili le proprie risorse descritte in modo analitico, è quello, nel caso si disponga di record Dublin Core relativi agli oggetti digitali di interesse, di convertirli nello schema specifico predisposto per quel determinato settore disciplinare, ricorrendo alla tecnica delle mappature, operando gli opportuni controlli, integrazioni e adattamenti.

I metadati descrittivi, come del resto i metadati amministrativi, gestionali e tecnici, non sono dati statici, ma in evoluzione. Ciò è dovuto non solo alla natura mutevole delle risorse, ma anche alla diversità di applicazioni in cui

gli oggetti digitali possono essere trattati. La preparazione dei metadati può avvenire al momento della creazione degli oggetti digitali, anche da parte degli autori stessi, oppure in un momento successivo, quando le risorse sono distribuite in rete da fornitori commerciali o istituzionali che ne organizzano la ricerca e l'accesso. E' frequente infatti il caso di applicazioni che procedono alla raccolta e selezione di oggetti digitali, integrandoli con metadati appropriati per offrire servizi a valore aggiunto. La preparazione dei metadati da parte degli autori, che è una delle ragioni, anche se non la principale, che ha inizialmente spinto alla definizione dello schema elementare Dublin Core, è una delle possibilità operative, specie in un contesto di *self-archiving* delle proprie risorse digitali in rete. Esistono alcune esperienze nel mondo accademico, ma questo metodo ha ancora oggi una scarsa diffusione e diverse indagini sull'argomento dimostrano una certa riluttanza da parte degli autori alla preparazione in proprio dei metadati<sup>84</sup>.

Alcuni metadati possono essere prodotti automaticamente: si tratta di dati gestionali o tecnici derivati dal software e dall'hardware utilizzati, ma anche alcuni elementi semantici possono essere generati da opportuni algoritmi di riconoscimento sulla base del contenuto delle risorse.

In applicazioni avanzate, come ad esempio quelle costituite da portali specializzati, *subject gateways* e servizi istituzionali, oggi i metadati sono per la maggior parte prodotti da esperti del settore che conoscono caratteristiche e contenuto delle risorse, da catalogatori professionisti e da addetti a servizi di indicizzazione di tipo commerciale. Questi sono accomunati dal fatto di operare in un ambiente, quello del web, che li mette in comunicazione, in quanto la rete è insieme il contenitore e lo strumento per l'accesso e la gestione dell'informazione, ma la diversità delle risorse disponibili, degli schemi utilizzati, delle esigenze e del livello di analiticità nella descrizione utile in differenti contesti richiede che questa varietà sia adeguatamente gestita e controllata. Gli strumenti predisposti a questo fine sono al momento il linguaggio comune offerto da Dublin Core nella definizione degli attributi degli oggetti digitali, sistemi architetturali

---

84 Semantic web construction : an inquiry on authors' views on collaborative metadata generation / Jane Greenberg. <http://www.bncf.net/dc2002/program/ft/paper5.pdf>

per l'espressione delle proprietà secondo RDF e sistemi di codifica strutturata dei record di metadati.

Esistono oggi diversi sistemi per la creazione e gestione dei metadati: molti progetti hanno creato i loro programmi di immissione e controllo sulla base degli schemi o profili adottati per la descrizione delle risorse trattate. Un esempio è costituito dal progetto Cooperative Online Resource Catalogue (CORC), illustrato più avanti. Sono disponibili anche software di pubblico dominio, come ad esempio Dublin Core metadata editor, conosciuto con il nome DC.dot<sup>85</sup>, messo a punto dal Centro UKOLN di Bath, in Gran Bretagna, che può essere utilizzato in rete o scaricato in locale per la preparazione di record Dublin Core. A seguito dell'immissione, da parte dell'operatore, dell'identificatore di una risorsa (attualmente è usato l'indirizzo di rete: URL (Uniform Resource Locator), il programma produce automaticamente un record Dublin Core con alcuni elementi riconosciuti a partire dalla struttura (generalmente in formato HTML) di un sito o documento. Funzioni di *editing* consentono poi di aggiornare il record, integrando o correggendo quanto del record è stato prodotto automaticamente. Si segnala inoltre il programma DC-assist<sup>86</sup>, anch'esso a cura del Centro UKOLN, che è una semplice applicazione, utilizzabile in linea, che offre assistenza nell'interpretazione del significato degli elementi, utile al catalogatore per l'assegnazione dei metadati distintivi di una risorsa.

Inoltre merita qui di essere ricordato un sistema messo a punto in Australia per l'indicizzazione e il recupero di materiale audiovisivo, denominato Veggie<sup>87</sup>, adatto per la produzione dei metadati di risorse digitali multimediali, che consente la generazione semi-automatica di un certo numero di elementi tipici dei documenti audiovisivi.

---

85 DC.dot. [www.ukoln.ac.uk/metadata/dcdot/](http://www.ukoln.ac.uk/metadata/dcdot/)

86 Dc assist. <http://www.ukoln.ac.uk/metadata/dcassist/>

87 Veggie : an indexing, browsing, search and retrieval system for audiovisual libraries /Jane Hunter. <http://archive.dstc.edu.au/RDU/staff/jane-hunter/INFOG99/paper.html>

Il sito web della DCMI segnala diversi strumenti<sup>88</sup> per la generazione automatica dei metadati e per la loro creazione esplicita, nonché per la conversione fra schemi di metadati.

E' recente la messa a punto di un prodotto particolare, Extensible Open RDF (EOR)<sup>89</sup>, nato per sviluppare una interfaccia di ricerca integrata su più risorse strutturate secondo l'architettura RDF, offrendo la possibilità di creare e gestire database costruiti a partire da record strutturati secondo le raccomandazioni RDF.

### **5.2.6 I sistemi di codifica dei record di metadati**

Per poter essere interpretati ed elaborati, i record di metadati devono avere una loro sintassi ed essere quindi strutturati in modo da essere riconosciuti in quanto appartenenti ad uno schema predefinito, a cui sia possibile accedere sia dall'uomo che da applicazioni. E' infatti necessario conoscere in modo preciso gli elementi e le loro caratteristiche, controllare la validità dei metadati secondo le regole fissate nello schema a cui appartengono, compararli con elementi di altri schemi per poi trasferirli in database per la gestione e l'accesso.

Mentre i record MARC sono ancora oggi strutturati in gran parte, ai fini del trasferimento e della comunicazione fra applicazioni, secondo le specifiche stabilite dalla norma ISO 2709 che prescrive una organizzazione dei dati (per dati si intende qui etichette identificative dei campi ed elementi propriamente bibliografici) orientata al trasferimento su supporti di tipo sequenziale (come ad esempio i nastri magnetici), tale organizzazione non è adatta per i record di metadati, che necessitano di una struttura di trasporto orientata alla comunicazione e all'elaborazione in rete. Anche nella comunità delle biblioteche dove il MARC ha un'enorme diffusione è sempre più forte la convinzione che i record MARC dovranno essere memorizzati secondo una struttura diversa, auspicabilmente in XML, di cui esistono già diverse applicazioni.

---

<sup>88</sup> DCMI : tools and software. <http://dublincore.org/tools/>

<sup>89</sup> The EOR toolkit. <http://eor.dublincore.org/index.html>

Oggi i record di metadati sono generalmente strutturati e codificati secondo il linguaggio di marcatura HTML e solo recenti applicazioni adottano la codifica in XML. L'etichettatura secondo le specifiche XML inizia a fare la sua comparsa all'interno di dichiarazioni RDF riferite alle proprietà assegnate alle risorse.

**HTML (Hypertext Mark up Language).** Dalla metà degli anni '90 fino ad oggi l'uso di HTML per etichettare i record di metadati è molto diffuso. Ancora oggi non è chiaro quale dei due sistemi, HTML o XML, prevarrà. Molti dei metadati prodotti per descrivere le risorse web è attualmente incorporata all'interno delle risorse stesse ed è codificata in HTML: il valore di questo approccio è la semplicità, ma il sistema consente scarsa flessibilità per la limitatezza e rigidità con cui le proprietà o attributi possono essere codificati: infatti i metadati e i relativi valori sono definiti all'interno dello specifico attributo HTML denominato '*meta content*'. Inoltre, adottando questa struttura, la manutenzione dei record risulta piuttosto complessa.

Seguono tre brevi esempi di codifica HTML di elementi e qualificatori all'interno di record Dublin Core.

a) Codifica dell'elemento (definibile anche proprietà, o attributo) *Date* in un record Dublin Core semplice: `<meta name="DC.date" content="2002-10-27" />`

b) Codifica di un qualificatore di raffinamento dell'elemento *Date*: `<meta name="DC.date.modified" content="2002-10-29" />`.

c) Codifica di un qualificatore di schema dell'elemento *Date*: `<meta name="DC.date." scheme="W3CDTF" content="2002-10-29" />`.<sup>90</sup>

Un record preparato secondo uno o più schemi di metadati deriva gli elementi dallo schema o schemi utilizzati e ciò deve essere dichiarato in testa al record, indicando l'identificatore persistente che consente di accedere alla lista dei metadati.

---

<sup>90</sup> Gli elementi e i qualificatori devono essere espressi con i nomi definiti ufficialmente nello schema e visibili al sito: <http://dublincore.org/usage/terms/dc/current-elements/>.

Ad esempio:

```
<link rel="schema.DC"
href=http://purl.org/dc/elements/1.1./" />
<link rel="schema.DCTERMS"
href=http://purl.org/dc/terms/" />
```

Un record che riporta al suo inizio tali dichiarazioni indica che i metadati sono derivati da DC e DCTERMS: questi sono i due prefissi che indicano i *namespace* (letteralmente: spazi di nomi), identificati da URI (Uniform Resource Identifier) con caratteristica di identificatori persistenti. Gli spazi di nomi sono i 'luoghi' sul web dove risiedono le descrizioni normalizzate dello schema dei metadati utilizzati nei record, ai quali si indirizzano le applicazioni per interpretare i metadati stessi.

Viene fornito un esempio di record di metadati Dublin Core in formato HTML, ripreso dal catalogo CORC, descritto nella sezione dedicata ai progetti internazionali.

```
<link rel="schema.DC" href=http://purl.org/dc/elements/1.1" />
<meta name="DC.title" content="Churches in Florence"/>
<meta name="DC.identifier"
content="http://www.arca.net/tourism/florence/churches.htm" />
<meta name="DC.type" content="Text data" />
<meta name="DC.type" content="[computer file]" />
<meta name="DC.type" content="World Wide Web Resource" />
<meta name="DC.description" content="Examines churches in Florence, Italy,
including the chartreuse, San Firenze, and the Badia, which is the oldest
monastery in Florence. Contains a history of each building and describes any
reconstruction work. Highlights the artwork within the buildings and provides
brief background information on the artists.
Features photographs of selected churches. Notes the address of each church
and links to a map of Florence. Offers access to the Your Way to Florence home
page, which includes information on transportation, accommodations, and the
arts in Florence" />
<meta name="DC.language" content="eng" />
<meta name="DC.publisher" content="Bonechi Casa Editrice" />
```

**XML (eXtensible Mark up Language).** E' un modello per strutturare risorse elettroniche in modo uniforme e indipendente dai dati che si

intende organizzare, siano essi dati di un documento o altro tipo di risorsa. XML usa il testo (e quindi delle etichette) per strutturare i dati, ma non prescrive i nomi degli elementi in quanto è un insieme di istruzioni da usare per definire gli elementi stessi, che saranno definiti dall'applicazione. L'uso dello standard per i record di metadati non è attualmente molto diffuso, ma le regole di marcatura XML bene si prestano al loro impiego in record di metadati per la possibilità offerta di codificare dati secondo una struttura gerarchica e definire etichette con funzione di marcatori secondo le necessità delle varie applicazioni, separando il contenuto dai marcatori, utili per funzioni diverse, come ad esempio la presentazione dei dati.

La proprietà eXtensible presente nel nome dello standard sta ad indicare la definizione di un linguaggio di contrassegni, e quindi di metadati, personalizzati. Generalmente un record di metadati in formato XML risiede esternamente alla risorsa descritta, in un database apposito, ma è ovviamente ad essa collegato mediante un puntatore.

La comunità degli implementatori di Dublin Core ha prodotto nove raccomandazioni per la codifica di Dublin Core in XML. Queste sono contenute in specifiche linee guida<sup>91</sup>:

**RDF.** Come sopra accennato, con l'espressione **Resource Description Framework** si intende un insieme di convenzioni messe a punto dal Consorzio W3C per consentire alle applicazioni di riconoscere e scambiare i metadati attribuiti alle risorse e mettere in tal modo in relazione le informazioni residenti sul web.

Un punto chiave di questa metodologia si basa sulla visione del web semantico, in cui comunità diverse comunicano fra loro sulla base di una reciproca comprensione dei loro dati, e quindi delle proprietà assegnate alle varie risorse sulla base degli schemi di metadati utilizzati.

Per realizzare questa reciproca comprensione due condizioni sono necessarie:

- a. adottare un linguaggio standard e quindi una sintassi comune con cui dichiarare le proprietà assegnate alle risorse

---

<sup>91</sup> Guidelines for implementing Dublin Core in XML".  
<http://dublincore.org/documents/2002/09/09/dc-xml-guidelines/>

- b. garantire delle corrispondenze fra la semantica dei diversi metadati, interpretando correttamente il significato dei metadati assegnati alle risorse e stabilendo delle corrispondenze logiche fra i termini utilizzati nei vari schemi.

Per quanto riguarda la sintassi delle dichiarazioni sulle proprietà delle risorse, la proposta RDF si basa su un semplice modello, composto da tre parti: il metadato si riferisce ad una *risorsa*, la risorsa ha una *proprietà*, ed ogni proprietà ha un *valore*. In altri termini il modello mette in relazione tre elementi: *Soggetto - Predicato - Oggetto*.

Relativamente alla comprensione della semantica dei metadati assegnati alle risorse e alla loro comparazione nel contesto diversificato del web, occorre precisare che si tratta di condizioni essenziali per sfruttare la ricchezza dei significati espressi dai metadati e che tale comprensione è resa possibile dalla presenza, in ogni dichiarazione RDF, del riferimento univoco tramite URI ai diversi *namespace* relativi agli schemi di metadati utilizzati, in quanto luoghi nella rete in cui sono memorizzati gli elementi usati per definire le proprietà assegnate alle risorse.

Inoltre RDF offre la possibilità di interpretare gli elementi dei vari schemi in un modo particolare, dichiarando le relazioni di corrispondenza e dipendenza tra i vari elementi.

Segue un semplice esempio di dichiarazione RDF relativa alla paternità del sito web dell'Istituto centrale per il catalogo unico. La dichiarazione letteralmente è espressa in questo modo: "http://www.iccu.sbn.it" (*risorsa* con funzione di *soggetto*) è stata creata (*proprietà* con funzione di *predicato*) dall'Istituto centrale per il catalogo unico (*valore* con funzione di *oggetto*).

La dichiarazione RDF è la seguente:

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc=http://purl.org/dc/elements/1.1/" >
  <rdf:Description about="http://www.iccu.sbn.it">
    <dc:creator> Istituto centrale per il catalogo unico </dc:creator>
  </rdf:Description>
</rdf:RDF>
```

In questo semplice esempio è riportata la versione XML usata, sono segnalati i due *namespace* utilizzati, indicando rispettivamente la URL dove risiede la specifica RDF e l'URI (l'URL persistente) relativo allo schema dei 15 elementi di Dublin Core. Viene inoltre espressa la tipica dichiarazione standard RDF ('Description about'), seguita dall'URL della risorsa descritta, che è il sito web dell'ICCU. Segue, conformemente al *namespace* dello schema Dublin Core richiamato in testa al record, l'elemento *creator*, preceduto dal prefisso dc (Dublin Core) con lo specifico valore: "Istituto centrale per il catalogo unico".

RDF si configura quindi come una metodologia per dichiarare le proprietà delle risorse, secondo una sintassi e precise regole che sono sfruttate da opportuni programmi per realizzare una serie di funzioni, quelle tipiche del web semantico, come ad esempio integrare le risorse fra di loro e fornire servizi di accesso unificato, riutilizzare i metadati assegnati a risorse diverse comparandoli in applicazioni costruite per scopi specifici, permettere la navigazione e la visualizzazione secondo aggregazioni e categorizzazioni che sfruttino il valore semantico contenuto nelle proprietà attribuite alle varie risorse.

Nell'ambiente del web ogni oggetto trattato, sia esso una risorsa, un metadato o proprietà ad essa attribuita, è identificabile da un URI, che deve essere un identificatore univoco e persistente all'interno del contesto globale del web. E' importante notare infatti che anche i singoli metadati o

proprietà, una volta codificati in uno schema di cui sia stata fornita in rete la codifica appropriata (generalmente secondo il linguaggio XML), hanno un loro identificatore, rappresentato dal relativo URI, configurandosi quindi come vere e proprie risorse. Ad esempio l'elemento *Title* dello schema Dublin Core, in quanto esistente nello specifico *namespace* riservato a Dublin Core, è identificato dall'URI: <http://purl.org/dc/elements/title>. Ciò consente a opportuni programmi di riconoscere e comparare il metadato 'title' con altri metadati, stabilendo relazioni fra documenti e navigando fra informazioni residenti nella ragnatela dei dati. In questa ottica *Semantic web* è sinonimo di *web of data*.

Oggi le applicazioni di RDF sono essenzialmente di tipo sperimentale, sviluppate in progetti di biblioteca digitale e nell'area della ricerca accademica, ma recentemente lo standard sta emergendo in applicazioni operative nel mondo aziendale e commerciale.

Nella pagina seguente viene fornito un esempio di record di metadati Dublin Core secondo la struttura XML-RDF (si tratta dello stesso record che precedentemente è stato esemplificato in formato HTML).

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcq="http://purl.org/dc/terms/">
<rdf:Description about="http://www.arca.net/tourism/florence/churches.htm">
<dc:title>Churches in Florence.</dc:title>
<dc:publisher>Bonechi Casa Editrice.</dc:publisher>
<dc:description>Examines churches in Florence, Italy, including the Chartreuse, San Firenze, and the
Badia, which is the oldest monastery in Florence. Contains a history of each building and describes any
reconstruction work. Highlights the artwork within the buildings and provides brief background information
on the artists. Features photographs of selected churches. Notes the address of each church and links to a
map of Florence. Offers access to the Your Way to Florence home page, which includes information on
transportation, accommodations, and the arts in Florence.</dc:description>
<dc:identifier>http://www.arca.net/tourism/florence/churches.htm</dc:identifier>
<dc:language>eng</dc:language>
<dc:subject> <rdf:Description> <dcq:subjectQualifier>class</dcq:subjectQualifier>
<rdf:value>726.5</rdf:value> </rdf:Description> </dc:subject>
<dc:subject> <rdf:Description> <dcq:subjectQualifier>class</dcq:subjectQualifier>
<rdf:value>291.657</rdf:value> </rdf:Description> </dc:subject>
<dc:subject> <rdf:Description> <dcq:subjectQualifier>topical</dcq:subjectQualifier> <rdf:value>Church
buildings</rdf:value> </rdf:Description> </dc:subject>
<dc:subject> <rdf:Description> <dcq:subjectQualifier>topical</dcq:subjectQualifier>
<rdf:value>Monasteries</rdf:value> </rdf:Description> </dc:subject>
<dc:subject> <rdf:Description> <dcq:subjectQualifier>geographic</dcq:subjectQualifier>
<rdf:value>Florence (Italy : Province)</rdf:value> </rdf:Description> </dc:subject>
<dc:type>Text data</dc:type>
<dc:type>[electronic resource]</dc:type>
<dc:type>World Wide Web Resource</dc:type>
</rdf:Description>
</rdf:RDF>

```

### **5.2.7 Gli identificatori e la loro codifica nei record di metadati**

Fra i metadati gli identificatori sono gli elementi che svolgono una funzione fondamentale per recuperare e utilizzare le risorse elettroniche, in quanto forniscono, in linea di principio, una chiave unica di accesso al materiale digitale. Sono ancora molte però le questioni aperte perché essi siano effettivamente strumenti univoci e persistenti, capaci di garantire l'accesso preciso, controllato e permanente al contenuto delle risorse.

Ancora oggi l'indirizzo di una risorsa (URL) è impiegato diffusamente come identificatore, ma per definizione questa funzione non può essere compiuta in modo valido da un indirizzo; infatti nella rete localizzare non equivale a identificare, dal momento che ancora oggi la volatilità e la mutevolezza dei documenti web pongono seri problemi all'accesso ed è difficile in molti casi realizzare funzioni quali il controllo delle versioni, l'autenticità e il recupero puntuale e rapido di una risorsa.

Se da una parte è in corso un lavoro molto intenso di studio e preparazione di standard per l'identificazione di documenti a vari livelli di granularità dei contenuti informativi e per funzioni diversificate che vanno dal recupero al commercio elettronico, dall'altra si deve riconoscere che l'utilizzo degli identificatori non ha ancora raggiunto la diffusione e l'efficienza che invece sarebbero necessarie a chi opera in rete a vari livelli, in veste di produttori, fornitori e utenti di materiale digitale.

Ciò dipende dallo scarso coinvolgimento che ancora dimostrano comunità di primaria importanza che operano in rete, come ad esempio gli editori, dalla complessità organizzativa derivante dal funzionamento dei servizi di risoluzione (servizi che a partire dagli identificatori indirizzano automaticamente alle risorse, e ciò su scala internazionale), ma anche da una certa riluttanza da parte di organismi produttori di risorse ad adottare standard specifici come ad esempio SICI (Serial Item and Contribution Identifier, riferito agli articoli di periodici), BICI (Book Item and Component Identifier per singole parti di un libro come capitoli o paragrafi), ISTC (International Standard Textual Work Code per opere testuali), ISAN (International Standard Audiovisual Standard per documenti audiovisivi), ISWC (International Standard Musical Work Code per documenti musicali).

Perché gli identificatori svolgano la loro funzione di recupero puntuale di risorse è necessario che vi sia un sistema efficiente che colleghi una risorsa con la sua descrizione. Questi sistemi sono chiamati servizi di risoluzione: essi svolgono quella funzione che in un sistema tradizionale è realizzata, ad esempio, collegando in un catalogo in linea (OPAC) la descrizione bibliografica con la collocazione di un libro, la cui disposizione sugli scaffali sia mantenuta fissa.

Anche per le risorse elettroniche sono quindi necessari sistemi di risoluzione che colleghino una risorsa con la sua descrizione o che associno direttamente due risorse. Data la mutevolezza della localizzazione delle risorse digitali e l'esigenza di gestire l'accesso sulla base delle loro condizioni di uso, confrontate con i diritti dei singoli utenti, i servizi di risoluzione devono avere particolari caratteristiche. Essi non dovrebbero essere, come frequentemente accade oggi, semplicemente statici, e quindi basarsi solo sull'indirizzo della risorsa stessa (URL), bensì dinamici, adattandosi ai mutevoli indirizzi delle risorse, nello spazio e nel tempo, considerando i diritti dell'utente e indirizzando a servizi che siano personalizzati sul profilo di interesse dell'utente stesso. Vi sono diverse applicazioni e progetti in corso, rappresentati da OpenURL<sup>92</sup> e CrossRef<sup>93</sup> che fanno uso di tali sistemi di risoluzione: questi sviluppi devono essere seguiti con attenzione, considerando che il loro funzionamento si basa in modo decisivo sulla persistenza e univocità degli identificatori delle risorse trattate.

La Internet Engineering Task Force (IETF) da tempo ha riconosciuto l'esigenza di sviluppare un'architettura per l'identificazione uniforme delle risorse (URI) e su questa base sono stati messi a punto alcuni standard come Uniform Resource Name (URN) e Digital Object Identifier (DOI). Anche se la loro diffusione non è certo capillare, questi identificatori vantano oggi diverse applicazioni importanti.

---

92 Open linking in the scholarly information environment using the OpenUrl framework /Herbert Van de Sompel.

<http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html> ;

SFX e OPENURL : gli esperimenti del team di Van de Soempel / Cinzia Bucchioni:

<http://www.spbo.unibo.it/bibliotime/num-v-2/bucchion.htm>

93 CrossRef turns one / Amy Brand: <http://www.dlib.org/dlib/may01/brand/05brand.html>

URN nasce come identificatore persistente e indipendente dalla localizzazione di una risorsa e ha finora trovato il suo impiego in vari progetti, fra i quali iniziative istituzionali in vari paesi come ad esempio la Finlandia e l'Australia. L'adozione di questo standard comporta lo sviluppo degli schemi di nomi da assegnare alle risorse e da ascrivere in appositi registri (che possono anche ospitare schemi esistenti come ISBN o ISSN) e il funzionamento di sistemi di risoluzione che siano in grado, appunto, di risolvere gli URN, traducendoli nella localizzazione di una risorsa (avvalendosi anche di un meccanismo di re-indirizzamento, come ad esempio PURL: Persistent Uniform Resource Locator), oppure puntando ad un record bibliografico che la descrive, o ad un modulo di ordine per ottenere la risorsa stessa. Un sistema di risoluzione assai diffuso, accessibile in rete e sviluppato negli Stati Uniti, è Handle<sup>94</sup>.

La sintassi di un URN è formata dalla stringa 'urn', dall'identificatore dei nomi assegnati (NID: *Namespace Identifier*), dalla singola stringa di caratteri che identifica la risorsa (NISS: *Namespace Specific String*). Un esempio è fornito di seguito ed è basato sull'ISBN di un documento, che può quindi essere incorporato in un URN a testimonianza del fatto che la sintassi URN può ospitare schemi di identificazione esistenti: "urn:ISBN:8884530008". In questo caso si tratta dell'ISBN della risorsa elettronica dal titolo: Una biblioteca in divenire, di Tommaso Urso, pubblicata dalla Florence University Press, all'indirizzo: <http://www.epress.unifi.it>

Il principio che sta alla base dell'assegnazione di un URN ad una risorsa è quello per cui il sistema di risoluzione deve essere separato dal modo in cui sono assegnati i nomi alle risorse. Questa indipendenza fra gli schemi dei nomi attribuiti alle risorse e i sistemi di risoluzione è in contrasto con la natura stessa di un URL, che contiene al suo interno il suo meccanismo di risoluzione, in quanto indica il protocollo di accesso e l'elaboratore da contattare per accedere alla risorsa.

---

<sup>94</sup> Handle system. [www.handle.net/ietf/handle/register\\_handle.html](http://www.handle.net/ietf/handle/register_handle.html)

DOI è un numero univoco assegnato ad un oggetto digitale, anch'esso, in via di principio, con caratteristiche di permanenza, destinato all'identificazione della proprietà intellettuale di una risorsa. Come l'ISBN che identifica una manifestazione di un'opera, DOI rappresenta la sequenza di bit di un documento elettronico in modo permanente, ma può essere assegnato a singole unità informative, come singole parti elementari di un documento.

L'assegnazione di questo identificatore è fatta in collaborazione fra l'editore e una delle cinque agenzie di registrazione che ad oggi operano in raccordo con la International DOI Foundation (IDF). DOI è infatti formato da due componenti: una è costituita da un prefisso che indica l'agenzia di registrazione e l'editore, l'altra da una stringa attribuita dall'editore.

Un esempio di DOI è costituito da: 10.1045/january99-bearman dove 10 indica che all'interno del sistema Handle la stringa DOI sarà risolta dal sistema Handle riservato per il DOI. Segue il prefisso assegnato all'editore e infine una directory nel database dell'editore.

La conoscenza del DOI di un articolo, ad esempio, consente la sua localizzazione in modo permanente grazie al meccanismo di comunicazione fra l'editore e il sistema di risoluzione (in questo caso il sistema Handle). E' evidente che l'efficienza del sistema dipende da aspetti organizzativi e quindi dai meccanismi di scambio fra i partecipanti al sistema e dall'impegno dell'editore a comunicare cambiamenti di indirizzo o di attività.

Al di là delle questioni aperte sulla persistenza ed univocità degli identificatori, che rappresenta una problematica a sé rispetto all'area dei metadati, a titolo di chiarimento vengono illustrati alcuni esempi di codifica di identificatori in record di metadati. Il riferimento è a record Dublin Core in formato HTML. Gli identificatori sono usati generalmente nei seguenti elementi:

- a. Identificatore (l'identificatore si riferisce alla risorsa descritta)
- b. Relazione (l'identificatore si riferisce ad una risorsa collegata a quella descritta)

- c. Fonte (l'identificatore si riferisce ad una risorsa che è la fonte da cui nasce la risorsa descritta)
- d. Diritti (l'identificatore si riferisce ad una risorsa di riferimento che chiarisce i diritti sulla risorsa descritta).

URI rappresenta lo schema da indicare a fianco di un identificatore presente nel record di metadati.

Ad esempio, nel record Dublin Core in formato HTML relativo al sito web del portale della Biblioteca Digitale Italiana, i due elementi costituiti da Identificatore (contenente l'URI del portale) e Relazione (contenente l'URI dello Studio di fattibilità della BDI, qui considerato come risorsa in relazione a quella costituita dal portale) saranno codificati nel modo seguente:

```
<meta name="DC.identifier" scheme="URI"
content="http://www.bditaliana.it/" >
<meta name="DC.relation" scheme="URI"
content="http://www.bditaliana.it/appl2/hometree.htm" > .
```

In questo caso i valori dell'identificatore utilizzato nei due elementi Identificatore e Relazione preannunciato dallo schema URI sono in ambedue i casi un URL.

Nel caso che il sito del portale abbia come fonte un documento analogico o digitale identificato da un ISBN, l'elemento fonte avrà la seguente codifica:

```
<meta name="DC.Source" scheme="URI" content="urn:isbn:.....">
```

In questo caso l'ISBN del documento fonte della risorsa descritta è codificato come URN.

### **5.3 Metadati gestionali, amministrativi e strutturali**

Un progetto di biblioteca digitale si fonda dunque sulla produzione e disponibilità di metadati per la gestione degli oggetti informativi che ne

costituiscono la materia prima. Tali oggetti, per essere identificati, recuperati, selezionati, accessibili e conservabili per l'uso nel tempo, devono essere associati, oltre alle informazioni di carattere descrittivo, anche a quelle di tipo gestionale, necessarie per assolvere a queste funzioni.

Nell'ambito della biblioteca tradizionale i record bibliografici relativi al materiale trattato contengono gli elementi per l'identificazione, il recupero e la selezione del materiale descritto. Le altre informazioni di tipo gestionale sono in parte contenute nel record stesso (collocazione e inventario), in parte residenti su archivi e strumenti diversi: inventari, schedari relativi allo stato di conservazione e a interventi di restauro, alla disponibilità delle pubblicazioni, alle regole di utilizzo e circolazione, etc. Anche se nei moderni sistemi di automazione utilizzati dalla maggior parte delle biblioteche tali informazioni sono fra loro collegate, l'integrazione in un insieme coerente di dati descrittivi e gestionali relativi ad un documento avviene all'interno di una biblioteca o al massimo di un sistema cooperativo.

Il contesto della biblioteca digitale è più ampio e variegato, vi è una maggiore e più articolata mole di dati da trattare per le diverse funzioni di gestione delle risorse elettroniche, che vengono eseguite essenzialmente in modo automatico. Gli oggetti digitali sono accessibili in un contesto distribuito e si rende necessario controllare la loro autenticità, l'effettiva disponibilità a fronte di un identificatore ad esse associato, consentire il loro utilizzo sulla base di condizionamenti tecnici e giuridici e garantirne la conservazione e l'uso nel tempo. Per questa ragione è necessario da una parte uno stretto collegamento fra tutte le informazioni che sono utili al trattamento delle risorse (fra metadati descrittivi e gestionali ed anche fra i vari tipi di metadati gestionali), dall'altra un buon livello di standardizzazione nella definizione degli elementi che identificano le caratteristiche tecniche ed amministrative degli oggetti digitali.

La tipologia dei metadati gestionali è quindi molto varia: un certo numero di essi sarà prodotto automaticamente da opportuni programmi che interpretano valori predefiniti collegati ad attività che vengono svolte nelle

varie fasi del ciclo di vita degli oggetti digitali, altri saranno immessi esplicitamente sulla base di esigenze e funzioni fissate dalle singole applicazioni.

L'attività a livello internazionale per la definizione dei dati necessari alla gestione delle risorse digitali si è indirizzata verso due fronti, in certo modo fra loro complementari: da una parte l'elaborazione di un modello di riferimento da adottare per la gestione di archivi digitali, con particolare attenzione alla conservazione, denominato OAIS: Open Archival Information System<sup>95</sup>. Dall'altra la definizione di metadati gestionali da parte di istituzioni e progetti di biblioteca digitale a cui gradualmente le varie comunità impegnate in esperienze simili tendono a conformarsi. Si segnalano a questo scopo:

- a. Le attività della Library of Congress, con lo sviluppo e la sperimentazione dello schema METS (Metadata Encoding & Transmission Standard)<sup>96</sup>
- b. Le proposte del gruppo congiunto OCLC-RLG (Online Computer Library Centre - Research Libraries Group) sui metadati per la conservazione digitale che ha recentemente finalizzato uno schema di metadati per la conservazione<sup>97</sup> che si basa sul modello OAIS. Tale lavoro ha le sue radici nel Rapporto sui metadati per la conservazione (denominato 'Libro bianco'), preparato dal gruppo di lavoro stesso nel gennaio 2001<sup>98</sup>.
- c. Oltre a queste due basilari iniziative molto lavoro è in corso da parte di istituzioni e progetti per la definizione e standardizzazione delle funzionalità e dei metadati necessari per controllare il ciclo di vita degli oggetti digitali e consentire lo svolgimento delle relative funzioni. Alcuni esempi sono forniti dalle sperimentazioni e proposte di insiemi di metadati definiti da progetti e istituzioni quali CEDARS (CURL Exemplars in Digital Archives), NLA (National Library of

---

95 Reference Model for an Open Archival Information System:CCSDS 650.0-R-2. Issue 2.June 2001. [http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)

96 METS Schema. <http://www.loc.gov/standards/mets/>

97 Preservation metadata, June 2002. <http://www.oclc.org/research/pmwg/>

98 Preservation metadata for digital objects : a review of the state of the art : a White Paper by the OCLC/RLG Working Group on Preservation Metadata. [http://www.oclc.org/research/pmwg/presmeta\\_wp.pdf](http://www.oclc.org/research/pmwg/presmeta_wp.pdf)

Australlia), NEDLIB, A-Core, DEF (Danmarks Elektroniske Forskningsbibliotek).

Va ricordato qui, relativamente all'impegno italiano, il lavoro del gruppo per la definizione di un set di metadati amministrativi, gestionali e strutturali creato dall'Istituto centrale per il catalogo unico<sup>99</sup>.

Inoltre sulla base del modello OAIS, dell'insieme di metadati definiti in METS ed anche del set di metadati per la conservazione messo a punto dal gruppo OCLC/RLG, la Biblioteca nazionale centrale di Firenze ha preparato delle specifiche per la fase di archiviazione dei metadati e dei dati gestionali, venendo incontro alle esigenze di chi implementa progetti. Sono stati esaminati quindi gli standard emergenti ed è stata utilizzata l'esperienza acquisita nel progetto di digitalizzazione di frontespizi e testi ARSBNI, da tempo operativo presso la BNCF. Sulla base di questo lavoro è stato elaborato uno schema XML<sup>100</sup> con sezioni apposite dove sono contenute le informazioni necessarie alla gestione del ciclo di vita degli oggetti digitali. In particolare sono state individuate le sezioni relative ai dati che documentano le caratteristiche generali del progetto che si sta attuando e il tipo di digitalizzazione, ai metadati descrittivi, a quelli relativi alle immagini fisse e al riconoscimento ottico del testo.

Sembra opportuno richiamare l'attenzione sulle caratteristiche principali di alcuni degli schemi e modelli sopra delineati, che sono oggi punto di riferimento essenziale per la definizione dei metadati amministrativi gestionali.

Il modello OAIS<sup>101</sup> (Open Archival Information System, tradotto dal Gruppo di lavoro italiano istituito dall'ICCU come 'Sistema Informativo Aperto per l'Archiviazione'), definisce le componenti essenziali dell'attività di immissione, archiviazione, gestione, accesso, amministrazione dell'archivio, pianificazione della conservazione (funzione ulteriore,

---

<sup>99</sup> Appunti per la definizione di un set di metadati gestionali amministrativi.

<http://www.iccu.sbn.it/metaAG1.pdf>

<sup>100</sup> MAG Schema. <http://www.bncf.firenze.sbn.it/progetti/mag/>

<sup>101</sup> Reference model for an Open Archival Information System (OAIS).

<http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>

suggerita dal progetto NEDLIB, rispetto alla proposta iniziale di OAIS). Queste attività, tipiche di un progetto di biblioteca digitale, sono concepite in modo distribuito e le entità che vi partecipano sono individuate nei produttori, nei consumatori, nel management e nell'archivio digitale.

Al centro dell'architettura prevista da OAIS vi è l'oggetto informativo, che è composto da dati (un insieme di bit) e da informazioni sulla rappresentazione (metadati che servono per rendere comprensibile l'oggetto e per potervi accedere). I metadati in questo contesto sono di vario tipo, relativi alla struttura della risorsa digitale, alla sua localizzazione, alle caratteristiche tecniche, al formato, al software necessario per l'accesso, al contenuto (per consentirne la ricerca e l'identificazione), alla conservazione. All'interno dell'informazione descrittiva per la conservazione il modello OAIS distingue dettagliatamente le informazioni per l'identificazione, il contesto, la provenienza e l'autenticazione delle risorse.

E' utile sottolineare la specifica natura di OAIS, che è un modello logico di riferimento, un insieme di specifiche ad alto livello che prevede dunque queste componenti essenziali:

- l'interazione delle entità sopra ricordate, le quali intervengono nei processi per la costruzione e gestione degli archivi digitali,
- le sei componenti funzionali sopra ricordate, che vanno dall'immissione fino alla pianificazione della conservazione,
- gli oggetti informativi, i quali compongono i vari pacchetti informativi (*information packages*), che circolano durante la fase di scambio di dati da e verso l'archivio. Questi pacchetti contengono le informazioni per le varie attività come l'acquisizione dei files (dati per l'immissione nell'archivio da parte dei produttori di materiale digitale), l'archiviazione (dati necessari per la conservazione a lungo termine), la distribuzione degli oggetti digitali (dati per l'accesso a seguito di una richiesta da parte dell'utente).

Il modello funzionale OAIS non si occupa di definire i metadati, ma rappresenta il contesto e la cornice in cui le attività funzionali di un archivio digitale si svolgono grazie ai metadati. Questi sono definiti ad un

altro livello, ad opera di istituzioni con un ruolo guida e da singoli progetti che forniscono specifiche sui singoli elementi necessari per garantire la gestione degli oggetti digitali da loro trattati.

Di seguito vengono presentate alcune iniziative determinanti in questo settore.

Lo schema METS fornisce un flessibile meccanismo per codificare metadati descrittivi, amministrativi e strutturali per oggetti digitali e per esprimere i complessi collegamenti fra questi metadati. Elaborato a seguito del progetto americano American Memory da parte della Library of Congress, METS è formato da quattro sezioni che includono rispettivamente:

i metadati descrittivi. Possono essere residenti al di fuori del documento METS, ad esempio in un OPAC, o all'interno della risorsa

i metadati amministrativi. I quattro tipi previsti sono metadati tecnici (informazioni sulla creazione dei file, formato e caratteristiche sull'uso), metadati sulla proprietà intellettuale (informazioni su copyright e licenze), metadati sulla fonte (metadati descrittivi e amministrativi sull'oggetto analogico da cui un oggetto digitale deriva), metadati sulla provenienza digitale (informazioni sulla relazione fra files fonte e di destinazione, su migrazioni e trasformazioni dell'oggetto digitale). Anche in questo caso i metadati possono essere residenti al di fuori del documento METS o al suo interno

Elementi relativi a gruppi di file. Questi elementi si riferiscono ai file che compongono una singola versione elettronica di un oggetto digitale e che devono essere messi in relazione fra loro (ad esempio ci possono essere singoli elementi per le immagini in formato ridotto e per quelle in formato esteso, per le versioni pdf, etc.).

Mappa strutturata. Sono definite singole parti, opportunamente codificate e gerarchicamente strutturate, che sono fra loro collegate e che formano un intero documento. Grazie a questa codifica è consentita la navigazione fra le singole componenti dell'oggetto digitale.

Il Gruppo congiunto OCLC-RLG sui metadati per la conservazione digitale ha prodotto un insieme di metadati da adottare in progetti di biblioteca

digitale nello svolgimento delle attività necessarie per assicurare l'accesso nel tempo e la conservazione del materiale digitale. Gli elementi definiti forniscono quindi le informazioni tecniche per le varie strategie, come la migrazione (trasferimento delle risorse digitali da una generazione di hardware e software alla successiva) ed emulazione (sviluppo di tecniche per simulare sistemi obsoleti su future generazioni di computer). In questo contesto i metadati spiegano l'ambiente tecnico necessario per utilizzare un oggetto, incluse le applicazioni e versioni software o hardware necessarie, gli schemi di decompressione, il riferimento ad altri file che devono essere fra loro collegati, etc. In linea con il modello OAIS nella definizione dei metadati si è tenuto conto, per garantire una loro ampia articolazione, che siano distinte l'identità e l'attività del creatore dell'archivio da chi ne ha la custodia.

A fronte del complesso e oneroso lavoro di predisposizione di uno schema come questo, sembra opportuno mettere in evidenza che la flessibilità dello strumento e la validità delle scelte fatte, come ad esempio l'indipendenza dei metadati di conservazione dal tipo o struttura delle risorse, oppure l'adattabilità dei metadati stessi a risorse di ogni tipo (come ad esempio i materiali di archivio istituzionali) dovranno essere testate in applicazioni specifiche.

Occorre tenere conto del fatto che gli schemi di metadati fin qui prodotti e brevemente illustrati sono orientati ad una notevole apertura e flessibilità: non distinguono gli oggetti nati digitali da quelli digitalizzati ed inoltre il modello OAIS non si occupa di valutare una tecnica di conservazione rispetto ad un'altra, come ad esempio la migrazione o l'emulazione. Si tratta quindi di specifiche e raccomandazioni di cui si devono accettare consapevolmente i principi, per poi utilizzarle in modo appropriato in progetti di biblioteca digitale. In questo modo sarà possibile disporre di una infrastruttura logica in cui le attività di gestione degli archivi sono inserite, sulla traccia del modello OAIS, e di un insieme articolato di metadati, pensati e definiti per il controllo e l'operatività del ciclo di vita degli archivi digitali secondo le varie necessità.

Anche per i dati relativi al copyright delle risorse trattate sono stati messi a punto specifici schemi che sono destinati ad essere gestiti in parallelo agli altri schemi di metadati descrittivi e tecnici oppure essere incorporati all'interno di questi. A tale proposito è utile precisare che una tecnica comune è quella di includere i metadati sui diritti digitali (così vengono spesso denominate le informazioni relative ai diritti e alle condizioni di uso degli oggetti trattati) nei profili di applicazione definiti da un particolare progetto o servizio e destinati a contenere tutti i dati gestiti da una applicazione, come i metadati descrittivi, tecnici e amministrativi gestionali, compresi quindi i diritti.

Vengono brevemente illustrati alcuni modelli di dati sui diritti di uso degli oggetti digitali: INDECS, XrML e ODLR. A parte INDECS, che vanta diverse applicazioni, si tratta di iniziative molto recenti che devono ancora essere diffusamente sperimentate per provarne l'efficacia ai fini della definizione dei dati per la gestione dei diritti di proprietà intellettuale e del commercio elettronico.

Interoperability of Data in E-Commerce Systems (INDECS)<sup>102</sup> è un organismo senza fini di lucro formato da aziende e agenzie di copyright che ha sviluppato un modello e ha elaborato un formato per la descrizione di varie componenti, come gli attori che intervengono nella negoziazione dei diritti, le transazioni compiute, la proprietà intellettuale stessa. Lo schema di metadati si propone dunque la definizione degli elementi necessari per garantire lo scambio e l'interoperabilità per la gestione dei diritti in un ambiente di commercio elettronico. Sono stati prodotti un dizionario dei dati inclusi nello schema e diverse mappature verso schemi quali FRBR, Dublin Core, ONIX.

eXensible rights Markup Language (XrML)<sup>103</sup> è un linguaggio che fornisce specifiche per esprimere diritti e condizioni di uso di contenuti e servizi digitali. Il suo utilizzo è subordinato a previa registrazione presso la ContentGuard Inc., che ha sviluppato il software per la gestione dello standard, basato sul linguaggio di marcatura XML.

---

102 INDECS. <http://www.indecs.org>

103 XrML. <http://www.xrml.org>

Open Digital Rights Language (ODRL)<sup>104</sup>, di recentissima elaborazione, è nato per fornire un linguaggio di pubblico dominio (contrapposto a XrML che richiede una licenza) per la gestione del commercio elettronico di risorse digitali, e si distingue per aver particolarmente sviluppato l'interoperabilità con materiale multimediale e il raccordo con lo schema MPEG.

#### **5.4 Standardizzazione e specificità dei metadati: l'evoluzione in atto**

Dal momento in cui è stata presa nella dovuta considerazione l'importanza, ai fini del recupero e della gestione delle risorse sul web, di definire gli elementi caratteristici degli oggetti digitali, vi è stato un fiorire di iniziative per la normalizzazione dei metadati e tale attività è in continua evoluzione. Essa riguarda gli elementi descrittivi e gestionali delle risorse, la codifica dei record di metadati, i vocabolari controllati e gli schemi utilizzati per attribuire determinati valori agli elementi stessi. Questi ultimi sono importanti strumenti, ben noti nel mondo delle biblioteche e in generale della documentazione, rappresentati da *authority files*, sistemi di classificazione, tesauri etc. La normalizzazione di questi strumenti, assolutamente essenziali anche per il recupero di risorse sul web, segue strade autonome di sviluppo.

Là dove si è concentrata l'attività di standardizzazione che interessa più da vicino il presente studio è l'area degli schemi di metadati, la struttura dei record che descrivono gli oggetti digitali, la definizione di termini o codici standard con cui individuare le varie tipologie di risorse e i dati necessari ad identificarle.

La standardizzazione a cui si è ispirata la definizione di Dublin Core è di un livello allo stesso tempo elementare e globale: si tratta di uno standard semplice e in certo senso povero, per la presenza di soli 15 elementi (al di là dell'utilizzo dei qualificatori che precisano gli elementi), che proprio per questa ragione risponde all'esigenza di garantire una condivisione di risorse più ampia possibile.

---

<sup>104</sup> ODRL. <http://odrl.net>

La normalizzazione operata per definire schemi più analitici e dettagliati, ma allo stesso tempo uniformi per particolari tipi di materiale, risponde all'esigenza di descrivere in modo controllato e preciso le relative risorse, garantendo un recupero puntuale a favore di comunità e applicazioni specifiche.

Si tratta quindi di due livelli di standardizzazione che sono ambedue necessari per garantire rispettivamente generalità e specificità, integrazione e autonomia.

A fronte della necessità di accesso a risorse eterogenee, descritte a livelli diversi e presenti in un unico grande sistema informativo, alle attività di normalizzazione nella definizione dei metadati si sono ben presto associate molte iniziative, come la nascita di organismi di coordinamento, l'istituzionalizzazione di gruppi di lavoro, ricerca e studio, lo sviluppo di forme di cooperazione a livello internazionale e intersistituzionale per raccordarsi su determinate applicazioni, fino alla predisposizione di strumenti nella forma di specifiche e prodotti software da condividere in rete.

#### ***5.4.1 L'impegno istituzionale nella definizione e nell'uso dei metadati : i profili di applicazioni e i registri***

Fra gli organismi e le iniziative più rilevanti per lo sviluppo in questo settore sono da evidenziare:

- a) La nascita dell'organismo che supporta e sviluppa lo standard elementare Dublin Core: Dublin Core Metadata Initiative (DCMI), formata da esperti internazionali di ogni continente ed affiliata all'OCLC, la grande rete di servizi americana per le biblioteche
- b) la definizione dei profili di applicazione e dei registri, strumenti essenziali per la documentazione, con funzione di supporto alla definizione di schemi e alla condivisione dei metadati
- c) Open Archive Initiative (OAI), un modello ed insieme un protocollo operativo per sviluppare e promuovere la disseminazione e l'utilizzo dei metadati.

**a) Dublin Core Metadata Initiative**, nata a seguito dell'adozione diffusa dello schema Dublin Core, ha come scopi istituzionali lo sviluppo di standard di metadati per l'accesso alla varietà delle risorse in rete, la promozione della normalizzazione nella descrizione delle risorse per l'interoperabilità, il sostegno alla definizione di schemi che siano orientati ai principi della condivisione e scambio fra diversi settori disciplinari e istituzionali. Si tratta di un modello di organizzazione e di lavoro aperto verso nuove idee che mira alla collaborazione attraverso il consenso su principi comuni, per realizzare l'accesso globale all'informazione salvaguardando la specificità delle risorse trattate.

L'adozione diffusa dello schema Dublin Core, specie nel settore pubblico della documentazione, ha spinto da una parte alla sua formalizzazione da parte degli esperti addetti al suo sviluppo e manutenzione, dall'altra ha affrettato il processo del suo riconoscimento come standard. Come tale è stato ratificato nel CEN (Comité Européen de Normalisation) Workshop Agreement 13874 ed è diventato standard ANSI Z39.85 ad opera della National Information Standards Organization americana (NISO). Il passo finale è l'approvazione da parte della International Standard Organisation (ISO), prevista nel 2003.

Numerose sono le aree di collaborazione e coordinamento in cui la DCMI è coinvolta: l'accordo con il Consorzio W3C e la partecipazione al programma del web semantico con sperimentazioni che utilizzano Dublin Core, la cooperazione con varie comunità impegnate nella creazione e diffusione dei metadati, come ad esempio il settore del governo elettronico, della documentazione e diffusione del materiale didattico. In questo ambito molto stretti sono i contatti con l'organismo IEEE-LOM: Learning Object Metadata, per lo sviluppo integrato di uno schema di metadati relativo ad oggetti relativi all'area della formazione e della didattica, che hanno oggi una notevole diffusione anche in seguito allo sviluppo dei corsi a distanza.

Per rendere un servizio efficace e concreto che promuova le applicazioni dei metadati Dublin Core sono stati definiti i *namespace*, con gli appropriati URI, per i tre tipi di dati che formano il nucleo centrale di

Dublin Core: i quindici elementi<sup>105</sup>, i qualificatori (definiti 'terms' nel contesto del loro utilizzo da parte delle applicazioni)<sup>106</sup>, i termini di vocabolari controllati da usare all'interno degli elementi e dei qualificatori (ad esempio la lista dei vari *tipi* di risorse)<sup>107</sup>. Si sottolinea ancora una volta l'importanza della definizione dei *namespace* in cui siano memorizzati i singoli schemi di metadati. Gli elementi degli schemi infatti devono trovare una loro sede specifica nello spazio del web; gli URI che li identificano sono gli strumenti grazie ai quali possono essere accessibili ed essere trattati da opportuni programmi preposti all'interpretazione e integrazione dei metadati.

Ulteriori iniziative riguardano l'elaborazione di uno schema XML-RDF per Dublin Core semplice<sup>108</sup> e qualificato<sup>109</sup>, linee guida continuamente aggiornate per l'uso del formato<sup>110</sup>, la creazione di strumenti software per la preparazione e l'elaborazione dei record<sup>111</sup>.

Una importante attività è costituita inoltre dalla elaborazione dei profili di applicazione Dublin Core, che sono insiemi di metadati ripresi da altri schemi che includono, oltre ai 15 elementi e ai qualificatori, ulteriori metadati specifici a singoli materiali e aree disciplinari. Questi vengono trattati di seguito.

## **b) I profili di applicazione e i registri**

La complessità e la varietà dei metadati sono oggi una realtà. La predisposizione di schemi generalmente definiti a livello internazionale per i singoli campi disciplinari e per i vari tipi di risorse rappresenta un risultato e un aiuto importante a vantaggio di chi crea i metadati e di chi li elabora per fornire servizi di accesso. D'altra parte la proliferazione degli schemi può costituire un ostacolo alla loro condivisione e all'integrazione di risorse eterogenee.

---

105 DC elements. <http://purl.org/dc/elements/1.1/>

106 DC terms. <http://purl.org/dc/terms/>

107 DC types. <http://purl.org/dc/dcmitype/>

108 Expressing simple Dublin Core in RDF/XML, 2002-07-31/ D. Beckett, E. Miller, D. Brickley. <http://dublincore.org/documents/dcmes-xml>

109 Expressing qualified Dublin Core in RDF/XML, 2002-04-14 / S. Kokkeliink, R.

Schwanzl. <http://dublincore.org/documents/dcq-rdf-xml/>

110 Using Dublin Core / Diane Hillmann, 2001-04-12.

<http://dublincore.org/documents/usageguide/>

111 Cit., DCMI : tools and software. <http://dublincore.org/tools/>

Altri fattori di complessità sono costituiti dalla spesso scarsa documentazione sulle caratteristiche degli schemi ed inoltre dall'uso, in uno stesso record di metadati, di più schemi per descrivere le risorse nelle varie applicazioni. L'utilizzo contemporaneo di più schemi per la gestione degli oggetti digitali è ormai una prassi molto frequente. Numerosi sono i progetti in cui una parte determinante del lavoro è costituita dalla definizione dei metadati che si rendono necessari alla gestione delle risorse trattate sulla base delle caratteristiche degli oggetti digitali, delle esigenze di determinate comunità di utenti e dei servizi da fornire: ciò si realizza sempre più spesso ricorrendo alla combinazione e integrazione di schemi di metadati diversi.

Questa tecnica si realizza mediante la definizione e l'uso di profili di applicazione. I profili sono veri e propri schemi di metadati che riutilizzano elementi di altri schemi, combinandoli in un insieme che contiene generalmente non solo metadati descrittivi, ma anche amministrativi e gestionali, ottimizzati per una particolare applicazione. Si può affermare che i profili sono veri e propri strumenti per l'armonizzazione e per l'accesso integrato agli oggetti digitali. Infatti questi schemi sono orientati al riutilizzo di elementi già definiti, in modo da non reinventare il lavoro già fatto da altri, garantire economicità nel lavoro e, cosa assai importante, assicurare una coerenza nel recupero delle risorse.

Non si tratta in realtà di un concetto nuovo. Molte applicazioni combinano elementi provenienti da più schemi, definiscono determinati elementi come 'locali', ne aggiungono di nuovi o modificano la semantica di schemi esistenti. La nascita dei vari formati MARC e l'accorgimento tipico del formato UNIMARC di consentire l'uso del numero 9 in etichette o codici di sottocampo per denotare un dato di tipo locale sono esempi di questa tecnica in uso da anni.

E' naturale che una precisa documentazione dei profili, ricorrendo a metodologie standard, consente di raggiungere questi obiettivi, primo fra tutti il riutilizzo dei profili da parte di implementatori di metadati, a garanzia di uniformità e risparmio di energie.

Nei profili sono dichiarati gli elementi usati in un'applicazione, con l'indicazione dei namespace degli schemi utilizzati, a cui si rimanda tramite gli URI pertinenti, che indirizzano appunto agli spazi di nomi dove risiedono le definizioni degli elementi stessi.

Il modello tipo nella costruzione di un profilo di applicazione è quello basato sull'uso di elementi di Dublin Core, se pertinenti, sull'utilizzazione di elementi specifici o qualificatori ripresi da altri schemi di metadati, sull'aggiunta di nuovi elementi o qualificatori. Recentemente molta attenzione viene posta alla regolamentazione di questo ultimo aspetto: gli ulteriori elementi, nuovi rispetto a schemi preesistenti, sono generalmente definiti in un *namespace* cosiddetto 'privato', quello della comunità o progetto che li ha definiti come propri. Le modalità con cui un namespace privato può diventare pubblico e quindi disponibile e condivisibile alle altre comunità sono allo studio e lo strumento per la loro pubblicazione è quello costituito dai registri di metadati.

Alcuni esempi di profili di applicazione e dei requisiti da rispettare nella definizione di elementi all'interno dei profili stessi sono forniti da DC-Education<sup>112</sup>, che è un profilo di applicazione sviluppato in stretta collaborazione con DCMI-Dublin Core Metadata Initiative, Collection Description-RSLP<sup>113</sup> per la descrizione a livello della collezione, DC-Gov per la descrizione dei servizi di governo elettronico, DC-Library<sup>114</sup> orientato al trattamento bibliografico delle risorse elettroniche, EULER<sup>115</sup> composto da elementi Dublin Core, risultanti dalla conversione di record in formati diversi, ad esempio MARC e UNIMARC, e da ulteriori elementi pertinenti all'area della matematica, Biblink, DAFNE, AGLS<sup>116</sup> messo a punto in Australia per la descrizione dei servizi di governo elettronico.

---

112 The DCMI Education metadata set. <http://www.schemas-forum.org/registry/schemas/DCMI-Education/>

113 RDF schema for the RSLP-CLD (Research Support Libraries Program-Collection Level Description) application profile. <http://www.ukoln.ac.uk/metadata/rspl/schema/>

114 DC-Library application profile.

<http://www.dublincore.org/documents/2002/04/16/library-application-profile/>

115 RDF schema for the EULER namespace and application profile.

<http://www.schemas-forum.org/registry/schemas/EULER/1.0/>

116 Australian Government Locator Service.

[http://www.naa.gov.au/recordkeeping/gov\\_online/agls/summary.html](http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html)

Coordinamento e integrazione sono dunque aspetti essenziali per lo sviluppo dei profili di applicazione. Ciò che è necessario è l'accordo su procedure e metodi per la dichiarazione e la documentazione dei profili, un'area in cui da qualche anno molto lavoro è in corso a livello internazionale.

Come sopra accennato un fattore di successo dei profili di applicazione è la creazione e l'istituzionalizzazione dei registri. Questi sono strumenti a cui è assegnato il compito di documentare in modo analitico, standardizzato e continuativo, gli schemi di metadati elaborati, i profili, le applicazioni ed anche, in prospettiva, i vocabolari controllati utilizzati nei singoli metadati. I registri di metadati sono una creazione recente; essi rispondono all'esigenza di informare in modo chiaro sull'esistenza e l'utilizzo degli schemi, configurandosi come strumenti assai utili per gli implementatori di servizi, per i creatori degli schemi che possono utilizzare il lavoro già fatto, ed anche per gli utenti per verificare il significato dei metadati.

Come risulta dalle raccomandazioni sulla codifica dei metadati illustrate precedentemente, per consentire alle applicazioni l'interpretazione e il riutilizzo dei metadati è necessario che nei registri figurino gli schemi di base e i profili per i quali siano state definite sintassi standard, come RDF Schema, DTD (Document Type Definition: dichiarazione dei termini utilizzati in uno schema), espresse in linguaggio XML e XML Schema.

Esistono oggi diverse iniziative di creazione di registri, portate avanti da istituzioni e progetti cooperativi. Il loro principale scopo è quello di dare accesso agli schemi offrendo funzioni di ricerca e scorrimento dei termini, chiarendo il significato dei termini utilizzati al fine di consentirne un uso appropriato in fase di costruzione di applicazioni. Ma i registri hanno anche la funzione di sostenere l'evoluzione degli schemi, mediante meccanismi di registrazione da parte di enti preposti alla standardizzazione. In tal modo si auspica che siano assolti i compiti che stanno manifestandosi in tutta la

loro urgenza: documentazione diffusa sugli schemi, recupero agevolato e loro riutilizzo in modo efficace, armonizzazione ai fini di economicità e coerenza nella descrizione delle risorse.

Fra le più significative attività nell'implementazione di registri si segnalano le seguenti: il progetto DESIRE<sup>117</sup>, DCMI registry<sup>118</sup>, SCHEMAS<sup>119</sup>, EU CORES<sup>120</sup>, Metaform<sup>121</sup>, che è una vera e propria base dati di schemi di metadati creata dalla Biblioteca universitaria di Göttingen, MEG (Metadata Education Group)<sup>122</sup> per i metadati relativi all'area didattica.

Le linee di tendenza a livello internazionale sono quelle rappresentate dagli obiettivi del progetto EU-CORES, il cui scopo è la creazione di registri autorevoli e il coordinamento nell'attività di creazione degli schemi, mediante la collaborazione fra diversi organismi preposti alla standardizzazione, come ad esempio DCMI, DOI, ONIX, GILS, MPEG, FGDC. Si tratta quindi di una cooperazione ad ampio raggio, non solo sul piano geografico, ma organizzativo e istituzionale. Ciò a cui si tende infatti è l'armonizzazione di ambienti diversi, non solo dal punto di vista disciplinare, quanto settoriale, economico e sociale, con il coinvolgimento del mondo commerciale a fianco di quello istituzionale.

L'attività di documentazione secondo le raccomandazioni definite nell'ambito di queste iniziative cooperative a livello internazionale rappresenta in certo modo un atto di fede verso il web semantico, visto come un ambiente popolato di metadati ricchi e precisi, semanticamente flessibili e derivati da fonti ufficiali e affidabili.

Viene fornito di seguito un esempio del profilo di applicazione MEG, documentato nell'apposito registro. Si tratta di un profilo ricco per l'uso di molti qualificatori Dublin Core, ma allo stesso tempo semplice, in quanto

---

117 DESIRE metadata registry. <http://desire.ukoln.ac.uk/registry/>

118 DCMI registry. [http://www.oclc.org/research/projects/dcmi\\_registry/index.shtm](http://www.oclc.org/research/projects/dcmi_registry/index.shtm)

119 SCHEMAS : forum for metadata schema implementers. <http://www.schemas-forum.org/>

120 EU-CORES project. <http://www.cores-eu.net/>

121 Metaform at the State and University Library at Göttingen, Germany (SUB). <http://www2.sub.uni-goettingen.de/metaform/>

122 Metadata Education Group registry project. <http://ukoln.ac.uk/metadata/education/regproj/>

combina solo due schemi: Dublin Core qualificato e National Curriculum Metadata Element Set.

## Elements

ID	Name	Element Set	
<a href="http://purl.org/dc/terms/abstract">http://purl.org/dc/terms/abstract</a>	Abstract	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/alternative">http://purl.org/dc/terms/alternative</a>	Alternative	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/audience">http://purl.org/dc/terms/audience</a>	Audience	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/available">http://purl.org/dc/terms/available</a>	Available	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/conformsTo">http://purl.org/dc/terms/conformsTo</a>	Conforms To	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/elements/1.1/contributor">http://purl.org/dc/elements/1.1/contributor</a>	Contributor	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/elements/1.1/coverage">http://purl.org/dc/elements/1.1/coverage</a>	Coverage	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/terms/created">http://purl.org/dc/terms/created</a>	Created	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a>	Creator	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/elements/1.1/date">http://purl.org/dc/elements/1.1/date</a>	Date	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/elements/1.1/description">http://purl.org/dc/elements/1.1/description</a>	Description	The Dublin Core Element Set v1.1	Detail
<a href="http://www.nc.uk.net/metadata/elements/duration">http://www.nc.uk.net/metadata/elements/duration</a>	Duration	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://www.nc.uk.net/metadata/elements/expiry">http://www.nc.uk.net/metadata/elements/expiry</a>	Expiry	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://purl.org/dc/terms/extent">http://purl.org/dc/terms/extent</a>	Extent	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/elements/1.1/format">http://purl.org/dc/elements/1.1/format</a>	Format	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/terms/hasFormat">http://purl.org/dc/terms/hasFormat</a>	Has Format	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/hasPart">http://purl.org/dc/terms/hasPart</a>	Has Part	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/hasVersion">http://purl.org/dc/terms/hasVersion</a>	Has Version	The Dublin Core Terms namespace	Detail
<a href="http://www.nc.uk.net/metadata/elements/inareas">http://www.nc.uk.net/metadata/elements/inareas</a>	Inareas	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://www.nc.uk.net/metadata/elements/infoType">http://www.nc.uk.net/metadata/elements/infoType</a>	InfoType	National Curriculum Metadata Element Set, 2.07	Detail

<b>ID</b>	<b>Name</b>	<b>Element Set</b>	
<a href="http://purl.org/dc/terms/isFormatOf">http://purl.org/dc/terms/isFormatOf</a>	Is Format Of	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/isPartOf">http://purl.org/dc/terms/isPartOf</a>	Is Part Of	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/isReferencedBy">http://purl.org/dc/terms/isReferencedBy</a>	Is Referenced By	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/isReplacedBy">http://purl.org/dc/terms/isReplacedBy</a>	Is Replaced By	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/isRequiredBy">http://purl.org/dc/terms/isRequiredBy</a>	Is Required By	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/isVersionOf">http://purl.org/dc/terms/isVersionOf</a>	Is Version Of	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/issued">http://purl.org/dc/terms/issued</a>	Issued	The Dublin Core Terms namespace	Detail
<a href="http://www.nc.uk.net/metadata/elements/keystage">http://www.nc.uk.net/metadata/elements/keystage</a>	Keystage	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://purl.org/dc/elements/1.1/language">http://purl.org/dc/elements/1.1/language</a>	Language	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/terms/mediator">http://purl.org/dc/terms/mediator</a>	Mediator	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/medium">http://purl.org/dc/terms/medium</a>	Medium	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/modified">http://purl.org/dc/terms/modified</a>	Modified	The Dublin Core Terms namespace	Detail
<a href="http://www.nc.uk.net/metadata/elements/publication">http://www.nc.uk.net/metadata/elements/publication</a>	Publication	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://purl.org/dc/elements/1.1/publisher">http://purl.org/dc/elements/1.1/publisher</a>	Publisher	The Dublin Core Element Set v1.1	Detail
<a href="http://www.nc.uk.net/metadata/elements/publisherEmail">http://www.nc.uk.net/metadata/elements/publisherEmail</a>	Publisher Email	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://purl.org/dc/terms/references">http://purl.org/dc/terms/references</a>	References	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/elements/1.1/relation">http://purl.org/dc/elements/1.1/relation</a>	Relation	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/terms/replaces">http://purl.org/dc/terms/replaces</a>	Replaces	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/terms/requires">http://purl.org/dc/terms/requires</a>	Requires	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/elements/1.1/identifier">http://purl.org/dc/elements/1.1/identifier</a>	Resource Identifier	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/elements/1.1/type">http://purl.org/dc/elements/1.1/type</a>	Resource Type	The Dublin Core Element Set v1.1	Detail

<b>ID</b>	<b>Name</b>	<b>Element Set</b>	
<a href="http://purl.org/dc/elements/1.1/source">http://purl.org/dc/elements/1.1/source</a>	Source	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/terms/spatial">http://purl.org/dc/terms/spatial</a>	Spatial	The Dublin Core Terms namespace	Detail
<a href="http://www.nc.uk.net/metadata/elements/specifier">http://www.nc.uk.net/metadata/elements/specifier</a>	Specifier	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://purl.org/dc/elements/1.1/subject">http://purl.org/dc/elements/1.1/subject</a>	Subject and keywords	The Dublin Core Element Set v1.1	Detail
<a href="http://purl.org/dc/terms/tableOfContents">http://purl.org/dc/terms/tableOfContents</a>	Table Of Contents	The Dublin Core Terms namespace	Detail
<a href="http://www.nc.uk.net/metadata/elements/teaching">http://www.nc.uk.net/metadata/elements/teaching</a>	Teaching	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://purl.org/dc/terms/temporal">http://purl.org/dc/terms/temporal</a>	Temporal	The Dublin Core Terms namespace	Detail
<a href="http://purl.org/dc/elements/1.1/title">http://purl.org/dc/elements/1.1/title</a>	Title	The Dublin Core Element Set v1.1	Detail
<a href="http://www.nc.uk.net/metadata/elements/users">http://www.nc.uk.net/metadata/elements/users</a>	Users	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://purl.org/dc/terms/valid">http://purl.org/dc/terms/valid</a>	Valid	The Dublin Core Terms namespace	Detail
<a href="http://www.nc.uk.net/metadata/elements/xareas">http://www.nc.uk.net/metadata/elements/xareas</a>	Xareas	National Curriculum Metadata Element Set, 2.07	Detail
<a href="http://www.nc.uk.net/metadata/elements/xtags">http://www.nc.uk.net/metadata/elements/xtags</a>	Xtags	National Curriculum Metadata Element Set, 2.07	Detail

### c) Open Archive Initiative<sup>123</sup>

Si tratta di un modello adottato oggi in numerosi progetti, applicato da diverse istituzioni che utilizzano metadati e che per mezzo di questi forniscono servizi di accesso ad oggetti digitali. Nata nell'ambito della comunicazione accademica per favorire l'accesso agli archivi di e-prints ai ricercatori su vasta scala, OAI è divenuta una metodologia standard, con un suo protocollo specifico (Metadata Harvesting Protocol: OAI-MHP)<sup>124</sup> per la raccolta e la gestione di metadati. Oggi il protocollo è esteso a risorse elettroniche di ogni tipo, residenti in archivi diversi e ha come caratteristica essenziale quella di consentire di catturare i metadati, selezionarli, trasferirli su database per fornire su di essi servizi diversificati e garantire l'accesso a varie risorse in modo unificato.

Il modello prevede due entità: i fornitori di metadati e i fornitori di servizi; questi comunicano fra loro grazie all'adozione del protocollo MHP. Nell'ambiente OAI i fornitori di metadati creano e gestiscono uno o più archivi o database di metadati (*repositories*) che supportano il protocollo; i fornitori di servizi li interrogano mediante le richieste formalizzate del protocollo, selezionandoli e catturandoli per costruire servizi a valore aggiunto, come ad esempio una interfaccia unificata per i propri utenti che non sono quindi obbligati a conoscere metadati e modalità di interrogazione dei sistemi originari. La richiesta fatta dai fornitori di servizi mediante le istruzioni e i comandi OAI dà come risultato una sequenza di byte codificata in XML (record OAI) che contiene, oltre ai metadati di una determinata risorsa accompagnati dall'identificatore dello schema o degli schemi utilizzati per la preparazione dei metadati stessi, l'indicazione di eventuali diritti sui metadati.

E' importante sottolineare le caratteristiche principali dell'approccio che sta alla base di OAI: il protocollo non si occupa dei modelli economici e degli accordi fra i fornitori dei metadati e degli erogatori dei servizi, che

---

123 Open Archive Initiative - OAI. <http://www.openarchives.org>

124 OAI-Protocol for Metadata Harvesting.

<http://www.openarchives.org/OAI/openarchivesprotocol.html>

intercorreranno in negoziazioni esterne al protocollo, anche se il record OAI può contenere informazioni sulle politiche di accesso ai metadati.

La fornitura di metadati in formato Dublin Core semplice è obbligatoria, per garantire una base comune su cui scambiare i metadati. Ma sono ovviamente possibili selezioni e catture di schemi diversi: infatti il protocollo consente la raccolta (*harvesting*) di molteplici formati di metadati, il cui requisito è quello di essere definiti in schemi XML per gestirne e controllarne la composizione e le proprietà.

E' utile sottolineare il ruolo di rottura che hanno rappresentato OAI e il relativo protocollo, in quanto forniscono una tecnologia agile e a basso costo per catturare concretamente i metadati e procedere al loro utilizzo.

Il monitoraggio sulle applicazioni, la costituzione di un Open Forum<sup>125</sup> per lo scambio di esperienze e la messa a disposizione di software per la gestione del protocollo rendono questa iniziativa estremamente importante per biblioteche e istituzioni che sono preposte alla fornitura di servizi di accesso integrato.

## 5.5. Alcuni progetti internazionali

La fioritura di progetti che utilizzano metadati è iniziata ormai dalla metà degli anni '90, ma gli sviluppi che oggi si susseguono con grande rapidità offrono un panorama molto più ricco e sofisticato rispetto alle applicazioni di qualche anno fa. A solo titolo di esempio vengono qui illustrate brevemente alcune realizzazioni, delineandone le scelte strategiche ed operative nell'area dei metadati.

### Cooperative Online Resource Catalogue (CORC)<sup>126</sup>

Messo a punto da OCLC con l'intento di promuovere la catalogazione delle risorse elettroniche fra le biblioteche in modo cooperativo, fornisce strumenti per la preparazione dei record in formato MARC21, Dublin Core semplice e qualificato, con codifica HTML e RDF. I formati sono fra loro

---

125 Open Archive Forum. <http://www.oaforum.org/workshops/>

126 Cooperative Online Resource Catalog. <http://connexion.oclc.org>

speculari, convertibili automaticamente al momento dell'immissione dei dati.

Il sistema offre, a fronte della digitazione di un identificatore (ad oggi, generalmente, un URL), la visualizzazione del sito web o altro tipo di risorsa, con il relativo record descrittivo che sarà completo se è già stato catalogato, oppure in forma minimale, come risultato di una conversione automatica operata da opportuni programmi, che poi ne consentono il completamento da parte del catalogatore.

Attualmente il catalogo delle risorse supera i 500.000 record e la ricerca è eseguibile nella base dati delle risorse elettroniche (Resource catalogue) e in Worldcat, che contiene le risorse elettroniche integrate con pubblicazioni tradizionali e che costituisce il più grande archivio bibliografico accessibile nel mondo.

Dopo aver sottoscritto la loro partecipazione le biblioteche aderenti si impegnano a catalogare le risorse sulla base dati cooperativa con la possibilità di trasferire i loro record nel proprio sistema.

L'importanza del progetto sta nell'opportunità offerta agli istituti partecipanti di disporre di strumenti per la descrizione di risorse a vari livelli, optando per standard catalografici e di formato come le AACR2 e il MARC o per lo schema di metadati Dublin Core, lasciando a soluzioni software ed a opportune mappature preparate sulla base di specifiche dettagliate, l'integrazione fra MARC e i metadati Dublin Core, sia per materiali tradizionali che elettronici.

#### The European Library (TEL)<sup>127</sup>

Sponsorizzato dalla Commissione europea all'interno del programma Information Society Technologies (IST) e formato dalla partecipazione di 10 biblioteche e istituzioni nazionali (fra i quali la BNC di Firenze e l'ICCU), TEL è un progetto in corso di realizzazione, che si propone di verificare gli aspetti tecnici e politici derivanti dalla condivisione di risorse digitali. In questo ambito, al fine di creare una infrastruttura di cooperazione per

---

127 The European Library : the gate to Europe's knowledge.  
<http://www.europeanlibrary.org/>

l'accesso alle maggiori collezioni europee, vengono utilizzati protocolli e schemi standard come Z39.50 per la ricerca su basi distribuite, il protocollo OAI e un profilo di applicazione in via di definizione, basato su Dublin Core Library application profile. E prevista la ricerca unificata su basi dati diverse ed è costruito anche un indice centralizzato, che conterrà record di metadati XML e che sarà accessibile su web tramite il protocollo http.

A partire dalla situazione corrente, in cui editori e biblioteche nazionali usano diversi schemi di metadati descrittivi, è stato deciso di basare la scelta del modello di dati essenzialmente sulle funzioni del portale di accesso alle risorse digitali trattate da TEL, ed è su tale modello che vengono rapportati i diversi schemi (MARC21, FinMARC, Dublin Core, UNIMARC, Pica3, PicaPlus, COMARC, alcuni formati proprietari). Nell'ambito del progetto quindi sono state definite dettagliatamente le funzionalità da raggiungere (ad esempio ricerca e recupero, identificazione, descrizione, servizi di collegamento e localizzazione alle risorse, accesso a livello delle collezioni, requisiti hardware e software, navigazione, catalogazione partecipata, etc.) e su questa base sono stati definiti i metadati e i relativi qualificatori.

Il progetto si sta indirizzando verso il profilo DC-Library, che è una combinazione di elementi di Dublin Core qualificato e di elementi di altri schemi. Esso sarà generato mediante conversione dai formati utilizzati dai singoli partecipanti (i test finora implementati hanno dato ottimi risultati, specie sui record della Biblioteca nazionale centrale di Firenze convertiti da UNIMARC e su quelli della British Library), ma per garantire tutti i servizi che il progetto si propone si renderanno necessari altri metadati. Saranno trattati anche i metadati che descrivono il livello della collezione, seguendo lo schema Collection Level Description sviluppato da RSLP<sup>128</sup>.

Anche se il profilo DC-Library costituisce il modello di dati di partenza, sarà sviluppato uno specifico profilo di applicazione TEL, in quanto il progetto tratta e dovrà ancora più estesamente trattare tutti i tipi di

---

128 RSLP Collection description Schema / A. Powell.  
<http://www.ukoln.ac.uk/metadatas/rslp/schema/>

materiali e dovrà essere aperto a nuove istituzioni ed esigenze. Un registro TEL in questo caso avrà la funzione di documentare il profilo DC-library ora adottato da TEL ed altri metadati e profili che si renderanno necessari.

Le decisioni sull'adozione di standard recentemente messi a punto insieme ad una attenta considerazione dei requisiti funzionali del portale TEL, sulla quale misurare le scelte sui metadati da utilizzare, fanno di questo progetto un punto di riferimento molto utile per le applicazioni di biblioteca digitale e quindi di accesso integrato a risorse distribuite.

### META engine<sup>129</sup>

Si tratta di un progetto cooperativo europeo di ricerca e sviluppo di una serie di programmi per il riconoscimento automatico e l'estrazione di metadati al momento della generazione delle immagini e della digitalizzazione di materiale a stampa (libri e periodici) combinando tecniche di riconoscimento ottico e di analisi dei documenti. Il motore META-e è progettato in modo da fornire i dati in XML, generando metadati descrittivi, tecnici e strutturali secondo gli schemi Dublin Core, METS e ALTO (Analyzed Layout and Text Object per la definizione degli elementi strutturali). Sulla base delle previsioni fatte al momento della progettazione e dell'analisi in corso di sperimentazione (il progetto terminerà nell'agosto 2003), sono stati previsti i diversi casi in cui vi è necessità o meno di integrazione, correzione, controllo dei metadati prodotti automaticamente, con predisposizione di raccomandazioni per la correzione essenzialmente dei metadati descrittivi e degli elementi logici della struttura dei documenti, come ad esempio i paragrafi, le note, le appendici, etc.

Il valore del progetto sta nella preparazione di strumenti integrati per l'esecuzione delle procedure di digitalizzazione, che includono la scannerizzazione, l'elaborazione delle immagini, l'analisi fisica e logica dei documenti, il controllo di qualità, le attività di configurazione e amministrazione. Grazie ad un'interfaccia grafica espressamente messa a

---

129 METAe : the metadata engine project. <http://meta-e.uibk.ac.at/>

punto è possibile verificare e correggere i metadati prodotti automaticamente.

Un altro modulo essenziale è quello che si basa sul database delle regole create per trattare i documenti; questi vengono analizzati nella loro struttura fisica e logica ed è stata creata una sorta di grammatica, espandibile ad altri tipi di documenti oltre ai libri e ai periodici, ai fini del riconoscimento della sintassi sulla quale si fonda la generazione dei metadati.

#### District Architecture for Networked Editions (DAFNE)<sup>130</sup>

DAFNE è un progetto in corso finanziato dal Ministero per l'educazione, università e ricerca, che vede la collaborazione di diverse università italiane (alcune delle quali, come ad esempio l'Università di Firenze, recentemente operative con ruolo di 'university press'), istituzioni nazionali come la Biblioteca nazionale di Firenze e l'Istituto di documentazione giuridica del CNR, editori e ditte di software e di servizi. L'intento è quello di sviluppare un prototipo di infrastruttura per l'editoria elettronica a livello nazionale. Il modello tecnico e architettonico si basa su OAIS e prevede l'interazione fra i fornitori di dati, i fornitori dei servizi e vari servizi esterni. A partire dai due modelli economici oggi esistenti per l'accesso all'informazione digitale, che vedono protagonisti gli editori commerciali e quella parte del mondo accademico aperto alla libera diffusione delle pubblicazioni accademiche, il progetto propone servizi in cui ambedue le modalità siano rese possibili, prevedendo varie funzionalità. Fra queste sono da segnalare la fornitura di documenti in formato digitale, la creazione e l'aggiornamento dei metadati, meccanismi di validazione dei contenuti, firma digitale e tecniche di crittografia, indicizzazione a testo pieno e ricerca per campi, modalità e condizioni di uso diversificate, interazione degli autori con il sistema, pagamenti on line e gestione dei diritti. Un prototipo del sistema previsto sarà sviluppato utilizzando come campione diversi documenti nell'area delle scienze sociali.

---

<sup>130</sup> District Architecture for networked edition : technical model and metadata.  
<http://www.bncf.net/dc2002/program/ft/paper1.pdf>

Si tratta quindi di una infrastruttura ad ampio spettro in cui i metadati per la descrizione e gestione delle risorse e dei processi in cui esse sono coinvolte hanno un ruolo essenziale.

Uno dei punti cardine del progetto è l'identificazione di un insieme di metadati che si basi su schemi esistenti, in nome dell'integrazione fra risorse eterogenee e della coerenza nella ricerca e nell'accesso. Il progetto si trova in una fase intermedia, in cui solo i metadati descrittivi e quelli relativi ai diritti di accesso sono stati definiti, mentre quelli tecnici, strutturali e di conservazione saranno analizzati in seguito.

DAFNE prende come punto di riferimento lo schema Publishing Requirements for Industry Standard Metadata (PRISM)<sup>131</sup>, messo a punto da editori, venditori e intermediari, che hanno definito un insieme di dati utili per lo scambio di documenti tanto in un contesto di editoria tradizionale che elettronica. Relativamente al contenuto delle risorse e quindi ai metadati descrittivi il modello scelto si basa su Dublin Core, ma contiene molti altri elementi per identificare, ad esempio, la provenienza degli oggetti, le informazioni sulle date, le relazioni con altre risorse. Inoltre nel profilo DAFNE sono stati inclusi tanto metadati sui diritti, ripresi dal linguaggio ODLR<sup>132</sup>, che una serie di elementi definiti dal progetto stesso (chiamati dai partecipanti al progetto 'elementi DAFNE'). Questi ultimi, una decina al momento, sono stati ritenuti utili per specificare le modalità di pagamento ed anche le modalità di accesso agli oggetti digitali.

Per quanto riguarda i metadati dunque, il valore del progetto sta nell'adozione di standard architetturali (il modello OAIS) e nell'attenta valutazione e considerazione degli schemi esistenti. PRISM registra un utilizzo piuttosto consolidato e ne esistono applicazioni XML; ciò facilita l'interpretazione e lo scambio dei record di metadati. Un altro elemento importante è costituito dalla metodologia affrontata: individuazione degli

---

131 Publishing Requirements for Industry Standard Metadata.  
<http://www.prismstandard.org/techdev/prismspec11.asp>

132 Open Digital Rights language. <http://odrl.net/>

attori e processi coinvolti, definizione delle risorse da trattare e delle problematiche per il loro accesso, studio dei metadati necessari basandosi sugli standard esistenti, costruzione di un profilo, che dovrà ovviamente essere testato con l'implementazione del prototipo.

### European Libraries and Electronic Resources in mathematical sciences (EULER)<sup>133</sup>

Pur trattandosi di un progetto cooperativo a livello europeo, EULER viene riportato in questa sezione dedicata ai progetti italiani per la partecipazione dell'Università di Firenze che ha avuto un ruolo molto importante nella definizione del profilo di metadati ai fini dell'integrazione di risorse.

Il portale per la matematica creato all'interno del progetto, che ora si avvia a diventare servizio operativo ad opera di un apposito consorzio, rappresenta un modello semplice e lineare di integrazione di risorse e servizi. Le risorse messe a disposizione dai partners (biblioteche universitarie, servizi di indicizzazione, centri di documentazione, editori) sono di vario tipo, come ad esempio cataloghi, basi dati specializzate, archivi di pre-prints, riviste elettroniche. Sulla base di un questionario destinato ai matematici che fanno uso di questo materiale nel loro lavoro e attività di ricerca, è stato definito il profilo EULER-DC, normalizzando i dati mediante conversione dai record originari dei partecipanti per offrire un accesso integrato alle varie risorse, che vengono descritte nei sistemi nativi con formati, regole di catalogazione e classificazione differenti. E' stato dunque compito dei partner convertire i propri metadati negli elementi descrittivi Dublin Core, a cui sono stati aggiunti altri elementi che non trovavano alcuna corrispondenza, ma utili alla ricerca e gestione dei documenti trattati dal progetto.

I record convertiti in Dublin Core sono memorizzati nei singoli sistemi distribuiti. Nel caso specifico dell'Università di Firenze il data base locale per EULER viene alimentato periodicamente da opportuni programmi che estraggono dall'OPAC generale, secondo algoritmi e vocabolari

---

<sup>133</sup> EULER. <http://www.emis.de/projects/EULER/>

opportunamente predisposti, i record pertinenti all'area della matematica. Questi vengono poi convertiti da UNIMARC a Dublin Core in formato XML. La ricerca effettuata grazie al motore EULER da parte dell'utente viene indirizzata ai vari database Dublin Core, che rispondono inviando i record corrispondenti ai parametri di ricerca. All'utente è presentata una lista di risultati unica, con deduplicazione e possibilità di puntare al servizio scelto. E' da notare la linearità dell'approccio adottato dal progetto, che è riuscito, sulla base delle opportune indagini sui bisogni dell'utenza, a definire un modello di dati comune, generato automaticamente a partire dai formati nativi, seguendo gli standard in uso e superando le difficoltà derivanti dalla eterogeneità di risorse, diverse per tipologia e per i sistemi che le ospitano, come OPAC, pagine web, indici di pre-print, etc.

### Biblioteca italiana telematica<sup>134</sup>

Il progetto si propone di rendere disponibili i testi della tradizione culturale e letteraria italiana: un numero consistente di documenti (circa 1700) è già disponibile in formato digitale grazie al lavoro svolto nell'ambito del progetto portato avanti dal Centro Universitario Biblioteca Italiana Telematica (CIBIT), frutto della collaborazione di diverse università italiane. L'obiettivo è quello di costituire un punto di servizio avanzato anche in collegamento con l'attività del consorzio ICoN, composto da un numero consistente di università e dal consorzio Nettuno, per la promozione della lingua e lo studio on line della cultura e letteratura italiane.

Nell'ambito di questo progetto è iniziata la codifica in linguaggio XML dei testi digitalizzati (ad oggi circa 70) e sono in corso di stesura le specifiche per l'utilizzo dei metadati necessari a descrivere il materiale trattato. L'orientamento è verso l'adozione dello schema METS, messo a punto dalla Library of Congress e precedentemente illustrato, unitamente al formato per i dati descrittivi MODS, anch'esso utilizzato in diversi progetti della Biblioteca di Washington e che ricalca la codifica MARC presente nei tradizionali record bibliografici.

Il progetto è caratterizzato dall'attenzione posta ai metadati gestionali e tecnici in quanto si tratta di testi digitalizzati in proprio, ma anche dalla preoccupazione di rendere la ricerca e l'accesso più puntuale possibile dal momento che l'obiettivo è quello di creare uno strumento di alto valore scientifico, orientato alla ricerca e alla didattica. Sono attualmente in corso le attività per la definizione di uno schema XML per i metadati che saranno scelti (che si configura come un vero e proprio profilo di applicazione per la Biblioteca italiana) ed è inoltre stato avviato lo studio sulle metodologie da adottare per il controllo di autorità degli accessi, come autori, titoli, etc. Si tratta quindi di una iniziativa dove particolare cura è posta all'identificazione degli elementi distintivi degli oggetti digitali, alle loro caratteristiche tecniche, strutturali e di contenuto, ma anche agli aspetti

---

<sup>134</sup> Biblioteca italiana. <http://www.bibliotecaitaliana.it> ; dall'inizio del 2003 il sito risulta purtroppo inaccessibile, né si danno spiegazioni su dove possa essere aperto.

propriamente semantici per quanto riguarda la forma delle intestazioni, al fine di garantire una ricerca di qualità, caratterizzata da coerenza e precisione.

## 5.6. Conclusioni

E' chiaro come in un qualsiasi progetto di biblioteca digitale, sia esso orientato alla produzione di risorse elettroniche, alla digitalizzazione di materiali documentari di tipo analogico o al recupero e all'accesso di risorse digitali già disponibili come tali, è indispensabile operare scelte oculate sulla scelta del materiale da trattare e sull'utenza da soddisfare in base agli obiettivi precedentemente fissati, misurati sulle risorse finanziarie e umane che si hanno a disposizione. Fra le decisioni che vanno prese nel corso del progetto è fondamentale quella che riguarda i metadati da adottare per l'esecuzione delle varie funzioni di acquisizione, produzione, accesso, utilizzo nel tempo degli oggetti digitali.

Il panorama dei metadati si presenta molto ricco e in continua evoluzione: sono stati messi a punto e sperimentati ormai numerosi schemi per la descrizione delle risorse e per la loro gestione; alcuni sono ampiamente utilizzati in applicazioni avviate da tempo ed hanno raggiunto un buon livello di standardizzazione, con la loro codifica XML in namespace specifici. Ciò indica il loro livello di ufficialità e diffusione, la possibilità per applicazioni di biblioteca digitale di adottarli per la descrizione delle risorse da trattare, di interpretarli e integrarli con altri schemi per la costruzione di servizi integrati. Altri hanno un'origine recente e si trovano in una fase di iniziale consenso e normalizzazione.

Sempre più numerosi sono i profili di applicazione, combinazioni di più schemi necessari per identificare elementi specifici a certi tipi di risorse e applicazioni, che allo stesso tempo utilizzano, fin dove possibile, schemi già esistenti e standardizzati per salvaguardare compatibilità e integrazione in servizi di accesso unificato.

L'adozione di Dublin Core nella sua funzione di minimo comune denominatore per l'identificazione delle risorse è consolidata, e ormai la

quasi totalità dei progetti di biblioteca digitale ne deve prevedere l'utilizzo, quanto meno come formato verso cui convertire schemi di metadati più ricchi e articolati.

Oltre all'elaborazione degli schemi di metadati, molte sono le iniziative e forte è l'impegno, diffuso a livello internazionale, per garantire un accesso coerente e più preciso possibile alle risorse web. Questi sono rappresentati da iniziative essenziali quali: a) l'attività di normalizzazione operata per la codifica dei record di metadati ricorrendo a linguaggi di marcatura HTML e XML; b) l'adozione sperimentale di RDF come architettura per lo scambio di metadati fra applicazioni e per l'espressione delle proprietà delle risorse in modo standard; c) la costruzione di prodotti software, commerciali e di pubblico dominio, per la gestione di metadati; d) l'istituzionalizzazione dei registri come strumenti per documentare l'esistenza degli schemi di metadati, le loro caratteristiche e i vocabolari controllati usati nei metadati stessi.

Ma non tutti i problemi sono risolti e occorre vigilare in modo continuo, esaminando gli sviluppi in corso su scala mondiale e procedere a sperimentazioni operando scelte oculate nel settore dei metadati, per la loro funzione essenziale di consentire la gestione delle risorse, lo scambio e l'integrazione dei servizi. Esattamente come la lingua che evolve con l'uso, anche i metadati sono tutt'altro che stabili e sono destinati a cambiare nel tempo. Le diverse esigenze di accesso e la nascita di nuovi tipi di oggetti digitali portano alla proliferazione dei metadati, mentre d'altra parte gli implementatori tendono spesso a dare agli elementi una interpretazione orientata ad una determinata applicazione.

La creazione e la gestione dei metadati hanno un costo e questo è tanto più alto quanto maggiore è la completezza nella descrizione delle risorse. La ricchezza o meno dei metadati è una scelta essenziale da compiere in ogni progetto di biblioteca digitale e tale strategia a sua volta si basa sui servizi da fornire e quindi sulle funzionalità che si intendono garantire. In linea generale si può affermare che una descrizione dettagliata degli

oggetti digitali ne favorisce notevolmente la ricerca, rendendola più precisa e migliorandone la gestione, ma richiede mezzi e risorse adeguate, rendendo d'altra parte più problematica la coerenza con altre applicazioni. Come è noto le soluzioni raccomandate in proposito sono al momento le mappature fra gli schemi, l'adozione di sistemi normalizzati nell'architettura dei record (come ad esempio XML e RDF), l'uso di un formato standard che serva come minimo comune denominatore. Una descrizione semplificata è senz'altro più agevole ed economica da creare; se da una parte può rendere la ricerca meno precisa o comunque richiedere da parte dell'utente specializzato uno sforzo maggiore per individuare i risultati più rilevanti, dall'altra rende più facile l'interoperabilità fra settori e risorse.

Fra la soluzione offerta dai motori di ricerca commerciali che raccolgono una buona parte di risorse Internet e forniscono una sorta di indicizzazione a basso costo, senza contestualizzare e delimitare l'informazione, e la soluzione fornita dalle biblioteche e centri di documentazione specializzati che controllano e catalogano analiticamente il materiale ricorrendo a schemi di descrizione consolidati, occorre esaminare attentamente i vantaggi e le implicazioni di una via di mezzo, quella offerta da metadati che possono in parte essere creati automaticamente o dagli autori stessi. La combinazione di questi diversi approcci sembra al momento la strada per migliorare l'accesso sul web e per ridurre i costi derivanti dalla preparazione esclusivamente manuale dei metadati.

Questo è dunque oggi il contesto in cui opera chi è impegnato in progetti di biblioteca digitale. Ciò non toglie che ulteriori sperimentazioni basate su strumenti software sempre più raffinati, architetture tecnologiche innovative e progressi importanti nell'elaborazione del linguaggio naturale emergano a breve o medio termine, offrendo soluzioni diverse, orientate ad una flessibilità nell'uso di metadati specifici ancora più spinta, capaci allo stesso tempo di consentire la condivisione delle risorse informative ad

un livello sempre più esteso e allo stesso tempo preciso. E' probabile che nei prossimi anni siano visibili sviluppi in questo senso.

Sul fronte italiano questi sviluppi e le relative problematiche sono seguite con interesse e vi sono ormai diverse sperimentazioni la cui importanza sta non solo nei risultati raggiunti e nei prodotti finora offerti, ma anche nell'esperienza che i partecipanti ai progetti stanno acquisendo. Si deve però riconoscere che manca ancora una conoscenza diffusa delle potenzialità e insieme delle problematiche dei metadati, nonché degli strumenti oggi disponibili, a pagamento e di pubblico dominio, che consentono di gestire i metadati all'interno di applicazioni per l'accesso a collezioni digitali.

Anche se le biblioteche, gli archivi, i musei e in generale il settore dei beni culturali, per tradizione attenti alla specificità delle loro risorse e alla costruzione di servizi per un pubblico selezionato, sono stati fra i primi a cimentarsi nella definizione dei metadati per il trattamento di oggetti digitali, non si è ancora creata nel nostro paese una massa critica di collezioni e progetti tale da costituire una base sicura su cui modellare le scelte relative ai metadati: occorre quindi da una parte esaminare attentamente e criticamente ciò che viene elaborato in materia a livello internazionale, dall'altra documentare analiticamente e verificare le soluzioni finora adottate nei progetti nazionali.

Inoltre altri settori dovranno essere sempre maggiormente coinvolti nella definizione e preparazione dei metadati, come le università per la messa a disposizione sul web di materiale didattico e di ricerca, la pubblica amministrazione centrale e gli enti locali per i loro servizi di governo elettronico destinati ai cittadini, il vasto e variegato mondo delle aziende per la documentazione e gestione dei loro dati, interni ed esterni, impiegati per il loro funzionamento, per l'interazione con i consumatori e per la gestione del commercio elettronico. Tutti questi mondi infatti trattano informazioni e come tali necessitano di metadati per identificarle e comunicarle in modo adeguato agli utenti della rete.

## 5.7. Raccomandazioni

Al termine di questa analisi sembra opportuno tracciare alcune linee guida, indirizzate da una parte agli organismi istituzionali perché garantiscano un supporto ai progetti di biblioteca digitale nella specifica area dei metadati e dall'altra destinate a chi è direttamente coinvolto in progetti di creazione e accesso a collezioni digitali.

### ***a) Supporto sistematico ai progetti di biblioteca digitale***

E' essenziale avviare un'attività ufficiale, con caratteristiche di stabilità e affidabilità, di documentazione, consulenza e formazione sullo specifico argomento dei metadati. Una iniziativa di rilievo in corso è costituita dai gruppi di studio creati dall'Istituto centrale per il catalogo unico, i cui lavori sono documentati nel sito web dell'ICCU. Il portale della Biblioteca Digitale Italiana istituito dal Ministero per i beni culturali, ora all'inizio della sua operatività, sembra la sede ottimale per offrire i servizi necessari a chi sviluppa progetti di biblioteca digitale, con una sezione apposita sulla vasta area dei metadati.

Ciò che è urgente è l'istituzionalizzazione del servizio stesso a favore di un pubblico italiano ormai vasto che necessita di guida e conoscenze scientifiche di alto livello in materia.

I contenuti dei servizi di supporto sono essenzialmente: a) segnalazione (e in parte traduzione) di documentazione aggiornata e opportunamente selezionata sugli schemi di metadati disponibili, su progetti (inclusi obiettivi, soluzioni adottate in tema di metadati, risultati, problemi risolti e da affrontare), su strumenti disponibili per la creazione e gestione di metadati; b) organizzazione di seminari e dibattiti su temi specifici, come singole problematiche comuni a più schemi, ad esempio la codifica dei metadati e le problematiche sull'interoperabilità, oppure il trattamento di particolari risorse tematiche o legate a specifici supporti digitali; c) attività di formazione e di fornitura di linee guida per la gestione dei metadati, per le mappature fra sistemi, per l'integrazione fra formati bibliografici e schemi di metadati

### ***b) Cooperazione su larga scala fra istituzioni e aziende***

La cooperazione nella definizione e nell'utilizzo di metadati relativi all'informazione trattata fra istituzioni operanti nel mondo delle biblioteche, musei, archivi, ma anche della pubblica amministrazione, centrale e periferica, nonché delle aziende e dell'industria è un'attività da avviare con urgenza, per l'inevitabile necessità di scambio di risorse nell'organizzazione di servizi di biblioteca digitale. In particolare le biblioteche, da sempre attive nella predisposizione di strumenti per la catalogazione, classificazione e accesso all'informazione, dovrebbero assumere un ruolo guida, sensibilizzando e coinvolgendo gli altri settori a documentare e gestire i loro materiali seguendo standard semantici e architettonici che consentano una integrazione di servizi ad un livello esteso.

### ***c) Linee guida e raccomandazioni per chi implementa servizi di biblioteca digitale***

Gli specifici progetti dovrebbero uniformarsi a regole che sono ormai piuttosto collaudate sulla scelta e sull'utilizzo dei metadati. Alla fase di definizione dei metadati occorre far precedere quella relativa all'analisi delle risorse da produrre o comunque gestire, dei servizi da fornire, dell'utenza a cui sono destinati e delle funzioni da svolgere sugli oggetti trattati. Solo allora si procederà alla scelta dei metadati, consultando i registri e la documentazione esistente, operando le scelte appropriate sulla base degli standard esistenti, di tipo generale e specifico, secondo le particolari risorse gestite. Un compito essenziale, oggi trascurato da molti progetti, è quello di documentare dettagliatamente e secondo gli standard gli schemi utilizzati e i profili eventualmente creati.

Si dovrà inoltre ricorrere a strumenti software sperimentati e disponibili sul mercato per la produzione, derivazione e controllo dei metadati, dopo un'attenta analisi dell'esistente e delle prestazioni in progetti simili.

Occorre pertanto attivare scambi e comunicazioni ad ampio spettro con il resto delle sperimentazioni e applicazioni a regime a livello internazionale, dalle quali il nostro paese può derivare indicazioni importanti, ma alle quali può aggiungere risultati preziosi, sia per quanto riguarda i contenuti e

quindi le risorse rese accessibili dai progetti nazionali, sia per le soluzioni tecniche e metodologiche che saranno adottate per la definizione e l'attribuzione dei metadati nelle specifiche applicazioni.