

L'importanza dei metadati nella costruzione della rete globale

Riflessioni ai margini della Conferenza mondiale del DCMI Firenze, 13-17 ottobre 2002

di Nicola Benvenuti

La Conferenza mondiale del Dublin Core Metadata Initiative (DCMI) tenutasi a Firenze nell'ottobre 2002, si è rivelata, come le precedenti, ricca di esperienze e progetti confermandosi come punto di osservazione privilegiato sullo sviluppo delle iniziative sull'informazione di rete e sulle tecnologie emergenti. Di grande interesse e per molti aspetti illuminanti sono risultati i *tutorials* su vari aspetti del DC (struttura, codifica, semantica, registri) e le *keynotes* di Van de Somple (uno dei fondatori di OAI, Open Archive Initiative), Eric Miller (membro del W3C, World Wide Web Consortium, per la costruzione del *semantic Web*), Stuart Weibel (presidente del DCMI)¹. La conferenza tuttavia non ha registrato le presenze che sarebbe stato lecito attendersi, soprattutto da parte di operatori italiani, sia bibliotecari che ricercatori e operatori di imprese, anche se un *tutorial* sul DC tenuto da Susanna Peruginelli nell'ambito di Bibliocom ha avuto notevole successo. Va detto che per una serie di sfortunate circostanze la Conferenza mondiale del DC cadeva in contemporanea o in stretta vicinanza con altri importanti appuntamenti: in Italia Bibliocom, l'appuntamento certamente più importante degli operatori delle biblioteche e dei servizi connessi, e a Ginevra il 2. convegno del CERN su OAI che ha attirato molte personalità e operatori internazionali, in parte passati anche da Firenze, come gli stessi Van de Sompel e Eric Miller.

Permane tuttavia una certa diffidenza verso l'uso stesso del DC, che viene visto erroneamente come concorrente di formati più complessi, quale lo stesso MARC, mentre non si coglie il ruolo di infrastruttura della biblioteca digitale che questo schema per tanti versi elementare e scontato può svolgere. Il fatto è che oggi l'informazione passa attraverso canali assai più articolati di quelli conosciuti fino a pochi anni fa, mentre lo sviluppo delle conoscenze dipende sempre più da un approccio sistemico, trasversale, alle fonti dell'informazione sia di tipo tradizionale che di tipo

NICOLA BENVENUTI, Università di Firenze, Biblioteche di scienze tecnologiche, via Micheli 2, 50121 Firenze, e-mail benvenuti_n@unifi.it.

Salvo quando diversamente indicato gli indirizzi Internet erano attivi al gennaio 2003.

¹ DC-2002: *Metadata for e-communities: supporting diversity and convergence*. Florence, October, 13-17, 2002, in linea <<http://www.bncf.net/dc2002/index.html>>.

nuovo². Inoltre la necessità di rendere più veloce diffusione, condivisione e trasferimento dei risultati delle ricerche e il mutamento dei meccanismi per la loro valorizzazione (la cosiddetta *value chain*), sta profondamente modificando i consolidati meccanismi di trasmissione delle conoscenze, mettendo in discussione i ruoli tradizionali del mondo del libro³. In questo contesto l'interoperabilità tra basi dati nate per esigenze specifiche o semplicemente in contesti applicativi diversi, diventa un punto cruciale e irrinunciabile di qualsiasi progetto di sistema informativo efficace.

Al di là di queste considerazioni è forse necessaria anche una riflessione su una situazione nazionale caratterizzata dalla scarsità di applicazioni che utilizzano i metadati attorno a cui costruire competenze e aggregare processi di apprendimento e innovazione. Non è casuale che la Conferenza mondiale del DCMI si sia svolta a Firenze grazie alla cooperazione della Biblioteca nazionale centrale, assai sensibile verso questi temi e impegnata in importanti progetti internazionali, come adesso TEL⁴, dell'Istituto universitario europeo che tra l'altro ha presentato alla conferenza un progetto di *database* federati⁵, dell'Università di Firenze, presente con l'interessante e assai apprezzato progetto di editoria elettronica DAFNE⁶, il Museo di storia della scienza promotore di progetti basati sulle nuove tecnologie, l'AIB Toscana, all'avanguardia nell'aggiornamento professionale, e ultimi ma certo non meno importanti, gli assessorati alla cultura del Comune di Firenze e alla Regione Toscana, da sempre attenti alle problematiche dell'informazione bibliografica e, nell'ambito dello *e-government*, dei servizi ai cittadini.

L'importanza dei metadati e l'interazione tra gli schemi

Il DC è stato forse il primo schema che ha attirato l'attenzione del mondo delle biblioteche, ma più in generale del vasto pubblico, sui metadati. Il loro utilizzo non è, per altro, una novità, dato che nasce dalla logica stessa della programmazione, fornendo i metadati istruzioni alla macchina su come trattare i dati: un esempio, in ambi-

2 Cfr. *The Information Grid. Ariadne reports on a one-day workshop on an "interoperable environment to support research, learning and teaching"* held at the e-Science Institute in Edinburgh, April 30, 2002, <<http://www.ariadne.ac.uk/issue32/information-grid/intro.html>>. Si veda in particolare: Tony Hey – Anne Trefethen, *The data deluge: an e-science perspective*, <<http://www.research-councils.ac.uk/escience/documents/DataDeluge.pdf>>.

3 Si veda il dibattito attorno al progetto OAI sintetizzato a Firenze nella *keynote* di H. De Sompel <<http://www.bncf.net/dc2002/presentations/DCMI-hvds-Florence-200210.ppt>> e i numerosi interventi di Stevan Harnard. Non insisto su questi temi su cui sono recentemente apparsi interventi nel «Bollettino AIB».

4 Theo van Veen – Robina Claypan, *Metadata in the context of the European Libraries Project*, in *DC-2002: metadata for e-communities: supporting diversity and convergence 2002. Proceedings of the international conference on Dublin Core and metadata for e-communities, 2002, October 13-17, 2002, Florence Italy*, Firenze: Firenze University Press, 2002, p. 19, <<http://www.bncf.net/dc2002/program/ft/paper2.pdf>>.

5 Marco Pirri – Chiara Pettenati – Dino Giuli, *Design of a Federation service for digital libraries: the case of historical archives in the PORTA EUROPA Portal (PEP) Pilot Project*, p. 157, <<http://www.bncf.net/dc2002/program/ft/paper18.pdf>>.

6 Antonella Farsetti – Valdo Pasqui, *District architecture for networked editions: technical model and metadata*, p. 7, <<http://www.bncf.net/dc2002/program/ft/paper1.pdf>>.

to bibliotecario, può essere costituito dalla sequenza di numeri che nel record MARC precedono il testo e che identificano i campi in cui posizionare le varie parti della registrazione. L'interesse per i metadati deriva oggi dal tentativo di utilizzarli non solo per definire caratteristiche, funzioni e comportamenti delle risorse informative ma anche per codificare le relazioni semantiche tra di loro, fino al tentativo di far sì che le macchine riproducano la capacità di fare dichiarazioni tipica del linguaggio umano e si comportino come se ne comprendessero il significato⁷.

Da un punto di vista sociologico la diffusione dei metadati indica la partecipazione e il protagonismo delle comunità di interesse (la comunità delle biblioteche, della *e-education*, dei musei, dello *e-government*, dello *e-commerce* ecc.) nello sviluppo di linguaggi per comunicare significati e valori in rete. Questa osservazione rinvia alla dimensione partecipativa della tecnologia di rete che fin dall'inizio caratterizza Internet ed evidenzia il fatto che il suo sviluppo ha bisogno di alimentarsi non solo di conoscenze specialistiche ma soprattutto della adesione e del consenso di queste comunità di interesse indipendenti, siano esse costituite da soggetti istituzionali (biblioteche, università, musei, enti pubblici ecc.) o operatori professionali (aziende di software, fornitori di servizi di rete, *webmaster* ecc.).

L'apparenza di Internet come luogo caotico e variegato rimanda al fatto che la rete si sviluppa secondo un flusso *bottom up* caratterizzato dalla creatività e dagli sforzi di molti soggetti, rendendo irrealistico un approccio *top down* alla standardizzazione, specialmente nell'ambito di molteplici domini. In questo senso la definizione dei metadati è più il risultato di un processo di ricerca dell'equilibrio tra le esigenze all'interno e tra diverse comunità, che non un risultato definito una volta per tutte. Questo stesso ragionamento impronta l'impostazione data dal Consorzio del W3C alla costruzione del "Web semantico" che, scontando il proliferare di una miriade di linguaggi in rete, pone come unica condizione che ad un indirizzo univoco e persistente sia indicato in modo normalizzato quali regole ne permettono la comprensione⁸.

Ciò che qui interessa sottolineare in base a queste brevi considerazioni è che da semplici formati o schemi per descrivere oggetti informativi, i metadati tendono a diventare le "parole" di un linguaggio relazionale per trasmettere significati leggibili dalle macchine, in modo che sia salvaguardato il principio fondamentale che sottende l'architettura del WWW, che cioè sia semplice per chiunque esprimere qualcosa su di una risorsa esistente⁹. L'implicazione di questo approccio è che, pur nell'ambito di questa libertà di espressione, si costituisca un accordo di fondo su ciò che le varie comunità di interessi ritengono utile esprimere ai vari livelli di complessità richiesti, in modo da rendere possibile la comunicazione fra sistemi.

Per permettere a chiunque libertà di costruire i propri linguaggi il W3C ha elaborato lo *Extensible Markup Language* (XML) un metalinguaggio che realizza la separa-

⁷ Il rapporto tra "rappresentazione della conoscenza" e *semantic Web* è affrontato da Kim H. Veltman del McLuhan Institute di Maastricht. Cfr. *Syntactic and semantic interoperability: new approaches to knowledge and the semantic Web*, «The New Review of Information Networking», vol. 7 (2001), <<http://www.i-massweb.org/Workpackages/syntacticNovember2001.doc>>; e *Digital reference rooms: access to historical and cultural dimension of knowledge*, <http://www.isoc.org/isoc/conferences/inet/99/proceedings/2b/2b_1.htm>. Per l'analisi del ruolo dei metadati nella rappresentazione della conoscenza, la riflessione sulla linguistica in un'ottica strutturalista offre importanti motivi di riflessione; cfr. Umberto Eco, *La struttura assente*, Milano: Bompiani, 1968, p. 285 e ss.

⁸ Si vedano le osservazioni di Tim Berners-Lee, in: <<http://www.w3.org/DesignIssues/Overview.html>>.

⁹ Cfr. <<http://www.w3.org/TR/RDF-schema/#RDFnotcite>>.

zione tra contenuto e visualizzazione. Esso è più semplice di SGML (Standard Generalized Markup Language), ma altrettanto flessibile e più maneggevole, e permette di creare liberamente metadati per le diverse esigenze delle comunità. Le potenzialità di XML non sono ancora pienamente sviluppate né sfruttate; in ambito bibliotecario la traduzione del MARC in Data Type Definition (DTD) di XML potrebbe sostituire il “formato” MARC e dar luogo a sistemi bibliotecari interamente scritti in XML¹⁰.

Un qualsiasi elemento, ad esempio “autore”, può però avere significati diversi in differenti contesti: per disambiguarne il significato si può far riferimento alla descrizione e struttura dello schema di metadati di cui l’elemento fa parte. XML prevede questo tipo di controllo attraverso il meccanismo del *namespace*, cioè attraverso la possibilità di indicare in modo univoco di quale schema (*scheme*) di metadati reperibile in rete ad un indirizzo persistente (URI) l’elemento fa parte e quindi quale area semantica copre e quali regole ne governano l’utilizzo: se i valori di un certo elemento devono assumere un determinato formato, oppure se devono essere scelti tra quelli di una lista predefinita, se l’elemento è/non è ripetibile ecc. In questo preciso senso i vari schemi (detti anche linguaggi, o interfacce) sono auto-descrittivi (*self-describing*) e perciò immediatamente manipolabili dalle macchine.

La conseguenza forse di maggior rilievo di questa architettura dei dati è che più schemi di metadati possono convivere in uno stesso file XML, realizzando ognuno una comprensione “parziale”, relativa cioè alla propria area di competenza semantica, e aprendo in tal modo la via all’uso modulare di più schemi per descrivere o gestire uno stesso documento.

L’aspetto più propriamente linguistico dei metadati si manifesta allorché diviene possibile fare affermazioni sulle risorse (si intende qui per risorsa qualsiasi oggetto identificabile in modo univoco da un URI). Tuttavia mentre per gli umani due affermazioni come: – l’autore del documento 1 è Silvio De Luca
– Silvio De Luca è l’autore del documento 1
sono assolutamente equivalenti, per la macchina sono stringhe completamente diverse.

Per costruire affermazioni un modo univoco è stato ideato il Resource Description Framework (RDF)¹¹, una applicazione XML definita dal W3C. Secondo RDF la codifica per esprimere affermazioni parte dalla considerazione di una risorsa, identificata da un URI, che svolge il ruolo di *soggetto*, a questa risorsa può essere imputata una proprietà, che assume il valore di *predicato* e il valore della proprietà costituisce un *oggetto*: ne deriva la forma normalizzata:

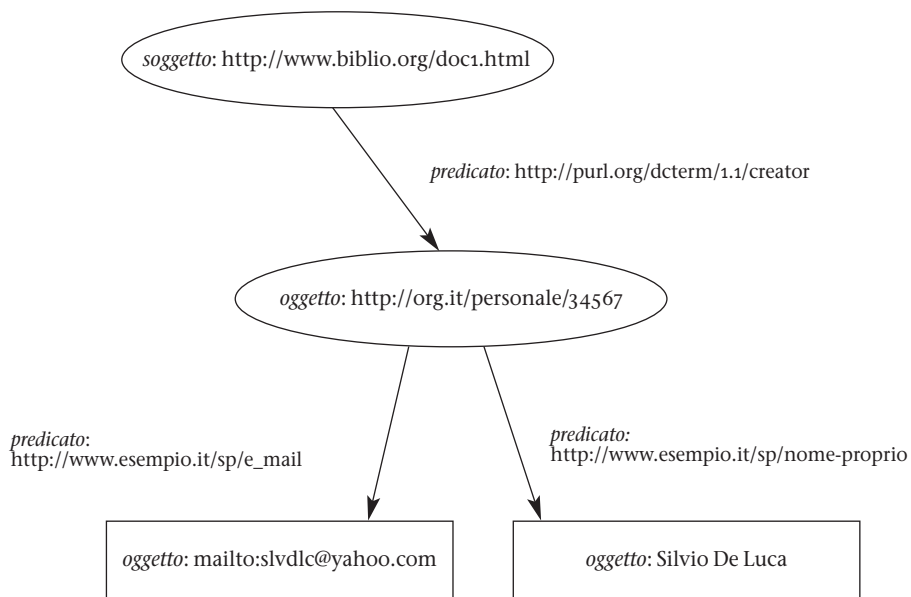
<i>soggetto</i>	<i>predicato</i>	<i>oggetto</i>
documento 1	ha come autore	Silvio De Luca

¹⁰ Cynthia Hodgson, *The RFP writer’s guide to standards for library systems*, Bethesda (MD): NISO Press, c2002, p. 47. Scaricabile da: <<http://www.niso.org/>>.

¹¹ *Resource Description Framework (RDF) modello e sintassi*, <<http://www.xml.it/RDF/REC-dc-syntax-19990222-it.html>>. Trad. it. di *Resource Description Framework (RDF) model and syntax specification*, <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>. Cfr. anche: Eric Miller, *An introduction to the Resource Description Framework*, in: «D-Lib Magazine», May 1998, <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Durante la scrittura di queste pagine è uscito: Frank Manola – Eric Miller, *RDF primer : W3C working draft 23 January 2003*, <<http://www.w3.org/TR/2003/WD-rdf-primer-20030123/>>.

Perché questa affermazione sia leggibile dalle macchine bisogna, come si è detto, che documento₁ sia identificabile in modo univoco sulla rete, ad esempio con l'URI: <http://www.biblio.org/doc1.html>. Il predicato è tipicamente un metadato perché esprime informazioni sul dato [doc1.html](http://www.biblio.org/doc1.html), ad esempio che ha un autore: questa proprietà è espressa facendo riferimento all'URI dello schema di metadati in questione, in questo caso, se usiamo il DC: <http://purl.org/dc/elements/1.1/creator>. Silvio De Luca è in questo caso un semplice valore costante detto *literal* (che con XML può essere specificato come stringa di caratteri, come numero o come data)

Si noti che potrebbe essere utile ai fini del sistema informativo che stiamo delineando fare affermazioni anche sull'oggetto, nel nostro caso Silvio De Luca, per indicare qualità specifiche: ad esempio che ha un indirizzo e-mail, lavora in un ente, ha una certa qualifica ecc. In tal caso però la relazione diventa più complessa perché quello che è l'oggetto, o valore, di una asserzione biunivoca diventa al contempo soggetto di altre proprietà che con il primo soggetto non hanno relazione diretta. Per risolvere questo problema RDF prevede che l'oggetto possa diventare un contenitore di affermazioni, in tal modo ricostituendo un gruppo di affermazioni logicamente separate. Il meccanismo più immediato per estendere l'asserzione iniziale è che Silvio De Luca possa a sua volta essere trattato come risorsa cioè, come abbiamo visto, abbia un identificatore univoco, e diventi soggetto di altre affermazioni. In questo caso avremmo stabilito una relazione tra affermazioni costruendo un ipotetico periodo che qui raffiguriamo in forma di grafo:



(In questo grafo utilizzo l'elemento *creator* dello schema *dcterm* e gli elementi *nome-proprio* ed e-mail di un ipotetico schema *sp*. Convenzionalmente le risorse sono rappresentate con ellissi, le proprietà con archi orientati etichettati e i valori *literal* con rettangoli).

Da un punto di vista concettuale, però, la risorsa <<http://org.it/personale/34567>> non ha alcuna funzione e infatti RDF permette con vari formalismi di costituire dei contenitori vuoti (dette *anonymous resources* o, nella rappresentazione grafica, *blank nodes*) a cui relazionare le proprietà. Ai fini di queste note l'importante è comprendere che con questi meccanismi la relazione biunivoca può estendersi utilizzando anche diversi schemi di metadati. RDF permette poi di fare affermazioni su affermazioni, ad esempio per confermare o smentire una dichiarazione (l'operazione è detta *reifificazione*).

Vale la pena di accennare allo schema RDF per la formalizzazione dei vocabolari degli schemi, cioè su come si possano articolare gli elementi costitutivi degli schemi di metadati. RDF propone di suddividerli in classi (*class*) e in sottoclassi (*subclass*), con legami di tipo gerarchico. Una classe particolare è costituita da proprietà (*property*) che non solo gode della struttura gerarchica tipica delle classi, cui si applica la proprietà transitiva, ma prevede anche la possibilità di contestualizzazione, con l'indicazione del dominio (*domain*) cui si riferisce e dell'intervallo di valori (*range*) possibili¹². Mirando ad una formalizzazione delle relazioni gerarchiche e contestuali del vocabolario usato per descrivere e gestire risorse di rete, RDF si prefigura come una grammatica comune ai fini di una migliore integrazione degli schemi di metadati delle diverse comunità¹³. Come si intuisce si tratta di un tentativo di costruire una sorta di "metacodice", o un sistema dei sistemi, che tuttavia per diventare operativo ha ancora bisogno di un lungo cammino.

Oggi per interoperare tra sistemi (cioè schemi) diversi l'unico metodo è quello di "mappare" gli elementi con funzione simile, creando i cosiddetti *crosswalk*. Nella loro applicazione però succede come con i traduttori automatici: nel passaggio da una lingua all'altra si hanno notevoli perdite di significato se non di senso, con risultati che rasentano l'assurdità¹⁴. Da qui il ruolo del DC come *pidgin* (Thomas Baker) o gergo di basso livello semantico su cui mappare, in una relazione uno a uno, gli schemi di metadati più complessi per permetterne l'interoperabilità.

Lo schema del DC

Gli elementi del DC nella versione 1.1 sono 15 e servono ad identificare le risorse informative per poterle localizzare e visualizzare: i 15 elementi non sono obbligatori e sono ripetibili; recentemente è stato aggiunto un nuovo elemento, *audience*, per indicare i destinatari di un determinato oggetto informativo¹⁵.

¹² *RDF vocabulary description language 1.0: RDF schema*. [W3C Working Draft 12 November 2002], <<http://www.w3.org/TR/2002/WD-RDF-schema-20021112/>>.

¹³ *Schemas, Third Workshop: Managing schemas in a multilingual semantic Web* in: <<http://www.schemas-forum.org/workshops/ws3/report.html>>. Si veda anche l'analisi della possibile codifica delle relazioni tra schemi di metadati e *application profile* in: Thomas Baker – Max Dekkers – Rachel Heery – Manjula Patel – Gauri Salokhe, *What terms does your metadata use? Application profiles as machine undestandable narratives*, «Journal of digital information», 2 (2001), n. 2, <<http://jodi.ecs.soton.ac.uk/Articles/v2/io2/Baker>>.

¹⁴ Si veda a proposito l'intervento di Lynne C. Howarth (University of Toronto), *Designing a metadata-enabled namespace for enhancing resource discovery in knowledge bases*, in: *Le risorse elettroniche: definizione selezione, catalogazione: atti del Convegno internazionale, Roma, 26-28 novembre 2001*, a cura di Mauro Guerrini, con la collaborazione di Stefano Gambari e Lucia Sardo. Milano: Bibliografica, 2002, <http://w3.uniroma1.it/ssab/er/relazioni/howarth_eng.pdf>.

¹⁵ La lista degli elementi, qualificatori e schemi del DC con link alle proprietà e agli schemi di codifica, si trova all'indirizzo: <<http://wip.dublincore.org/registry/registryXSLTServlet>>.

Il DC fu creato per descrivere i file di Internet in modo che potessero essere utilizzate dai motori di ricerca. L'idea era che gli autori stessi inserissero delle informazioni nell'area «head» dei file HTML per facilitare la localizzazione e il recupero dei propri documenti. Come è noto il progetto ha funzionato solo in parte. I valori inseriti nei *metatag* da autori e *webmaster* rispecchiano la Babele di linguaggi presenti in rete, evidenziando la molteplicità di modi per indicare una stessa informazione, per non parlare dell'ambito semantico da assegnare a ciascun termine. Per comparire nelle prime posizioni dei motori di ricerca sfruttando il criterio della frequenza delle citazioni, qualcuno ripeteva più volte nei *metatag* le parole che presumibilmente un utente avrebbe cercato per giungere ad una determinata informazione, indipendentemente dal contenuto effettivo del file, falsando il risultato della ricerca¹⁶.

Come era già emerso nella pratica bibliotecaria, si riproponeva così la necessità di una catalogazione di risorse selezionate in base a criteri di qualità ad opera di specialisti anche se, data la dimensione di espansione del WWW e le continue modifiche dei file, in un formato assai più semplice e quindi meno costoso dei formati catalografici tradizionali quali il MARC. La prima applicazione del DC è stata infatti nella costruzione dei *Subject Gateway*, cataloghi di risorse di rete selezionate secondo criteri di qualità e di tipo disciplinare, per facilitare la ricerca e condivisione di informazioni in ambienti accademici.

Il secondo impulso per la creazione del DC fu quello di disporre di uno schema di elementi essenziali e di larghissima comprensione per condurre ricerche tra risorse di diversa tipologia (*cross domain search*): gli elementi del DC permettono infatti sia di rendere reperibili file in formati non testuali (identificabili nel file html solo dalle parole del contesto al link o dal testo alternativo, nel caso di immagini), sia di condurre ricerche su tipologie di documenti catalogati secondo schemi specifici e complessi. Questa seconda funzione rinvia al ruolo del DC di linguaggio di basso livello semantico che abbiamo segnalato prima. Grazie a questo ruolo il DC si presta a costituire lo strumento per strutturare i documenti in contesti di *knowledge management* e in generale di archiviazione e recupero di documenti senza la mediazione di bibliotecari.

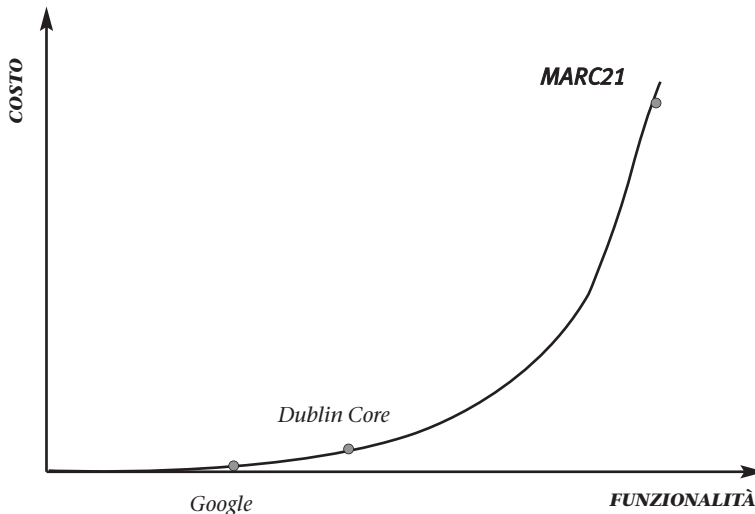
Via via che diverse comunità articolavano la propria presenza in rete sono nati nuovi schemi di metadati e nuove esigenze e funzionalità venivano ad essere soddisfatte per questa via: a seconda della funzione i metadati si possono suddividere in descrittivi, amministrativi, strutturali, per la usabilità ecc. Naturalmente il grado di complessità degli schemi varia a seconda della precisione richiesta dal servizio che si vuol implementare. Sulla scorta di un intervento di Lorcan Dempsey e Rachel Heery¹⁷ si possono distinguere diverse tipologie di metadati a seconda del loro uso e complessità (la tabella è riportata in modo molto semplificato):

16 Analogamente oggi si creano pagine html solo per linkare a risorse di rete e forzare il criterio di ordinamento in base alla popolarità di motori di ricerca come *google*; si veda Danny Sullivan, *What is a bridge or doorway page?* <<http://searchenginewatch.com/webmasters/bridge.html>>.

17 Lorcan Dempsey – Rachel Heery, *Metadata: a current view of practice and issues*, apparso in «Journal of Documentation», 54 (1998), n. 2, p. 145-172, <<http://www.ukoln.ac.uk/metadata/publications/jdmetadata/>>.

	Semplici	Strutturati	Ricchi
Esempi	Costruiti in modo automatico dai robot dei motori di ricerca	DC, IAFA, GILS	MARC, CIMI, EAD, TEI, LOM
Ambiente d'uso	Servizi di ricerca globale su Internet; Servizi di indicizzazione del Web	Servizi di ricerca selettivi di Internet e liste di siti	Descrizione di collezioni museali accademiche e importanti archivi di testi
Funzione	Identificazione di un documento	Identificazione e selezione	Identificazione, selezione, valutazione, analisi, documentazione
Creazione	Generati da robot	Robot ed elaborazione manuale	Richiede un'altra qualificazione intellettuale spesso da parte di personale appositamente addestrato
Struttura	Non strutturato	Coppia elemento valore	Qualificatori di sottocampo; Marcatura strutturata

Per visualizzare in modo più preciso la posizione del DC tra gli schemi di metadati esistenti evidenziandone nel frattempo una valutazione del rapporto costi/benefici, è utile questa immagine tratta dal *tutorial* sul DC tenuto da Thomas Baker alla Conferenza di Firenze. A livello di ricerca dell'informazione nella vasta e indistinta rete, il motore di ricerca Google, grazie ai metadati che costruisce in modo automatico (soprattutto quello che conta i link diretti ad un sito, per inferire criteri di "popolarità"), prefigura un costo assai basso per discrete funzionalità di ricerca. Via via che si richiedono maggiori servizi e precisione è però necessario far ricorso a operazioni manuali condotte da personale specializzato e la curva dei costi si impenna con decisione. Il DC, che può essere creato in parte automaticamente e in parte manualmente, ed è costituito da pochi semplici elementi, rappresenta un punto di equilibrio per ricerche tra sistemi informativi complessi di diversi domini.



I 15 elementi della versione 1.1 – che malgrado sia dichiarata obsoleta prendiamo qui a riferimento perché approvata come standard NISO ed è in ogni caso la più conosciuta – sono stati considerati, per alcune esigenze, troppo generici e sono stati quindi affiancati da “qualificatori” (ad esempio *alternative* come qualificatore dell’elemento *title*). Per garantire l’interoperabilità tra sistemi che adottano il *simple DC* (solo i 15 elementi) con quelli che adottano il *qualified DC* è però necessario applicare la cosiddetta regola del *dumb down*, per garantire piena coerenza tra i valori dei qualificatori e quelli degli elementi di primo livello: infatti, se una macchina non avesse implementato lo schema per interpretare un qualificatore (*qualifier*), il valore ad esso riferito dovrebbe comunque ricadere nell’area semantica dell’elemento superiore. La funzione dei qualificatori è, infatti, quella di specificare il significato degli elementi, non di estenderlo.

Sempre ai fini di garantire univocità e coerenza nella descrizione degli oggetti informativi, sono poi stati indicati degli schemi che specificano quali valori singoli elementi dovrebbero assumere: ad esempio per l’elemento *dc.subject* sono consigliati come schemi da utilizzare il LCSH, DDC, MeSH, LCC e UDC. Per l’elemento *dc.type* il DCMI ha elaborato una propria lista di termini¹⁸.

Il processo di approvazione degli elementi e dei qualificatori conosce vari passaggi segnalati da indicatori di status: si parte dallo status di “locale”, per termini usati in singoli progetti, a quello di “proposto” per i termini proposti dai gruppi di lavoro come importanti per la comunità di riferimento, e in corso di esame da parte del Comitato per l’applicazione, fino a “conforme”, per i termini approvati e infine “raccomandato”.

Il seguente diagramma esprime la struttura del DC che abbiamo descritto, avvertendo che in tondo sono indicati i termini con status “raccomandati” e in corsivo i termini “conformi”¹⁹:

Elementi del set di metadati del Dublin Core	Qualificatori degli elementi	Schemi di codifica degli elementi
Title	Alternative	-
Creator	-	- LCSH MeSH
Subject	-	DDC LCC UDC
Description	TableOfContents abstract	-
Publisher	-	-
Contributor	-	-
Date	Created valid available issued modified <i>Accepted</i> <i>Copyrighted</i> <i>Submitted</i>	DCMI Period W3C-DTF

¹⁸ DCMI Usage Board, *DCMI grammatical principles*, [2002-10-06], in linea <<http://dublincore.org/usage/documents/principles/>>.

¹⁹ *Dublin Core qualifiers*, [DCMI Recommendation], <<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>>.

Type	-	Vocabolario delle tipologie del DCMI
Format	Extent	-
	Medium	IMT
Identifier	-	URI
Source	-	URI
Language	-	ISO 639-2
	-	RFC 1766
Relation	isVersionOf	
	hasVersion	
	isReplacedBy	
	replaces	
	isRequiredBy	
	Requires	
	isPartOf	URI
	hasPart	
	isReferencedBy	
	references	
	isFormatOf	
hasFormat		
ConformsTo		
Coverage	Spatial	DCMI Point
		ISO 3166
		DCMI Box
	Temporal	TGN
		DCMI Period
		W3C-DTF
Rights	-	-
Audience	mediator	
	<i>educationLevel</i>	-

I metadati possono essere collegati al documento in diversi modi: possono essere inclusi nel file o possono essere contenuti in un file separato richiamato da un iperlink, magari come *record* di una base dati. Queste diverse soluzioni richiedono codifiche specifiche, così come lo richiede l'adeguamento a nuove versioni dei vari linguaggi utilizzabili o alla comparsa di nuove tecnologie. Ad esempio sono state descritte le codifiche da utilizzare in HTML e il Gruppo di lavoro per l'architettura del DCMI ha elaborato tre *namespace* in RDF, *dc* per il DC versione 1.1, *dcterm* per i qualificatori e *dctype* per il vocabolario delle tipologie. In tal modo il DC si integra nel contesto del *semantic Web*.

Il DC come processo

La "manutenzione" in questo quadro di continuo mutamento appare una funzione strategica di primo piano per adeguare il DC all'innovazione e sviluppo di Internet. Accanto a questo compito necessario per ogni schema di metadati va sottolineata la necessità anche di adottare una serie di strategie per corrispondere all'evoluzione della rete non soltanto come tecnologie, quanto come nuove esigenze e progetti.

A questo duplice compito è preposto, come si è visto, il DCMI. Un breve sguardo alla sua organizzazione²⁰ può forse dare un'idea più chiara di cosa intendiamo quando abbiamo fatto riferimento al valore del DC come processo e non soltanto come prodotto.

²⁰ Cfr. <<http://dublincore.org/>>.

Il DCMI, una emanazione di OCLC, che è proprietaria del marchio, è l'organizzazione che si occupa di mantenere e aggiornare il set di metadati. Si tratta di una organizzazione di tipo "leggero" e reticolare, con una struttura flessibile e aperta, costituita da una Direzione di tre membri (Stuart Weibel, Makx Dekkers e Tom Baker), un Comitato di garanti rappresentato attualmente da 7 personalità internazionali (Denise Bedford, Joseph Busch, Michael Crandall, Lorcan Dempsey, Juha Hakala, Nathalie Leroy, Shigeo Sugimoto) con la funzione di sovrintendere all'indirizzo politico complessivo del DCMI, un Comitato consultivo costituito da 40 esperti di diverse discipline, e un Comitato per l'applicazione (*usage board*) del DC (Thomas Baker, Rebecca Guenther, Diane Hillmann, Traugott Koch, Haruki Nagata, Andy Powell, Roland Schwaenzl, Stuart Sutton; ne fanno poi parte, d'ufficio, i membri della direzione).

Inoltre il DCMI ha relazioni continue con altre attività di standardizzazione e altre iniziative per la definizione di metadati tra cui il CEN (Comité Européen de Normalisation), IEEE/LOM (Institute of Electrical and Electronics Engineers/Learning Object Metadata), IETF (Internet Engineering Task Force), ISO/IEC JTC1/SC29: MPEG Standardization of Coded Representation of Audio, Picture, Multimedia and Hypermedia Information, NDLTD (Networked Digital Library of Theses and Dissertations), OAI (Open Archives Initiative che rende obbligatorio, ma non esclusivo, l'uso del DC), RDF Core Working Group, W3C (World Wide Web Consortium).

Il risultato di questa attività è rappresentato dalla adozione del DC in progetti di istituzioni educative e governative in Australia, Canada, Danimarca, Finlandia, Svezia, Germania, Irlanda, Nuova Zelanda e Regno Unito. Come standard ufficiale è stato approvato dall'ente di standardizzazione americano ANSI/NISO e da quello finlandese e si attende l'approvazione da parte dell'ISO.

Il centro dell'attività del DCMI è rappresentato dai gruppi di lavoro e dai gruppi di interesse, ognuno su un tema attuale e cogente, costituiti da personalità impegnate in progetti di costruzione e definizione di servizi di rete, e dalle Conferenze annuali. Gran parte dell'attività si svolge naturalmente in rete attraverso numerose mailing list e disseminazione di documenti. I gruppi di lavoro sono attualmente 13; vi sono poi 5 gruppi di interesse. I temi su cui operano questi gruppi sono i seguenti:

1. Metadati amministrativi
2. Agenti
3. Architettura
4. Citazioni
5. Descrizione della collezione
6. Metadati per l'educazione
7. Biblioteche
8. Metadati per la funzione pubblica (*government*)
9. Registri
10. Standard
11. Software (*tools*)
12. Vocabolario delle tipologie
13. Guida utenti

I 5 gruppi di interesse riguardano: accessibilità, aziende, ambiente, kernel/ERC, localizzazione e internazionalizzazione.

Tutti i gruppi di lavoro sono cruciali per monitorare e discutere le esigenze che nascono all'interno delle varie comunità e valutarne le conseguenze e ricadute sulla struttura del DC in modo da mantenere il suo ruolo di gergo comune, tra le varie

applicazioni. Dai gruppi di lavoro nascono anche i cosiddetti *application profile* (AP) cioè adattamenti del DC alle esigenze della comunità di riferimento con guide e per la loro implementazione. Gli schemi di metadati ispirati al DC ma integrati con elementi locali sono invece detti “dialetti”²¹. L’organizzazione del DCMI si caratterizza pertanto per essere oltre che leggera e reticolare, altamente collaborativa: grazie a questa articolazione è possibile attuare strategie di adattamento e innovazione.

Strategia di adattamento

Il DC è sottoposto a diverse tensioni che richiedono scelte ponderate e la ricostituzione di nuovi equilibri: tra queste tensioni abbiamo visto ad esempio quella tra semplice e generico verso ricco e specifico o quella tra comune a più domini e specifico di un dominio, o ancora quella che deriva dall’adeguamento alla evoluzione della tecnologia di rete ecc. Per ovviare a queste tensioni sono state definite una serie di strategie alcune delle quali abbiamo già esaminato: ad esempio l’uso di qualificatori che specificano la semantica degli elementi principali secondo le esigenze delle varie comunità, e l’adozione di vocabolari controllati per esprimere i valori da attribuire agli elementi e ai qualificatori.

A proposito dei vocabolari non si tratta solo degli schemi già approvati nella tabella di cui sopra, ma di nuove liste di termini anche brevi ma univoche e controllate: ad esempio il qualificatore *audienceLevel* nato in ambito *education* per indicare il destinatario di un corso di lezioni, richiede un vocabolario tale che si possano stabilire corrispondenze nel modo più univoco, tra i sistemi educativi di diversi paesi; analogo problema sorge nell’applicazione di questo qualificatore all’informazione di fonte pubblica, data la diversità dei sistemi politici, anche solo occidentali. In altri casi, come per il profilo di applicazione per le biblioteche, le liste sono riprese da altri sistemi, soprattutto MARC21 (ad esempio *relator list* per l’elemento *dc.contributor*; v. avanti). Abbiamo anche visto come i lavori preparatori di ricognizione sul campo ed elaborazione autonoma conduca a proposte che sono sottoposte ad un articolato iter di approvazione.

Questa attività di adattamento e aggiornamento culmina nella definizione dei profili di applicazione (*Application Profile* o AP): questo concetto è emerso all’interno del DCMI come modo per dichiarare quali elementi sono usati in una determinata applicazione o progetto. Essi sono elementi presi da uno o più *namespace*, combinati insieme e ottimizzati per una particolare applicazione locale²².

Ma, oltre agli adattamenti e specificazioni interni al modello DC, vi è la necessità di confronto e di interoperabilità con schemi di metadati nati ad opera di comunità diverse dal DCMI in modo autonomo o per esigenze non ancora emerse. Per garantire interoperabilità tra gli schemi di metadati è necessario definire con precisione l’area semantica che ciascun elemento è destinato a esprimere in modo da evitare la proliferazione di elementi per la stessa funzione. Uno strumento per raggiungere questo obiettivo è il registro (*registry*) dei metadati, cioè la fissazione in forma ufficiale e pubblica degli identificatori e della semantica degli elementi adottati: nel definire uno schema per esigenze particolari, devono essere consultati i registri per verificare se gli elementi voluti non siano già stati attivati in altri schemi.

21 Una lista di tali “dialetti” è presentata in Metaform, la base dati con funzioni di registro dei formati di metadati costituito nell’ambito del progetto META-LIB delle biblioteche tedesche: <<http://www2.sub.uni-goettingen.de/metaform/>>.

22 Rebecca Guenther, *Library application profile*: <<http://dublincore.org/documents/library-application-profile/>>.

Per meglio comprendere il ruolo e l'attività del DCMI è opportuno esaminare più da vicino l'attività almeno di alcuni gruppi di lavoro.

Il gruppo per il registro si occupa della costruzione del registro del DC, inteso come vocabolario degli elementi, qualificatori e schemi, definiti in modo comprensibile sia da esseri umani che leggibile dalle macchine, quindi espresso sia in linguaggio naturale che in XML/RDF. A tale scopo nel sito del DCMI è implementato un motore di ricerca multilingue sugli elementi, qualificatori e schemi del DC, con lo schema RDF relativo e la spiegazione dell'uso consentito²³.

L'attività del gruppo di lavoro può far riferimento a importanti precedenti come quello di Ukoln (UK On Line Network) che ha sviluppato un registro con funzioni di *crosswalk*²⁴, il GMR (German Metadata Registry)²⁵, e SCHEMAS adesso sostituito da CORES: a Forum on Shared Metadata vocabularies, per progetti condotti nell'ambito del programma IST della CE²⁶. Alla conferenza di Firenze è stato presentato il prototipo del software SCART (Schema Creation And Registration Tool) per aiutare i progettisti di sistemi informativi a costruire i propri schemi in modo corretto e soprattutto evitando sovrapposizioni con altri schemi esistenti: chi si propone di costruire un sistema informativo utilizzando i metadati per esprimere le proprie ontologie, dovrebbe innanzitutto consultare i registri per trarne gli elementi di cui ha bisogno, o eventualmente per crearne di nuovi che coprano un'area semantica non contemplata o non sufficientemente definita²⁷.

Altro gruppo cruciale è quello sulla architettura dei dati perché dalla sua capacità di tradurre il DC nei linguaggi emergenti dipende la possibilità di diffusione e operabilità dello schema nel tempo. L'obiettivo del gruppo è pertanto quello di elaborare modelli, strategie e guide per l'applicazione del DC tramite le principali tecnologie Web: XML, RDF, XHTML ecc. Il primo documento del gruppo è il noto articolo di Kunze sulla codifica del DC in HTML²⁸, ancora valido anche se attualmente sono in corso di revisione le specifiche per esprimere il DC in HTML/XHTML ed esiste già un *working draft* ad opera di Andy Powell, datato 11 agosto 2003²⁹. Il gruppo ha poi elaborato lo schema del DC in XML/RDF, che ha raggiunto lo status di raccomandazione del DCMI. Oltre al *namespace dc*, sono stati descritti in RDF i *namespace dcterms* che contiene la definizione e le specifiche dei qualificatori degli elementi principali, e *dctype* che codifica il vocabolario delle tipologie dell'elemento *type*³⁰.

Il 2 aprile 2003 è uscita la versione finale delle *Guidelines for implementing Dublin*

23 Cfr. <<http://dublincore.org/search/searchServlet>>; nel sito esiste anche la possibilità di cercare, registrare e gestire vocabolari controllati: <<http://wip.dublincore.org/schemes/index.html>>.

24 Cfr. <<http://desire.ukoln.ac.uk/registry/index.php3>>.

25 Cfr. <<http://www.mpib-berlin.mpg.de/dok/metadata/gmr/gmr1d.htm>>.

26 Cfr. <<http://www.schemas-forum.org/>>; <<http://www.cores-eu.net/>>.

27 Cfr. <<http://www.bncf.net/dc2002/program/ft/paper14.pdf>>.

28 John A. Kunze, *Encoding Dublin Core metadata in HTML*: <<http://www.ietf.org/rfc/rfc2731.txt>>.

29 Andy Powell, *Expressing qualified Dublin Core in HTML/XHTML meta and link elements*, <<http://www.ukoln.ac.uk/metadata/dcmi/dcq-html/>>.

30 Beckett, Dave – Eric Miller – Dan Brickley, *Expressing simple Dublin Core in RDF/XML*: <<http://dublincore.org/documents/2002/07/31/dcmes-XML/>>; il *ns dcterms* si trova a: <<http://dublincore.org/2002/08/13/dcq#>>; il *ns dctype* si trova a: <<http://dublincore.org/2002/09/20/dctype#>>.

Core in XML (Andy Powell – Pete Johnston) e il 15 maggio la proposta di Raccomandazione dal titolo: *Expressing Qualified Dublin Core in RDF / XML* (Stephan Kokkelink – Roland Schwänzl). È stata poi presa in considerazione l'eventualità di elaborare delle guide all'architettura del DC (Stu Weibel – Makx Dekkers) e per esprimere il DC in XML/RDF (Roland Schwänzl – Ana Alice Baptista – Harry Wagner)³¹.

Molto attivo è anche il gruppo che si occupa dei metadati per le biblioteche, presieduto da Rebecca Guenther della Library of Congress. In questo settore già così affollato di iniziative su formati e schemi di metadati, il possibile uso del DC nelle biblioteche può essere:

- favorire gli scambi tra diversi sistemi che usano differenti formati o standards di metadati (col nome di *Bath profile* il DC è già utilizzato come uno dei profili di Z39.50);
- raccogliere (*harvesting*) metadati da fonti interne ed esterne al settore bibliotecario per costituire cataloghi di risorse di rete (il riferimento potrebbe essere qui alla attività di costruzione dei Subject Gateway: lo schema di metadati del progetto della CE, Renardus, è riconosciuto come *AP* del DC);
- permettere la creazione di semplici *record* catalografici di risorse in una varietà di sistemi (vengono qui in mente le librerie *on line*, come Amazon);
- trasmettere dati in MARC ad altre comunità, attraverso la conversione in DC (come avviene anche nel progetto CORC);
- permettere l'acquisizione di metadati creati da non bibliotecari che usino il DC (il riferimento potrebbe essere all'attività di auto archiviazione come prevista da OAI in cui un *subset* del DC è richiesto come struttura minima per l'identificazione dei documenti).

Il gruppo ha sviluppato un *application profile* (*AP*) chiamato: *DC lib*. L'attenzione in questo momento è concentrata sulla questione della rappresentazione degli agenti (a cui ci si riferisce con l'acronimo CCP: *Creator, Contributor, Publisher* e su cui esiste un apposito gruppo di lavoro) e nella discussione di progetti in corso, tra cui soprattutto *TEL The European Library* che mira a costruire un sistema di accesso alle biblioteche nazionali europee in cui il *DC-lib AP* è usato come punto di partenza della struttura dei metadati ma con altri elementi ritenuti necessari. Alla Conferenza di Firenze è stato discusso anche lo schema di metadati per la descrizione delle collezioni possedute dalle biblioteche. Il fatto è che lo *AP* non è ancora completamente stabilizzato anche se utilizzato in diversi progetti. Attualmente il *DC-lib AP*, oltre alle qualificazioni (di cui specifiche: *descriptionURI, dateCopyrighted, dateSubmitted, dateAccepted, dateCaptured, dateISO8601*; viene poi accolto il qualificatore proposto dal gruppo di lavoro sulle citazioni bibliografiche, e cioè *identifierCitation*), prevede 18 elementi, cioè oltre ai 15 dello schema base:

- *audience*: elemento ormai acquisito nello schema base;
- *edition*: informazione che indica la versione o edizione di un'opera;
- *location*: elemento che identifica l'organizzazione che ha la risorsa o da cui si ottiene l'accesso. Questo elemento è ripreso da MODS (v. sotto) e ha uno schema di codifica che si richiama alla Lista dei codici delle organizzazioni adottato nel MARC.

Il settore delle biblioteche mostra un certo fermento, per i molteplici usi richiesti ai metadati e per il vasto dislivello di complessità che differenzia MARC21 dal semplice DC. L'“Ufficio per lo standard MARC e lo sviluppo della rete” della Library of Congress ha infatti emesso una bozza di metadati tratti da MARC21 che prende il nome di Metada-

³¹ DCMI Architecture Working Group, URL: <<http://dublincore.org/groups/architecture/>>.

ta Object Description Schema (MODS) ed è espresso in XML. La sua funzione principale è sia di trasportare dati selezionati del MARC21, sia creare *record* descrittivi originali. Rispetto al *DC-lib AP* presenta un set di elementi più ricco, anche se assai più semplice del MARC21 completo, con cui condivide pienamente la semantica: esso si presta quindi a dialogare in modo più puntuale con applicazioni in MARC ed è pensato per essere utilizzato anche come formato per Z39.50. In pratica la conversione MARC21-MODS riduce fortemente la perdita di specificità (notevole nella conversione in DC) e quindi di informazione (ma non è ugualmente completamente reversibile).

Abbiamo accennato al gruppo sugli agenti che ha avuto fin dall'inizio il compito di proporre qualificatori per i tre elementi del DC che registrano una qualche responsabilità sulla risorsa descritta, il *CCP* (*Creator, Contributor, Publisher*). A questo scopo è stata approvata la proposta del Comitato per l'applicazione, che per specificare i ruoli degli agenti prevede di fare riferimento alla lista di termini stabilita per i record MARC dalla Library of Congress designando con un codice di tre lettere la relazione tra un nome e una attività: si tratta di circa 185 termini della cosiddetta *relators list*³².

Inoltre è in corso di discussione la proposta dello stesso Comitato per l'applicazione di considerare il *dc.creator* come qualificatore del *dc.contributor*, anche se l'elemento *dc.creator* potrà continuare ad essere utilizzato come elemento di primo livello³³. In realtà la proposta tende a rendere obsoleto l'elemento *dc.creator* cioè, ci sembra, il criterio di responsabilità intellettuale principale, per sostituirlo con la designazione del ruolo svolto. La discussione è tuttora in corso.

In secondo luogo compito del gruppo è di stabilire un set di elementi per definire la forma del nome dell'agente, su cui esiste una proposta basata su 10 elementi, tra cui oltre al nome autorizzato o conosciuto, vi sono specifiche per lo "schema" del nome, con riferimento al sistema o alle regole in base alle quali il nome è formulato, l'agenzia responsabile, quale criterio di *authority control* (ad esempio AACR2; LC *names authorities*), il nome alternativo, il tipo di nome (*personal, corporate*), fino alle informazioni per contattare l'agente (con valori ispirati allo schema vCard MIME Directory Profile-RFC 2426)³⁴.

Il gruppo sui metadati gestionali era nato per definire un set essenziale di metadati di tipo gestionale da utilizzare in progetti di basso livello, detto *A-Core* o *Admin Core-Administrative Container Metadata*. In seguito a discussioni interne e con altre comunità di rete, il gruppo di lavoro ha rinunciato a definire un *core* set di elementi per la gestione delle risorse – di questo tema di occupa già il gruppo ISO TC46 / SC11 – per concentrarsi sulla gestione dei metadati, privilegiando l'interoperabilità tra differenti sistemi. Mentre la gestione degli oggetti digitali di tipo bibliotecario è affidata al *Metadata Encoding & Transmission Standard (METS)* messo a punto dalla Library of Congress, nessuno si era finora posto il problema del controllo dello scambio di *record* di metadati tra sistemi. Il gruppo si è quindi concentrato su quest'ultimo aspetto definendo un set di elementi per rispondere a tre tipologie di metadati: metadati per tutto il record (8 elementi), per l'aggiornamento e la modifica dei record (1 solo elemento, *activity*, per indicare l'azione condotta dall'entità responsabile sui metadati del record, un po' la storia delle azioni compiute su di esso, con 5 *refine-*

32 Cfr. <<http://www.loc.gov/marc/sourcecode/relators/relators.html>>.

33 Agents Working Group Report - DC 2002, Florence, <<http://dublincore.org/groups/agents/report-200210.html>>.

34 Draft DC Agent Element Set, <<http://www.nla.gov.au/meta/drafts/dcagent2.html>>.

ments) e per l'interscambio dei record in *batch* (7 elementi significativi per lo scambio di dati tramite file)³⁵. Il nuovo set è stato rinominato *AC-Administrative Components* e in ambiente XML dovrebbe dar luogo a un *namespace ac*.

Il gruppo sulle citazioni ha come scopo la definizione delle proprietà per il recupero di citazioni da risorse bibliografiche (ad esempio articoli da riviste *on line*) e a tal scopo sta discutendo una serie di possibilità tra cui l'introduzione dell'elemento *bibliographic.Citation* da considerare come qualificatore di *dc.identifier*. Un problema di rilievo è lo studio di raccomandazioni per la codifica delle citazioni bibliografiche in accordo con OpenURL il software elaborato in ambito OAI e in corso di approvazione come standard NISO, che permette di creare link *context sensitive* ai *full text* di risorse. Per gli stili delle citazioni è poi in corso di elaborazione una lista di risorse utilizzabili³⁶.

Il gruppo sulle collezioni si occupa di raccolte di oggetti fisici, o di surrogati di oggetti fisici, di oggetti nati come digitali e dei rispettivi cataloghi. Insomma rientrano in questa categoria collezioni museali, bibliotecarie, archivi, depositi digitali, *subject gateways*, indici Web, collezioni di immagini, suoni, set di dati, software ecc. L'importanza della descrizione delle collezioni è aumentata con lo sviluppo dei portali e dei servizi di biblioteca digitale, che richiedono l'accesso a più depositi e cataloghi di materiale informativo e progetti di *knowledge management* come ad esempio il progetto Dspace del MIT³⁷. Lo scopo del gruppo è infatti di definire i metadati per permettere il ritrovamento e la localizzazione di collezioni di diverse tipologie e discipline, e di condurre ricerche tra più collezioni in modo controllato. In assenza di cataloghi dettagliati come succede spesso negli archivi, la descrizione della collezione può inoltre offrire informazioni sui contenuti e la completezza della collezione. Gli standard di riferimento sono RSLP (Research Support Libraries Program) Collection Description Schema e EAD (Encoded Archival Description)³⁸.

Il gruppo *education* ha come obiettivo quello di sviluppare proposte per l'uso del DC nella descrizione delle risorse educative, ai diversi livelli, dall'asilo, alle elementari, medie, alta educazione, ma anche educazione nel corso della vita, educazione professionale e tecnica ecc. Il gruppo di lavoro ha proposto l'elemento *audience* (con relativi sotto elementi) e adesso sta elaborando un vocabolario appropriato per esprimerne i valori confrontando quelli usati in diversi progetti, in particolare il vocabolario di LOM di IEEE (*Primary Education, Secondary Education, Higher Education, University First Cycle, University Second Cycle, University Postgrade, Technical School First Cycle, Technical School Second Cycle, Professional Formation, Continuous Formation, Vocational Training*)³⁹. Si tratta cioè di elaborare qualificatori e vocabolari per i qualificatori per descrivere tali materiali al fine di favorirne il reperimento. Il prossimo impegno del gruppo dovrebbe essere quello di definire uno schema dello AP leggibile dalle macchine e compatibile coi metadati approvati dal Learning Technology

35 Cfr. <<http://dublincore.org/groups/admin/proposal-20021007.shtml>>.

36 Ann Apps, *Citation styles*, <<http://epub.mimas.ac.uk/DC/citstyles.html>>.

37 Cfr. <<http://www.dspace.org/>>.

38 Cfr. <<http://www.ukoln.ac.uk/metadata/rslp/schema/>; <<http://www.loc.gov/ead/tglib/index.html>>.

39 IEEE 1484 Learning Objects Metadata (IEEE LOM): *IEEE 1484 LOM mappings to Dublin Core: Learning Object Metadata: Draft Document v3.6, IEEE Learning Technology Standards Committee (LTSC), 5 September 1999.

Mappings by Stuart A. Sutton, <<http://ltsc.ieee.org/doc/wg12/LOM3.6.html>>.

Standards Committee (LTSC) della autorevole IEEE (IEEE.LOM).

Il gruppo *e-government* si occupa della applicazione del DC per acquisire informazioni sui servizi e i documenti delle agenzie governative, ovvero, si direbbe in Italia, dell'amministrazione pubblica, e mira pertanto a identificare elementi comuni nelle applicazioni in corso e a promuovere l'uso del DC in questa comunità. Il primo schema di questo tipo è stato certamente AGIS (*Australian Government Locator Service*), basato sul DC con 4 elementi addizionali. Analoghi schemi, basati o largamente compatibili col DC sono stati sviluppati in Canada, in Irlanda e in alcuni stati degli USA oltre che in UK e Danimarca⁴⁰. Il gruppo di lavoro del DCMI propone per il profilo applicativo dell'informazione pubblica un nuovo elemento, il *dc.audience*, ripreso dal profilo elaborato dal gruppo di lavoro per il settore education, e una serie di nuovi qualificatori dell'elemento *dc.date*, *dc.relation* (“*isBasedOn*” e “*isBasisFor*”) e soprattutto del *dc.rights*, con l'introduzione dei qualificatori “*securityClassification*”, “*previousSecurityClassification*”, “*accessRight*”, “*copyright*”. Infine il qualificatore del *dc.type* “*aggregationLevel*”⁴¹. Ancora a livello di bozza è invece l'identificazione delle caratteristiche per la descrizione dei servizi forniti dalle agenzie governative⁴².

Recentemente l'attenzione del gruppo di lavoro si è spostata soprattutto verso i metadati per il mantenimento dei record (*recordkeeping metadata*), settore che comprende sia metadati per la gestione dei record, in modo di mantenere completa, accurata e affidabile evidenza alle transazioni basate sulle informazioni archiviate, che nel senso di gestione degli archivi. Alla funzionalità principale della “scoperta” (*discovery*) dell'informazione se ne aggiungono altre. *Recordkeeping metadata* è “informazione strutturata e semistrutturata che permette la creazione, la gestione e l'uso di record nel tempo, all'interno e tra i domini in cui è creata. Essa può essere usata per identificare, autenticare e contestualizzare i record ed anche le persone, processi e sistemi che creano, gestiscono e usano questi record”⁴³. Tale studio è quindi cruciale per la gestione sia dei record correnti che degli archivi.

Gli argomenti di cui si occupano gli altri gruppi di lavoro o di interesse sono altrettanto importanti (utili i software elencati dal gruppo *tools*⁴⁴) ma la loro attività o è chiaramente intuibile o è meno stabilizzata sia come dibattito che come risultati.

Il quadro proposto è dunque abbastanza ricco e articolato e tale da comportare per tutti coloro che si interessano alla gestione dell'informazione – e quindi anche le biblioteche – la necessità di relazionarsi in modo non episodico e saltuario con questa realtà in movimento. Questa accentuata attenzione dovrebbe anzi inserirsi in una attività sistematica di analisi e supporto alle iniziative di costruzione della biblioteca digitale e dei sistemi informativi, in modo da garantire comunicazione,

⁴⁰ Una tabella con le applicazioni del DC per lo *e-government* nel mondo si trova in: Max Dekkers – Palle Aagaard – Andrew Wilson, *Deliverable: survey of government implementation of DC*, <<http://dublincore.org/groups/government/survey-200106.shtml>>.

⁴¹ *DC-Gov application profile*, <<http://dublincore.org/groups/government/profile-200111.shtml>>.

⁴² Una prima analisi dei requisiti di un simile schema si trova in: John Roberts, *Functional requirements for describing services. Discussion paper for DC-Gov*, <<http://dublincore.org/groups/government/service-requirements.html>>.

⁴³ John Roberts – Andrew Wilson, *International recordkeeping/records management metadata initiatives: report and recommendations for DC Advisory Board*, <<http://dublincore.org/groups/government/DCMIResourceManagement.html>>.

⁴⁴ Cfr. <<http://dublincore.org/tools/>>.

interoperabilità, scambio di esperienze, collaborazione ecc., non solo in Italia ma in un contesto europeo e internazionale. È utile ricordare che un ruolo chiave in questo processo vede protagonisti sistemi bibliotecari efficienti e disposti a investire sulle nuove tecnologie con lo sguardo rivolto anche oltre ai confini della propria organizzazione e in modo autonomo sia da eccessivi centralismi che da posizioni di predominio sul mercato. È poi necessario avviare una cooperazione su larga scala tra istituzioni e aziende per poter coordinare progetti di largo respiro entro e tra comunità di interessi (nel senso delineato in queste pagine) incrementando l'adesione agli standard internazionali, diffondendo linee guida e *roadmaps* per la implementazione di sistemi informativi, organizzando l'attività di diffusione e trasferimento dei risultati della ricerca e delle esperienze internazionali.

The importance of metadata in the construction of the global network

by Nicola Benvenuti

This text has been written further to the International Conference of the Dublin Core Management Initiative (DCMI) 2002, held in Florence (Italy) in October 2002 to inform Italian librarians about scope, results and applications of the most common metadata scheme. The article reports on the main themes tackled in the various sections of the Conference such as keynote speeches (speakers: De Sompel about OAI and OpenURL, E. Miller, about the Semantic Web, S. Weibel, about DCMI), tutorials and the proceedings of the conference.

An important issue concerned the function of Dublin Core, frequently considered as alternative to the traditional MARC cataloguing and not as a format to meet the problem of interoperability between traditional book collections and new types and formats of information objects.

The article retraces the debate on the function of metadata and outlines the project of Semantic Web supported by W3C. The function of Extensible Markup Language (XML) as meta-language for creating user-oriented metadata and the role of "namespace" as formal declaration on the net of metadata schemes is described, as well as the proposed way to establish semantic relationships between resources and to express ontologies through the Resource Description Framework (RDF).

The role of the DC as a language to achieve interoperability between metadata formats is presented together with DC elements, their vocabularies and refinement rules. Reference is made to the different structure and function of the existing metadata schemes from the automatic metadata created by search engines through DC up to complex schemes like TEI, EAD or MARC21.

The activity of DCMI to maintain the position of the DC in the professional information community as the most used scheme is illustrated, characterising his organisation as a light, open and collaborative environment. The article addresses also the adaptation strategy of the DCMI to make different languages and metadata schemes interoperable, focusing on the role of application profiles as a means to meet different user requirements in changing frameworks.

The work of the interest groups within DCMI is also examined: the Registry group intended for promoting the public registration of semantic and the reuse of elements from different domains; the Architecture group that licensed prescriptions for expressing DC in HTML, XML, RDF; the Library group that created the library profile for multiple use, as well as the Agent group that is actually working on the refinement

NICOLA BENVENUTI, Università di Firenze, Biblioteche di scienze tecnologiche, via Micheli 2, 50121 Firenze, e-mail benvenuti_n@unifi.it.

of the creator element. The Administrative metadata group concentrates on the core set for the administrative metadata, while the Collection group deals with collection level description, a main task for digital libraries and knowledge organisations. Brief reports on the activity of the Education group, E-governments group, Citation group etc. are given.

The article on the whole focuses on the opportunity to operate in the open, collaborative environment facilitated by DCMI for extending the role and function of the libraries and building efficient interoperable information systems.