

“Metadata boys”

Nell'agosto 1998 si tenne il XVI Colloquium on Library Science Research, “Internet, metadata and information access to libraries and networks in the electronic age”, organizzato al University Center for Library Science Research della National Autonomous University of Mexico. Michael Gorman, direttore dei servizi bibliotecari della California State University, Fresno, California, presentò in quell'occasione la relazione *Metadati o catalogazione? Una falsa alternativa*, edita negli atti del convegno e nel “Journal of Internet cataloging”, 2 (1999), 1, p. 5-22, fascicolo che raccoglie contributi di altri relatori al Colloquium.

Il saggio di Gorman, che riproponiamo volentieri ai lettori di “Biblioteche oggi”, costituisce un'importante riflessione sull'utilizzo dei metadata e sulla loro efficacia nel ricupero delle risorse elettroniche (ER) disponibili in rete; significativa è la correlazione fra la problematica del controllo bibliografico delle ER remote e la pianificazione del processo della loro selezione. Ciò evidenzia il problema della valutazione dei documenti che meritano di essere conservati o resi disponibili nelle biblioteche, che riguarda la loro durata, la loro stabilità, la loro universalità e altro. La classificazione delle risorse disponibili in relazione al loro valore condiziona la scelta del metodo di catalogazione economicamente più conveniente. Gorman è consapevole di essere prossimi a un effettivo controllo bibliografico grazie all'estesa adozione di protocolli e di standard comuni (ad esempio, MARC, Z39.50), ma dichiara la superiorità della catalogazione completa e dei record Dublin Core arricchiti rispetto alla prospettiva minimalista, ad esempio, uso dei 15 elementi con contenuti non normalizzati. Afferma, inoltre, che è compito di ogni istituto mantenere queste due forme di catalogazione per le risorse che intende conservare. Introduce il concetto di conoscenza registrata che si oppone al concetto di dato o informazione, nozione che permette di focalizzare il significato delle raccolte in quanto insiemi, in quanto sistemi di documenti correlati fra loro tramite l'architettura del controllo bibliografico. Gorman ha difeso l'importanza della catalogazione anche in altri interventi recenti; in From

card catalogues to WebPACS: celebrating cataloguing in the 20th century, relazione presentata alla Bicentennial Conference on Bibliographic Control promossa dalla Library of Congress il 15 novembre 2000 (cfr. <http://lcweb.loc.gov/catdir/bibcontrol/gorman_paper.html>), Gorman commenta con fine umorismo le posizioni di alcuni documentalisti e informatici che prefiguravano un mondo senza storia in cui i documenti digitali “trascendenti e trasformativi [...] misteriosamente, catalogheranno se stessi” e appella sarcasticamente “metadata boys” coloro che confon-

dono la funzione del formato MARC con le regole di catalogazione. La comunità bibliotecaria ha oggi la possibilità di “incorporare i documenti elettronici nel controllo bibliografico universale, come ha incorporato tutte le altre forme di comunicazione umana, inaugurando un'altra età dell'oro della catalogazione che rinforzi il nostro unico impegno in quanto bibliotecari, quello di conservare e di trasmettere ai posteri le registrazioni dell'uomo”.

La traduzione cerca di mantenere il tono discorsivo del testo originale.

(A cura di Stefano Gambari e Mauro Guerrini)

Metadati o catalogazione?

Una falsa alternativa

di Michael Gorman

Introduzione

Le biblioteche sono istituzioni che selezionano, acquisiscono, danno accesso, ordinano e conservano l'informazione e la conoscenza registrata in qualsiasi formato, fornendo assistenza e istruzione all'uso di tale informazione e conoscenza. Nel perseguire questa missione, ogni biblioteca dispone di tre inestimabili risorse: uno staff addestrato e ben informato di bibliotecari e di altri operatori di biblioteca, le raccolte e l'architettura del controllo bibliografico che con-

sente l'accesso a tali raccolte. Una biblioteca che manchi di una di queste tre caratteristiche non è una biblioteca. Senza controllo bibliografico, i bibliotecari e gli altri membri dello staff non possono fare il loro lavoro e gli utenti della biblioteca non possono usare con efficacia le raccolte. Senza i bibliotecari e gli altri membri dello staff, la biblioteca è semplicemente un deposito. Senza raccolte, una biblioteca si riduce al livello di un ufficio informazione. Si dovrebbero vedere questi tre tratti essenziali come congiunti e interdipendenti,

ciascuno prezioso e necessario al pari degli altri. Come dice il vecchio indovinello: qual è la gamba più importante di uno sgabello a tre gambe? Tratterò brevemente di ciascuno di questi fattori. In primo luogo i ...

Bibliotecari

In molte nazioni oggi, il termine "bibliotecario" è applicato a persone con un titolo (spesso un titolo avanzato) nelle discipline biblioteconomiche, sebbene sia spesso usato per designare amministratori di biblioteche più anziani, senza considerare se posseggano o meno tale titolo. In ogni caso, il lavoro di bibliotecario consiste nell'applicare la conoscenza professionale e l'esperienza a una o più componenti la missione della biblioteca. Queste componenti sono in genere chiamate sviluppo delle raccolte, controllo bibliografico, reference, istruzioni all'uso della biblioteca e amministrazione. Ognuno di questi ambiti di lavoro può essere applicato alla biblioteca come a un tutto o a raccolte e servizi specializzati al suo interno. La natura del lavoro del bibliotecario è mutata negli ultimi vent'anni. Questo cambiamento è stato in parte prodotto dalla tecnologia e, in parte, dalla concentrazione della professionalità e dalla conseguente crescita come numero e importanza del personale para-professionale nelle biblioteche. In breve, abbiamo meno bibliotecari, ma il loro lavoro è più professionale di prima. La successiva, essenziale, caratteristica di una biblioteca sono le sue ...

Raccolte

È mutato parimenti il concetto di raccolta di biblioteca. Non troppo tempo fa, una raccolta di biblioteca poteva essere definita come i

materiali (quasi tutti testi a stampa, su carta) conservati all'interno dell'edificio o degli edifici della biblioteca. Al giorno d'oggi dobbiamo ammettere una concezione molto più estesa delle raccolte di biblioteca. La tecnologia ha portato nuove forme di conoscenza registrata e di informazione e ci ha messo in grado di condividere l'informazione sulle risorse disponibili e, in definitiva, di condividere le risorse stesse.

La raccolta di una biblioteca moderna consiste di quattro parti:

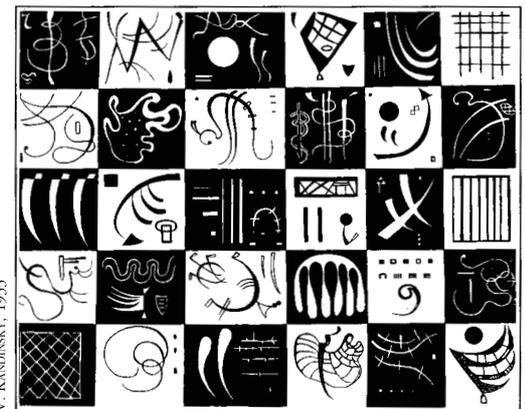
- oggetti tangibili (libri, registrazioni, mappe, video, ecc.) posseduti dalla biblioteca e conservati in una determinata localizzazione fisica;
- risorse (elettroniche) intangibili¹ (cd-rom, ecc.) possedute dalla biblioteca e disponibili per l'uso in biblioteca;
- oggetti tangibili (libri, ecc.) posseduti da altre biblioteche ma disponibili per la biblioteca attraverso sistemi interbibliotecari di condivisione delle risorse e cataloghi collettivi;
- risorse (elettroniche) remote, intangibili non possedute dalla biblioteca ma alle quali essa fornisce accesso.

Quando facciamo nostra quest'ampia nozione, possiamo vedere la biblioteca come il centro di una serie di cerchi concentrici più estesi che alla fine comprendono la conoscenza registrata e l'informazione del mondo intero. In tale rappresentazione, la propria biblioteca è sia una risorsa locale, sia parte di una vasta rete di basi dati e di biblioteche intercollegate. Prima di essere portati altrove dalla grandezza di questa metafora, dovremmo ricordare che essa è per l'appunto una metafora e che la grande biblioteca universale consiste, in realtà, in un *patchwork* di un numero quasi infinito di programmi, servizi, schemi ecc. individuali, tutti guidati da interessi locali. Dovremmo anche ricordare che

gli utenti delle biblioteche preferiscono raccolte locali e basi dati disponibili localmente perché più verosimilmente producono ciò che l'utente richiede, un accesso rapido alle risorse di cui ella o egli ha bisogno.

Un altro aspetto importante della visione allargata delle raccolte di biblioteca è la necessità di operare distinzioni tra conoscenza registrata da una parte e informazione dall'altra. Quando è usata in espressioni quali "era dell'informazione" e "professionista dell'informazione" (un sinonimo moderno, impreciso, di bibliotecario), la parola "informazione" è quasi priva di significato. La mancanza di senso conduce all'incapacità di distinguere, in termini di importanza o valore, tra le poesie di Garcia Lorca e un elenco telefonico. Le parole hanno significati e quei significati portano un peso, quindi è importante definire i termini fondamentali su cui poggiano tutte le nostre speculazioni sul futuro.

È vitale per noi comprendere la distinzione tra *informazione* da una lato e *conoscenza registrata* dall'altro. L'informazione consiste di dati, pacchetti discreti di testo e immagini. La caratteristica più importante dell'informazione è che può essere usata fuori dal contesto. Possiamo prendere il numero della popolazione che vive in America centrale, un record da un catalogo di biblioteca, una breve descri- ➤



zione della geografia dell'Honduras che possiamo trovare in un almanacco o in altra opera di pronta consultazione, o una fotografia della Casa Bianca e usarle, senza rapporto a qualsiasi altro fatto o informazione, per aggiungerle a un deposito di informazioni. La conoscenza registrata, d'altra parte, è discorsiva e cumulativa. Proceede da un punto a un altro, nel modo in cui è stata formata dall'autore o creatore del testo o della sequenza visiva. In altre parole, una pagina tratta da un libro scolastico o un'immagine da un documentario non ha un significato indipendente; il suo uso dipende da ciò che precede e che segue (non sto parlando qui della scrittura creativa o delle opere d'arte visive, sebbene anch'esse condividano la natura discorsiva, modellata, della conoscenza registrata e debbano essere lette o viste come un intero e nel contesto). Il punto che sto cercando di fare non è meramente semantico. Il fatto è che la tecnologia elettronica è straordinariamente valida per elaborare l'informazione e inadatta a trattare la conoscenza registrata. Inoltre, la tecnologia elettronica fornisce poche risposte a uno dei più importanti aspetti della biblioteconomia, la conservazione e trasmissione alle future generazioni delle testimonianze della specie umana. Quando discutiamo delle raccolte, inoltre, dobbiamo tener presente che alcuni mezzi di comunicazione hanno punti di forza di cui altri difettano e mostrano debolezze che richiedono altri media per essere perfezionate. Il futuro delle biblioteche dipende dalla capacità di capire la distinzione tra informazione e conoscenza registrata, perché ciascuna delle tre componenti principali di cui abbiamo parlato verrebbe ad essere completamente stravolta se questa distinzione e le sue conseguenze non fossero comprese. Se assumiamo, come qualcuno ha fatto, che il

futuro si basa sui mezzi digitali e solo su di essi, per diffondere e conservare sia la conoscenza registrata che l'informazione, i bibliotecari diverrebbero dei chierici elettronici, la nozione estesa delle raccolte di cui abbiamo parlato si ridurrebbe a successioni di basi dati e il controllo bibliografico diventerebbe solo materia di motori di ricerca, ricerca per parole² e meta-dati. Parliamo ora del...

Controllo bibliografico

Il controllo bibliografico è la linfa vitale della biblioteca. Nel corso dei secoli, i bibliotecari hanno creato un'elegante e tuttora funzionale architettura di catalogazione, classificazione, soggettazione e, nelle decadi recenti, dei sistemi automatizzati che hanno notevolmente accresciuto la funzionalità di quell'architettura. La pressione verso la standardizzazione del controllo bibliografico è stata accelerata dalla tecnologia. Ora, per la prima volta nella storia delle biblioteche, siamo a un punto in cui il sogno del Controllo bibliografico universale (UBC) è realizzabile nel mondo reale. Il mondo di lingua inglese è dominato dalle *Anglo-American cataloguing rules, second edition*, (AACR2) e quel codice si è rivelato assai prestigioso nell'influenzare i codici di catalogazione in molti paesi non di lingua inglese. Tutte le nazioni hanno adottato l'International standard bibliographic description (ISBD). Migliaia di biblioteche nel mondo usano la Classificazione decimale Dewey (DDC) o la Classificazione della Library of Congress o la Classificazione decimale universale

(CDU). I paesi di lingua inglese usano la *List of subject headings* della Library of Congress o qualche sua variante. Il sistema MARC di codifica dei record in formato elettronico è basato su uno standard dell'International Standards Organization (ISO). È usato per scambiare record tramite UNIMARC, universalmente accettato, e quasi ogni paese, regione o gruppo linguistico ha la propria versione di MARC oppure utilizza il formato USMARC per mantenere e trasmettere record. Il MARC è la base per quasi tutti i cataloghi di biblioteca in linea e il record MARC ha reso possibile la costituzione di immense e aggiornate basi dati nazionali e internazionali che sono necessarie per poter raggiungere il controllo bibliografico universale. In breve, si sta delineando un accordo a livello internazionale sulla struttura e il contenuto dei record. Abbiamo, o potremmo avere, gli strumenti per creare un nuovo ordine bibliografico mondiale.

Catalogazione

Catalogare è un'attività centrale nel campo del controllo bibliografico. Consiste di un numero di passi complessi, il primo dei quali è la descrizione – in accordo con le ISBD – dell'entità bibliografica che si sta catalogando. Tale entità bibliografica può essere un oggetto tangibile o un insieme di oggetti (un libro, un microfilm, una serie di videocassette in contenitore, una raccolta di manoscritti) oppure un assemblaggio di dati elettronici che troviamo in un oggetto tangibile localmente posseduto (cd-rom, nastro



elettronico, ecc.) o ancora una risorsa remota contenuta in una base dati distante. Una volta descritta l'entità bibliografica, si guarda l'opera di cui quell'entità è una manifestazione e si assegnano punti d'accesso a nomi e a titoli uniformi, intestazioni di soggetto e numeri di classificazione in accordo con regole standard e tavole di classificazione. Infine, si prendono tutti i dati bibliografici relativi all'entità bibliografica e all'opera per codificarli secondo la versione di MARC usata nella propria biblioteca, sistema o paese. Tutto questo processo catalografico è svolto con riferimento non solo alle regole e agli standard che ho già menzionato ma conformemente alla pratica del passato quale è registrata negli authority file (che contengono i nomi e i titoli standard di persone, enti e opere) e nei manuali di applicazione delle intestazioni di soggetto e dei numeri di classificazione.

Da questa descrizione risulta chiaro che la catalogazione richiede per sua natura un lavoro intensivo e professionale. Si tratta, quindi, di un processo costoso. La cosa evidente è che si può impiegare il denaro all'inizio del processo, in modo tale che migliaia di futuri utenti della biblioteca possano trovare ciò che cercano rapidamente e con un accettabile grado di precisione e richiamo, oppure, come nel caso di Internet e del web oggi, il denaro sarà speso nel tempo sprecato da quelle migliaia di utenti che si dibatteranno con scarsi risultati in un caos privo di controllo con il solo fallibile aiuto della ricerca per parole. Si confronti per l'appunto l'esperienza di localizzare la conoscenza e l'informazione desiderata in una biblioteca ben ordinata con l'assordante rumore prodotto da una semplice ricerca nel web.

Per esempio, ho inserito la parola "Honduras" come termine di soggetto nel catalogo in linea della mia biblioteca e ho ottenuto come

risultato un elenco di intestazioni di soggetto ordinate alfabeticamente, ognuna accompagnata da un'indicazione del numero dei documenti che erano presenti nel catalogo su quello specifico soggetto. Ecco le prime voci:

- Honduras (4)
- Honduras – Administrative and political divisions – Maps (3)
- Honduras – Antiquities (7)
- [See Also Narrower Term] Cerro Palenque Site (Honduras) (1)
- [See Also Narrower Term] Copán Site (Honduras) (10)
- Honduras – Armed Forces – Political activity (1)
- Honduras – Armed Forces – Political activity – 20th century (1)
- Honduras – Constitutional history (1)
- Honduras – Description and travel (5)
- Honduras – Discovery and exploration – Maps (1)
- Honduras – Economic conditions (3)
- Honduras – Economic conditions – 1918- (3)
- Honduras – Economic conditions – 1918-1980 (1)
- Honduras – Economic conditions – 1918 – Maps (1)
- Honduras – Economic policy (1)

Ho scelto quindi una di queste voci "Honduras – Antiquities" e ho trovato una lista breve di sette titoli, fra cui:

- Pottery of prehistoric Honduras : regional classification and analysis / Los Angeles : Institute of Archaeology, University of California, Los Angeles, 1993.
- 312 p. : ill. ; 28 cm.
- Monograph (University of California, Los Angeles. Institute of Archaeology) ; 35.
- Subjects
- Indian pottery – Honduras.
- Indians of Central America – Honduras – Antiquities.

- Pottery, Prehistoric – Honduras – Analysis.
- Pottery, Prehistoric – Honduras – Classification.
- Honduras – Antiquities.
- CSUF STACK F1505.3.P6 P68 1993

[Si notino le ulteriori voci di soggetto, la descrizione ISBD, i numeri di classificazione, tutti elementi che richiedono la conoscenza di regole di catalogazione, MARC, ecc.]

Successivamente ho cercato la stessa parola ("Honduras") su AltaVista, uno dei più popolari motori di ricerca. Il risultato è stato: 266.970 pagine che soddisfano il criterio di ricerca, tra cui queste quattro voci considerate come "le più rilevanti":

- 1) 31.07.95.BID Y HONDURAS FIRMAN CONTRATOS POR US\$79 MILLONES
[URL:www.iadb.org/prensa/cp15195c.html]
31.07.95.BID Y HONDURAS FIRMAN CONTRATOS POR US\$79 MILLONES. Samuel Silva: (202) 623-3731 Correo electrónico: samuels@iadb.org CP-151/95 PARA ...
- 2) Dive in Honduras – Main Page
[URL: www.diveinhonduras.com/] nbsp; Why us? The Resorts. Islands info. Online Request. Feedback. Mainland Tours. You are Diver number. to dive deeply into our Web site Since...
Last modified 6-Feb-98 – page size 9K – in English [Translate]
- 3) A Special Advertising Section on Honduras
[URL:www.washtimes.com/international/honduras/coffee.html]
A Special Advertising Section on Honduras. A medium size producer will net about \$80.000 this year. Honduran coffee industry is filling up its cup. The... ➤



Last modified 15-Aug-97 – page size 11K – in English [Translate]

4) Central America Business Directory: Guatemala, El Salvador, Honduras, Nicaragua
[URL:www.centramerica.com]
Business Tourism Exporters Importers Services Banking Directory: Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica

AltaVista, e motori simili, usano una ricerca per parole a pieno testo sulle risorse elettroniche oggetto della ricerca. Per definizione, le parole non sono parte di un vocabolario controllato come quello prodotto da codici di catalogazione, repertori di voci di soggetto e classificazioni. Il risultato è il “rumore” di cui siamo stati per l'appunto testimoni. L'esito di una catalogazione accurata (e costosa) è che l'utente viene condotto molto rapidamente al documento desiderato e ha fiducia nel fatto che la ricerca abbia restituito una buona proporzione dei materiali rilevanti posseduti dalla biblioteca. Una ricerca in MELVYL, il gigantesco catalogo collettivo di tutte le biblioteche della University of California,

con il soggetto “Honduras” ha recuperato molti più record (825) di quelli rintracciati nel catalogo della mia biblioteca. In ogni caso, quei record erano resi più utilizzabili dall'essere presentati in ordine alfabetico per autore e per la possibilità di restringere facilmente la ricerca con l'aggiunta di termini di soggetto più minuziosi.

Quando ho aggiunto il termine “Antiquities” alla ricerca su MELVYL, mi furono restituiti 50 record, di nuovo presentati in ordine alfabetico.

Cercando sul web con AltaVista e aggiungendo la parola “Antiquities” a “Honduras”, il risultato fu di limitare i documenti recuperati (da oltre 250.000 a “soli” 86.030), presentati di nuovo in apparente ordine casuale. La disposizione è, naturalmente, basata sul conteggio delle parole, una regola che ha prodotto, come quinto più rilevante documento degli 86.000:

5) POPYRUS BOOKS, books on Ancient Art, Antiquities, Archaeology & Numismatics
[URL:www.papyrusbooks.com/] new, rare and out of print books on the Arts, Archaeology, and Numismatics from Ancient Greece, Rome, Egypt, Near East, Byzantium

È ovvio, in termini logici e anche per esperienza, che la catalogazione completa delle risorse elettroniche basata sugli standard e sui vocabolari controllati è infinitamente preferibile al recupero e alla visualizzazione basati sulla ricerca per parole a pieno testo, con vocabolari non controllati e conteggio delle parole. Rimangono, in ogni caso, alcuni problemi.

Possiamo sostenere una catalogazione completa?

Esiste una terza via fra la catalogazione completa e la palude della ricerca per parole?

Gli standard di catalogazione esistenti possono essere applicati alle risorse elettroniche?

Quali risorse elettroniche sono suf-

ficientemente utili per meritare di essere catalogate?

Come possiamo conservare le risorse elettroniche una volta catalogate?

Cosa catalogare? (non: come catalogare?)

Prima di intraprendere una disamina dei metadati e di come essi possano o meno rientrare nel controllo bibliografico universale, consideriamo la natura dei documenti elettronici disponibili su Internet e sul web. Come sa chiunque abbia cercato di districarsi tra i risultati di un certo numero di ricerche nel web, gran parte dei documenti che si ottengono da tali ricerche sono inutili, d'uso ridotto o estremamente limitato nel tempo e nello spazio (ad esempio, quelli che sono effimeri o d'interesse estremamente locale o entrambe le cose). Non ho alcun mezzo per stabilire la percentuale delle risorse elettroniche di valore generale e duraturo, ma sarei molto sorpreso se fossero più dell'uno o due per cento. Ciò richiama all'attenzione la domanda di un approccio a due livelli nel recupero delle risorse elettroniche: catalogazione completa per una minuscola minoranza e recupero migliorato ma ancora poco costoso per la gran maggioranza? Esistono criteri per rendere più efficiente la ricerca per parole anche per ricercatori privi delle necessarie abilità. Ad ogni modo, il punto fondamentale è che le parole hanno i seguenti, intrinseci inconvenienti:

- non sono, per definizione, estratte da un vocabolario controllato;
- sono prive di contesto: ad esempio, una ricerca con “Honduras” può portare a ottenere un documento che contiene la frase “questo studio non attiene all'Honduras”, proprio con la stessa facilità con cui recupera un documento relativo, in effetti, all'Honduras;



G. CARROZZINI, 1959

– non hanno relazioni con altre parole: ciò è in netto contrasto con gli intricati collegamenti che esistono in un elenco di voci di soggetto e con le capacità che permettono all'utente di spostarsi da un soggetto ad altri soggetti correlati;

– al di fuori delle scienze tradizionalmente consolidate, la terminologia è ben lungi dall'essere precisa e gli studenti e gli scrittori usano le parole con moltissime sfumature e mancanza di accuratezza.

Fintanto che non possediamo macchine effettivamente in grado di pensare e “leggere” testi nello stesso modo di un sofisticato essere umano,³ nessun incremento nella programmazione renderà la ricerca per parole così proficua per gli utenti come la catalogazione biblioteconomica “tradizionale”. D'altra parte, se siamo capaci di identificare la piccola minoranza di risorse elettroniche di valore permanente, forse una ricerca per parole un po' più avanzata di quella attuale sarà sufficiente per la massa dei documenti elettronici effimeri di valore limitato e locale.

Standard (struttura e contenuto)

Abbiamo in precedenza menzionato gli standard che governano la catalogazione. È molto importante per la nostra discussione capire la differenza tra gli standard *strutturali* e quelli che prescrivono il *contenuto* del record. MARC è il primo esempio di uno standard strutturale. Assegna nomi ai campi e sottocampi del record leggibile da una macchina, senza specificare i dati che devono essere inseriti in quei campi e sottocampi. Per esempio, il campo 100 è per le intestazioni principali al nome personale e il campo 700 per i campi per le intestazioni personali aggiunte (la distinzione tra i due è un residuo della catalogazione dei

tempi precedenti l'uso del computer, ed è stata trasportata in modo bizzarro nei sistemi di metadati, ma questa è un'altra storia ...). MARC non dice al catalogatore quali nomi dovrebbero andare nel campo 100 e quali nel 700, né la forma dei nomi che deve usare. Per questo bisogna rivolgersi a uno standard di contenuto, come AACR2. Le ISBD sono essenzialmente standard strutturali; sebbene prescrivano il contenuto in termini generali, si deve ricorrere al proprio codice di catalogazione nazionale per trovare la forma precisa in cui il dato deve essere inserito nella struttura ISBD. Schemi di classificazione e liste di voci di soggetto sono naturalmente standard di contenuto. I primi periodi di sviluppo dei metadati (ad esempio, quelli di cui siamo testimoni) sembrano essere caratterizzati dalla confusione per l'incomprensione – da parte degli autori di metadati – della distinzione tra standard strutturali e standard di contenuto e dalla conseguente incapacità di decidere se i metadati hanno a che fare con il contenuto e da che cosa dovrebbe discendere quel contenuto.

Possono essere applicati alle risorse elettroniche gli standard di catalogazione attuali?

I metadati, ovviamente, sono destinati a rispondere a bisogni percepiti. Dal mio intervento risulta che quei bisogni consistono:

– nell'assicurare un modo per la catalogazione delle risorse elettroniche;

– nel fornire una terza via alternativa al costo della catalogazione tradizionale e alla futilità della ricerca per parola.

È del tutto possibile, dato che il concetto originale di “metadati” non ha avuto origine tra i bibliote-

cari, che non sia stato considerato l'uso della catalogazione “tradizionale” e che, seppure i bibliotecari siano ora coinvolti nei progetti, ci si trincerino ancora fermamente dietro l'idea che le risorse elettroniche non possano essere catalogate con l'uso degli standard tradizionali. Sia quel che sia, il fatto è che le entità bibliografiche elettroniche hanno gli stessi attributi delle altre entità bibliografiche. È perfettamente possibile catalogare le risorse elettroniche in modo tale che i relativi record possano essere pienamente integrati nei cataloghi di biblioteca. Esiste un nuovo ISBD per le risorse elettroniche⁴ che costituirà la base della revisione del cap. 9 delle AACR2R; le risorse elettroniche hanno titoli e creatori (autori) che possono essere usati per garantire punti standard di accesso, possiedono soggetti esprimibili in numeri di classificazione e in intestazioni di soggetto, e tutti questi dati possono essere incorporati in un record MARC. In breve, se una delle giustificazioni per l'invenzione dei metadati fosse che essa è resa necessaria per facilitare l'accesso alle risorse elettroniche in assenza di standard di catalogazione, ebbene quella giustificazione è semplicemente erranea.

Metadata e MARC

Forse la decisione è stata presa senza pensarci a fondo. Quella decisione sembra essere: poiché la “catalogazione tradizionale” è troppo costosa, deve esistere un compromesso – una qualche terza via – che ci darà i benefici della catalogazione senza lo sforzo e la spesa che essa comporta. Si legge nelle parole dell'*Introduzione* al rapporto finale del Nordic Metadata Project:⁵

Molti specialisti credono che una qualsiasi forma di metadati sia preferibile

all'assenza dei metadati: non abbiamo bisogno di rimanere bloccati dai *rigorosi requisiti di qualità e dai formati complessi dei sistemi di catalogazione bibliotecari*. Invece, è possibile vivere con qualcosa di semplice, che sarà facilmente comprensibile a editori, autori e altre figure coinvolte nella pubblicazione dei documenti elettronici (*il corsivo è mio*).

È uno dei pochi riferimenti in questo lungo resoconto delle necessità esistenti e della natura dei metadati come alternativa alla catalogazione. È preso per buono che esista qualcosa tra i "rigorosi requisiti di qualità" e "l'assenza completa di qualità", e che ci sia qualcosa tra i "formati complessi" e l'assenza di qualsiasi formato. La questione della catalogazione è introdotta più avanti nel rapporto in un modo per certi versi bizzarro. Il comitato sta discutendo cambiamenti al Dublin Core influenzati dal Nordic Project:

Su un piano di dettaglio, il Nordic Metadata Project ha avuto una forte influenza sul modo in cui i metadati Dublin Core vengono attualmente forniti. Per esempio, si decise che *la forma preferita per i nomi di autore è quella normalizzata* ("cognome, nome"). Una ragione di questa decisione era il fatto che il nostro progetto sosteneva fortemente questa concezione. Ponemmo in rilievo nella discussione che questa forma del nome è superiore per l'ordinamento degli indici e la conversione da Dublin Core a formati quali MARC, che richiedono forme normalizzate dei nomi (*il corsivo è mio*).

Si noti l'assunzione che la sola "normalizzazione" che i nomi richiedono per funzionare come efficaci mezzi di ricupero è l'inversione (cognome, nome) che è la norma per molti nomi di origine europea. In effetti, come qualsiasi catalogatore può riferirmi, nomi non occidentali non sono spesso invertiti ed esistono numerose varianti di tutti i tipi di nome per i

quali le regole di catalogazione sono necessarie per scegliere la forma normalizzata. Questo concetto ricorre più avanti nel testo della relazione:

Al fine di mettere in grado gli utenti di preparare record in formato Dublin Core, *un aspetto essenziale è di gran lunga carente, ossia le regole di catalogazione. Non esiste alcun testo autorevole* che affermi da quale punto della pubblicazione sia permesso andare a prendere il nome dell'autore, o che definisca come sistemare il nome nell'etichetta Creatore ("cognome, nome" o "nome cognome"). Questo tipo di informazione è generalmente fornito dalle guide per l'utente specifiche di ogni progetto (*il corsivo è mio*).

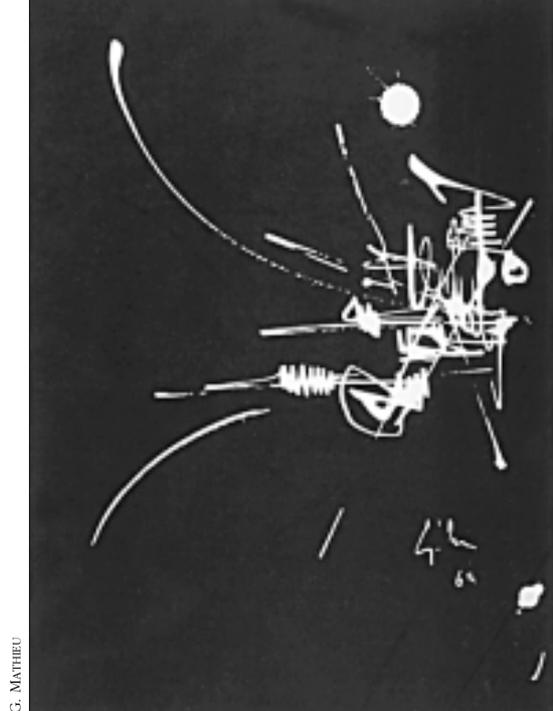
Di nuovo osserviamo l'idea che le regole di catalogazione riguardano solo il nome dell'autore e se esso sia invertito o meno. Non è fatto alcun cenno agli innumerevoli problemi di catalogazione che si presentano nel fornire punti di accesso autore e titolo.

Il formato Dublin Core

Sembra universalmente accettato che Dublin Core sia l'applicazione di metadati più sviluppata e che sia sul punto di essere in genere favorevolmente accolto. Fu sviluppato da OCLC nel suo centro situato a Dublin, Ohio, e chiamato così dal nome di quella municipalità. Consiste di 15 elementi descrittivi etichettati. Considererò ciascuno di questi elementi e mostrerò come ognuno abbia relazioni con gli standard e la pratica di catalogazione esistente.

L'insieme di elementi dei metadati Dublin Core

1. Titolo
Etichetta: "Title"
Il nome dato alla risorsa, di norma dal Creatore o Editore



G. MATHIEU

[Il titolo proprio come definito da AACR2, campo MARC 245]

2. Autore o Creatore
Etichetta: "Creator"

La persona o organizzazione che ha la responsabilità principale della produzione del contenuto intellettuale della risorsa. Per esempio, autori nel caso di documenti scritti, artisti, fotografi, o illustratori nel caso di risorse visive. [L'autore (definito quasi nello stesso modo di AACR2, ma senza "contenuto artistico"). Campi MARC 1XX]

3. Soggetto e Parole chiave
Etichetta: "Subject"

L'argomento della risorsa. Tipicamente, il soggetto sarà espresso da parole chiave o frasi che descrivono l'argomento o il contenuto della risorsa. Viene incoraggiato l'uso dei vocabolari controllati e degli schemi formali di classificazione.

[Il soggetto come definito da elenchi di voci di soggetto/thesauri. Campi MARC 6XX]

4. Descrizione
Etichetta: "Description"

Una descrizione testuale del contenuto della risorsa, che può includere un riassunto analitico nel caso di oggetti simili a documenti, o descrizioni di contenuto nel caso di risorse visive.

[Note di contenuto o sommario come definite da AACR2. Campi MARC 505 o 520]

5. Editore
Etichetta: "Publisher"

L'entità responsabile di rendere disponibile la risorsa nella sua forma presente, come una casa editrice, un dipartimento universitario o un'entità collettiva.

[Editore come definito in AACR2. Sottocampo MARC 260#b]

6. Autore di contributo subordinato

Etichetta: "Contributor"

Una persona o organizzazione non specificata nell'elemento Creator che ha reso contributi intellettuali significativi alla risorsa, ma secondari rispetto alle persone o organizzazioni specificate nell'elemento Creator (per esempio, curatore, trascrittore e illustratore).

[Voci aggiuntive come definite da AACR2. Campi MARC 7XX]

7. Data

Etichetta: "Date"

Una data associata alla creazione o alla disponibilità della risorsa. Tale data non deve essere confusa con quella appartenente all'elemento Coverage, che sarebbe associata alla risorsa solo in quanto il contenuto intellettuale fosse in qualche modo relativo a quella data.

[Data della pubblicazione come definita in AACR2. Sottocampo MARC 260#c]

8. Tipo di risorsa

Etichetta: "Type"

La categoria della risorsa, come home page, romanzo, poesia, relazione, rapporto tecnico, saggio, dizionario.

[Campo MARC 008 a lunghezza fissa, posizione 27]

9. Formato

Etichetta: "Format"

Il formato dati della risorsa, usato per determinare il software e se possibile l'hardware che potrebbe essere necessario alla visualizzazione o al funzionamento della risorsa.

[Area 7.5 di ISBD(ER), campo MARC 516]

10. Identificatore della risorsa

Etichetta: "Identifier"

Una sequenza di caratteri alfabetici o numerici usati per identificare la risorsa in modo univoco. Esempi per risorse di rete includono URL e URN (quando sarà reso effettivo). Altri identificatori unici globalmente riconosciuti, come l'International Standard Book Number

(ISBN) o altri nomi formali sono inoltre candidati per questo elemento.

[Come definito da AACR2; ISBD(ER) 7.5.2 e area 8; campi del formato MARC 02X]

11. Fonte

Etichetta: "Source"

Informazione relativa a una seconda risorsa dalla quale è derivata la risorsa principale. Mentre è in genere raccomandato che gli elementi contengano informazione unicamente relativa alla risorsa presente, questo elemento può contenere una data, un creatore, un identificatore o altro metadato relativo alla seconda risorsa, considerato importante per il recupero della risorsa.

[Come definito da AACR2; campi del formato MARC 76X-78X]

12. Lingua

Etichetta: "Language"

La lingua del contenuto intellettuale della risorsa.

[Campo a lunghezza fissa del formato MARC 008, posizione 35-37]

13. Relazione

Etichetta: "Relazione"

Un identificatore di una seconda risorsa e della relazione con la risorsa principale. Questo elemento permette di indicare collegamenti fra le risorse collegate e le descrizioni delle risorse. Gli

esempi includono l'edizione di un'opera (versione di), la traduzione di un'opera (basata su), il capitolo di un libro (parte di) e la trasformazione meccanica di un dataset in un'immagine (formato di).

[Come definito da AACR2; campi del formato MARC 76X-78X]

14. Copertura

Etichetta: "Coverage"

Le caratteristiche spaziali e temporali del contenuto intellettuale della risorsa. La localizzazione spaziale si riferisce a una regione fisica (ad es., un settore celeste); si utilizzino coordinate (ad es., longitudine e latitudine) o nomi di luogo derivati da un repertorio di voci controllate e formulate per esteso. La copertura temporale è relativa a ciò di cui tratta la risorsa piuttosto che alla data di creazione o disponibilità (queste ultime appartenenti all'elemento Date).

[Campi a lunghezza fissa del formato MARC 008, posizioni 033 e 043]

15. Gestione dei diritti

Etichetta: "Rights"

Una formulazione dei diritti esercitati sulla risorsa, un identificatore che fa riferimento a una formulazione di tali diritti o a un servizio che fornisce questa informazione.

[Campi del formato MARC 506 e 561]

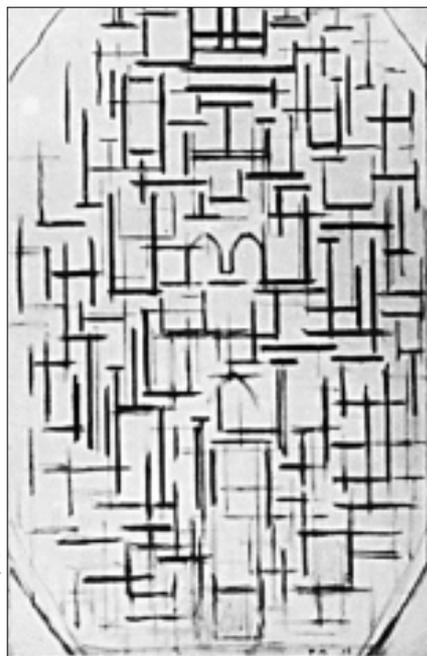
È facile vedere che ciò che abbiamo qui è, sostanzialmente, un sottoinsieme di MARC. Ognuno di questi elementi ha la sua controparte in MARC e il contenuto di ciascuno di essi è regolato da codici in campi MARC di lunghezza fissa o da codici di catalogazione/ISBD e/o da elenchi/thesauri di voci di soggetto.

Secondo un documento web rilasciato dal Surrey Institute for Art and Design per conto dell'ADAM Consortium,⁶ il Dublin Core possiede i seguenti attributi positivi:

è molto semplice da apprendere
ha elementi ripetibili

è provvisto di elementi opzionali
può essere esteso per applicazioni più complesse

può essere incorporato in modo invisibile nelle pagine web



P. MONDRIAN, 1912

è riconosciuto dal World Wide Web Consortium

Tutto ciò è vero, ma scarsamente rilevante riguardo all'interesse fondamentale per i metadati, considerato che nessuno parla in proposito dei punti centrali del contenuto del record o della natura limitata del sottoinsieme rappresentato dai 15 elementi.

Il bibliotecario australiano Warwick Cathro ha scritto un breve, interessante articolo in cui esamina il dibattito su Dublin Core relativo alla semplicità e alle altre caratteristiche della sua struttura e del suo contenuto.⁷ Egli cita Stuart Weibel di OCLC quando scrive:⁸

Il punto di vista minimalista rappresenta una forte predisposizione al concetto che la principale caratteristica qualificante di DC è la sua semplicità. Questa semplicità è importante sia per la creazione dei metadati (per esempio, da parte di autori privi di formazione nelle arti della catalogazione), sia per l'uso dei metadati da parte di strumenti (quali i software per la creazione degli indici, che non faranno probabilmente uso di qualificatori di dettaglio o di schemi di codifica). L'obiettivo ... può essere raggiunto solo se fosse disponibile un semplice nucleo di elementi che conservino lo stesso significato in ogni caso.

Il punto di vista opposto è presente in una citazione di Roger Clarke:⁹

Le priorità dei proponenti sono state espressamente orientate verso la semplicità, e lontano da strutture sofisticate. È implicito nel loro approccio che le due cose sono incompatibili. Il ... desiderio di semplicità ha dato luogo a un meccanismo incapace di rappresentare la ricchezza delle sfide del mondo reale che si presentano... è necessario che un modello più ricco, più sofisticato non sia sconvolgente o scomodo.

La visione minimalista, allora, è che Dublin Core dovrebbe contenere solo le sue attuali 15 catego-

rie e che il contenuto non dovrebbe essere normalizzato secondo l'uso delle persone esperte nell'arte della catalogazione" (per usare il termine di Weibel). Il punto di vista opposto esige maggiore ricchezza di contenuto e, implicitamente, l'esercizio di quelle autentiche "arti catalografiche" che hanno mostrato: a) di aumentare grandemente richiamo e precisione nel ricupero e b) di costare più denaro e sforzo di quelli che i minimalisti sono disposti a spendere e sostenere. Nelle parole di Peter Graham della Rutgers University:¹⁰

Ciò che è sotteso alla discussione "anti-MARC" è un desiderio di mantenere semplice il Dublin Core. Tutto ciò è buono e giusto, ma i desideri degli utenti sono complessi... Quelli che vogliono conservare semplice il record Dublin Core... hanno l'obbligo di descrivere quali sono gli ambiti di ciò che il record Dublin Core potrebbe fornire. Quelli (come me) che vogliono vedere il Dublin Core capace di espandersi sino a una considerevole capacità hanno l'obbligo di descrivere come ciò possa essere fatto preservando insieme il concetto di nucleo.

Dobbiamo essere chiari nella nostra comprensione di quale sia l'oggetto di questa discussione. Essa si riduce a una scelta tra una forma di catalogazione non costosa e inefficace in cui i 15 elementi del Dublin Core sono riempiti con testo libero non qualificato e non controllato – da una parte – e una forma di catalogazione dispendiosa e valida in cui almeno qualcuno degli elementi del Dublin Core è caricato con dati normalizzati e controllati, decisi sulla base di un esame professionale della risorsa. Nell'ultimo caso, e come minimo, i seguenti elementi del Dublin Core richiederebbero un intervento umano.

Titolo (DC#1): si dovrà decidere seguendo le regole di catalogazione (ad es., AACR2, regola 1.1B).



THEO VAN DOESBERG

Autore o creatore (DC#2): da determinare in accordo alle regole di catalogazione sia nei termini della scelta dell'autore/creatore che della forma in cui il nome è presentato.

Soggetto e parole chiave (DC#3): dovrà essere scelto all'interno dei repertori di voci di soggetto e dei thesauri.

Autore di altro contributo (DC#6): da decidere seguendo le regole di catalogazione per la scelta e la forma.

Tipo di risorsa (DC#8): una parola o frase derivata da un elenco standard di designazioni.

Fonte (DC#11): si dovrà determinare dall'esame della risorsa e formulare in modo standardizzato.

Lingua (DC#12): un codice dall'elenco MARC dei codici di lingua.

Relazione (DC#13): da determinare a partire dall'esame della risorsa formulandola in modo standardizzato.

Copertura (DC#14): codici di luogo e tempo dall'elenco MARC dei codici di località e data.

Tale intervento umano non dovrebbe, in tutta probabilità, essere così dispendioso in termini di tempo e costoso come la catalogazio-

ne completa, ma andrà certo oltre la semplicità e il basso costo desiderati dai fautori del punto di vista minimalista.

Cosa dovremmo fare?

Questa analisi mi porta a credere che abbiamo ora quattro approcci al controllo bibliografico delle risorse elettroniche:

- catalogazione completa
 - record Dublin Core arricchiti
 - record Dublin Core minimi
 - fare affidamento sulla ricerca per parole a pieno testo non strutturato.
- Se siamo preparati per formulare giudizi di valore e se possiamo raggiungere un vasto consenso (entrambi i “se” sono molto grandi), mi sembra che potremmo filtrare le risorse elettroniche in base al loro valore, ora e nel futuro – un processo non molto dissimile alla selezione dei materiali – e applicare questi quattro livelli di controllo bibliografico quando appropriati. Per esempio, quelle risorse elettroniche relativamente scarse (2%?) che si è visto possiedono un valore duraturo potrebbero essere catalogate in modo completo in accordo con gli standard nazionali e internazionali che si risolvono in record MARC esaustivi; il livello successivo potrebbe ricevere una

“catalogazione” Dublin Core arricchita; il livello seguente potrebbe consistere in un record Dublin Core minimo, e l’ultimo (che probabilmente riguarda la maggior parte delle risorse) potrebbe essere lasciato alla mercé dei motori di ricerca. Non desidero minimizzare la difficoltà di raggiungere una comprensione del valore delle singole risorse elettroniche, ma credo che il dibattito catalogazione/Dublin Core non possa essere risolto senza un piano coerente e un approccio sistematico al controllo bibliografico.

Conservazione

È importante notare che tutti i sistemi di controllo bibliografico hanno due funzioni primarie. Queste sono: a) fornire un accesso veloce ed efficiente ai materiali desiderati, e b) assistere nella conservazione e nella trasmissione ai posteri delle registrazioni umane. Nell’ultimo caso, è ovvio che, se non sappiamo cosa abbiamo, non possiamo conservare quei documenti. Sebbene una certa attenzione sia stata riservata alla conservazione dei documenti elettronici, gran parte degli operatori è stata scoraggiata dalla vastità del problema e dall’ulteriore fardello imposto dal mutamento tecnologico.

Credo che, in quanto bibliotecari, si abbia un dovere unico, quello di assicurare che le future generazioni sappiano ciò che noi sappiamo e che dovremmo, come priorità, elaborare dei piani per assicurare che ciò accada.

Proposte per giganteschi archivi elettronici e sistemi elettronici di archiviazione sembrano passarci sopra la testa nelle pro-

blematiche relative al finanziamento e in quelle tecnologiche.¹¹ Chi deve pagare per tale impresa, specialmente se progettata per continuare nel tempo? Dato che i record elettronici creati solo dieci anni fa sono inutilizzabili per molti a causa del continuo cambiamento nell’hardware e nel software, come facciamo ad essere sicuri che le risorse conservate in tale archivio saranno utilizzabili tra dieci, cinquanta, cento o duecento anni? Dobbiamo conservare hardware e software obsoleti (un compito verosimilmente impossibile per molte decadi) oppure aggiornare e “migrare” tutti i documenti e le risorse elettroniche dopo pochi anni? Se ogni risorsa viene conservata in un solo luogo, che cosa ci garantisce che questa unica copia sopravviverà indefinitamente? Queste e altre domande richiederanno una risposta e una risposta tra breve se non vogliamo perdere per sempre l’informazione e la conoscenza registrata di valore.

Quasi per caso, editori e bibliotecari hanno sviluppato un modo quasi a prova di tutto per preservare la conoscenza registrata e l’informazione nel corso degli ultimi cinquecento anni. Quella risposta consiste, naturalmente, nello stampare il testo e/o le immagini su carta alcalina, fare molteplici copie, e distribuire largamente quelle copie. È significativo notare che esistono molti libri perfettamente utilizzabili che hanno più di quattrocento anni. C’è qualcuno che pensa seriamente che i loro equivalenti elettronici di oggi saranno disponibili e utilizzabili alla fine del XXIV secolo? Esistono persone che non prenderanno mai in considerazione alcuna soluzione che non sia dipendente dalla tecnologia digitale, ma io propongo la seguente a voi e a loro in tutta serietà. La cosa migliore che potremmo fare, una volta selezionate le risorse elettroniche che stiamo per catalogare e/o per le ➤



R. DENNY, 1957

quali stiamo per costruire record Dublin Core arricchiti, è quella di stampare quelle risorse su carta priva di acidi, farne molte copie e distribuirle alle biblioteche di ricerca nel mondo. ■

Note

¹ “Comprende la questione che i cd-rom sono letteralmente ‘tangibili’ ma lo sono anche i nastri e i dischi rigidi dei computer. Il punto è che l’utente non tocca e non può toccare l’oggetto ma semplicemente usa la risorsa ‘intangibile’ di cui il cd-rom è un supporto. Credo ancora che ‘intangibile’ sia una parola appropriata per ogni risorsa elettronica visualizzata sugli schermi, indipendentemente dai mezzi attraverso i quali ci arrivi” (e-mail a Mauro Guerrini del 25 gennaio 2001).
² *Keyword* possiede anche il significato di parola chiave, ma in tutto l’articolo è reso con “parola”, con il significato che è quello ampiamente impie-

gato nell’ambito delle ricerche a pieno testo effettuate attraverso i motori di ricerca.

³ Per un’interessante discussione della distanza che ci separa dalle “macchine pensanti” cfr. E. J. GREGORY RAWLINS, *Slaves of the machine*, MIT Press, 1997 [N.d.t.: traduzione italiana: *Schiavi del computer?*, Bari, Laterza, 1999].

⁴ ISBD(ER) : International Standard Bibliographic Description for Electronic Resources – München : K. G. Saur, 1997. [N.d.t.: la pubblicazione in linea dello standard è del 2 luglio 1999 presso l’istituto dell’IFLA, <<http://www.ifla.org/VII/s13/pubs/isbd.htm>> e successivamente presso il sito mirror europeo <<http://ifla.inist.fr/VII/s13/pubs/isbd.htm>>. La traduzione italiana (colma di errori) è a cura dall’ICCU. ISBD(ER) : International Standard Bibliographic Description for Electronic Resources revised from the ISBD(CF): International Standard Bibliographic Description for Computer Files / International Federation of Library Association and Institutions, recommended by the ISBD(CF) Rewi-

ew Group. – Ed. italiana / a cura dell’Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche. – Roma : ICCU, 2000. – 122 p. ; 24 cm. – ISBN 88-7107-092-5].

⁵ <<http://www.lib.helsinki.fi/meta/nm-final.htm>> (datato luglio 1998).

⁶ <<http://www.adam.ac.uk/adam/metadata.html>> (datato dicembre 1997).

⁷ WARWICK CATHRO, *The Dublin Core: Simplicity or complexity?*, National Library of Australia. Staff paper, <www.nla.gov.au/staffpaper/cathro2.html> (datato 4 agosto 1997)).

⁸ STUART WEIBEL, *The 4th Dublin Core metadata report*, “D-Lib magazine”, July 1997, <<http://www.dlib.org/dlib/june97/metadata/06weibel.html>>.

⁹ ROGER CLARKE, *Beyond the Dublin Core*, <<http://www.anu.edu.au/people/Roger.Clarke/II/DublinCore.html>>.

¹⁰ Al 4° Dublin Core Workshop tenutosi a Camberra, Australia, nel marzo 1997.

¹¹ Si veda, ad esempio, *Preserving digital information*, <<http://www.rlg.org/ArchTF/>>.