

## **L'*authority control* nel contesto del controllo bibliografico in ambiente elettronico**

Michael Gorman  
California State University, Fresno

Esiste un significato per il quale *authority control* e controllo bibliografico sono sinonimi – due facce della stessa medaglia. Come minimo, il controllo bibliografico è letteralmente impossibile senza l'*authority control*. La catalogazione non può esistere senza punti d'accesso standardizzati e l'*authority control* è il meccanismo mediante il quale otteniamo il livello necessario di standardizzazione. La catalogazione ha a che fare con l'ordine, la logica, l'obiettività, la denotazione precisa e la coerenza, e deve possedere dei meccanismi che assicurino queste caratteristiche. Lo stesso nome, il titolo, il soggetto, dovrebbero avere sempre la stessa denotazione (nel linguaggio naturale o nei linguaggi artificiali della classificazione) ogni volta che ricorrono in una registrazione bibliografica, a prescindere da quante volte possa accadere. Senza accurate registrazioni di ciascuna forma autorevole, delle sue varianti, delle forme precedenti e delle regole sulle quali è basata (cioè l'*authority control*), è impossibile ottenere la desiderata e necessaria standardizzazione.

Iniziamo analizzando i fondamenti della catalogazione. Una registrazione catalogafica è costituita da tre parti:

- un punto d'accesso;
- una descrizione bibliografica;
- una localizzazione oppure (attualmente) il documento stesso<sup>1</sup>.

Il punto d'accesso guida l'utente alla registrazione; la descrizione permette all'utente di decidere se il documento desiderato è quello che si sta cercando di ottenere e la localizzazione porta l'utente al documento desiderato. Questa è una formulazione semplice e profonda, ed è la base di tutta la catalogazione. Anche l'infelice *Dublin core*, di cui parleremo diffusamente in seguito, contiene punti di accesso, elementi descrittivi e

---

<sup>1</sup> Nel caso di molti sistemi che danno accesso a documenti e risorse elettronici, la "localizzazione" è un URL, o qualcosa di simile, cliccando sopra il quale l'utente è condotto al documento o alla risorsa stessa.

informazioni di localizzazione. Ciascun elemento di una voce catalografica è standardizzato. La descrizione e la localizzazione sono presentati in forma standardizzata (altrimenti non potrebbero essere capiti). Gli standard che regolano la descrizione (soprattutto, le *International Standard Bibliographic Description*) e gli standard locali per la designazione della localizzazione non fanno parte dell'*authority control*. Quindi, l'*authority control* e il controllo del vocabolario ad esso correlato riguardano i punti di accesso e la loro standardizzazione. Il punto di accesso ha due funzioni basilari: permette all'utente di trovare la registrazione e raggruppa insieme le registrazioni che condividono una caratteristica comune [1]. Per poter svolgere la prima funzione deve essere standardizzato (ovviamente l'utente dovrebbe sempre trovare *Il gattopardo* sotto "Tomasi di Lampedusa, Giuseppe [Gattopardo]", e non qualche volta sotto questo punto d'accesso e in altre volte sotto "Lampedusa, Giuseppe Tomasi di" e/o qualche traduzione e variante del titolo). Questo è quello che conosciamo come controllo del vocabolario: la presentazione di tutte le denominazioni di nome (personale o collettivo), il titolo uniforme, la serie e il soggetto in una singola forma standardizzata. La ragione per cui abbiamo regole per la catalogazione descrittiva è che in questo modo qualsiasi catalogatore, seguendole, dovrebbe teoricamente ottenere, in ogni caso, lo stesso risultato standardizzato.

### ***Authority record***

Il controllo del vocabolario è vitale per l'*authority control* ma è soltanto il primo aspetto dell'*authority work*, sebbene sia il più importante. Robert Burger discute il concetto di *authority record* – il veicolo (tramite/mezzo di trasmissione) che contiene i risultati dell'*authority work* [2, p. 5]. In sintesi, Burger afferma che l'*authority record* ha cinque funzioni:

- registrare la forma standardizzata di ciascun punto d'accesso;
- assicurare il raggruppamento di tutte le registrazioni con lo stesso punto d'accesso;
- consentire registrazioni catalografiche standardizzate;

- documentare le fonti e le decisioni prese in merito ai punti d'accesso;
- registrare tutte le forme del punto d'accesso diverse da quella scelta come forma normative (cioè le forme dalle quali si fa rinvio).

A queste vorrei aggiungere:

- registrare tutte le forme precedenti e gli altri usi del punto d'accesso standardizzato affinché possano costituire una guida per i catalogatori.

In molti cataloghi cartacei e pre-OPAC, l'*authority record* esisteva solo implicitamente, nell'evidenza delle stesse voci catalografiche. I cataloghi *online* richiedono l'esplicita formulazione degli *authority record*, collegati alle voci catalografiche e contenenti almeno gli elementi necessari alle funzioni sopra elencate. Cioè, devono contenere:

- il punto d'accesso standardizzato;
- tutte le forme dalle quali si fanno rinvii del tipo “vedi”;
- un collegamento (un rinvio del tipo “vedi anche”) a tutti gli *authority record* collegati;
- le fonti dalle quali è stato ricavato il punto d'accesso standardizzato;
- l'elenco dei precedenti e degli altri usi della forma standardizzata.

In aggiunta, nelle registrazioni automatizzate l'*authority record* sarà collegato a tutte le registrazioni bibliografiche alle quali si riferisce<sup>2</sup> [3]. Il database creato riunendo tutti gli *authority record* usati in un catalogo è chiamato *authority file*, o, osservato sotto un altro punto di vista, *thesauro*.

### **Da dove proviene il contenuto degli *authority record*?**

Il contenuto degli *authority record* (la forma autorevole, le forme varianti, i collegamenti e le note di vario tipo) è ovviamente di grandissima importanza. Quando esistono forme

---

<sup>2</sup> Sto sostenendo questo sistema evoluto da più di un quarto di secolo. Si veda, ad esempio: [3].

varianti, esiste sempre una ragione per sceglierne una rispetto alle altre e, elemento cruciale, una fonte d'informazione tra tutte. L'agente principale in queste scelte è il codice di regole catalografiche in uso nell'area in cui si cataloga. Dal momento che non abbiamo un codice catalografico universale (sebbene le *Anglo-American cataloguing rules*, 2. ed. – AACR2 – abbiano portata mondiale) o liste di intestazioni per soggetto o sistemi di classificazione usati in tutto il mondo, i catalogatori in luoghi diversi possono giungere a conclusioni molto differenti, anche quando partono dallo stesso presupposto. Questo presupposto è un misto di oggettivo (l'evidenza dei materiali stessi e delle fonti di riferimento) e soggettivo (l'interpretazione del catalogatore delle regole o del significato del soggetto del documento catalogato).

Esistono alcune fonti che devono essere prese in considerazione quando si creano registrazioni catalografiche:

- gli *authority file* esistenti, nazionali e locali;
- il codice di catalogazione e liste di soggetto applicabili, ecc.;
- il documento in corso di catalogazione;
- le fonti di riferimento (nel senso più ampio del termine, qualsiasi fonte che fornisca dati utili).

Ciascuna di esse va valutata rispetto all'altra e, anche all'interno di ciascuna categoria, alcune fonti sono più autorevoli di altre. Nessuna fonte può essere considerata sempre la migliore. Ad esempio, l'evidenza del documento stesso può essere soppiantata da notizie trovate in altre fonti di riferimento se le regole del codice di catalogazione lo permettono. Ancora, un *authority file* nazionale può essere più autorevole in un caso, ma un *authority file* locale può esserlo in un altro (per conoscenze locali particolari). All'interno del documento stesso le informazioni trovate in una parte possono essere in conflitto con quelle presenti in un'altra. Infine, esiste un'evidente gerarchia nelle fonti di riferimento. Alcuni editori producono opere di qualità superiore di altri e la maggior parte delle fonti a stampa è più autorevole di gran parte delle risorse elettroniche. Il risultato è che il catalogatore deve essere capace di superare queste ambiguità grazie alla propria bravura, ad una buona capacità di giudizio e all'esperienza. La bravura consiste nella conoscenza della tipologia di materiale da catalogare, delle regole catalografiche, delle interpretazioni

date nel passato a queste regole e delle possibili fonti di riferimento con i loro punti forti e punti deboli. La buona capacità di giudizio consiste nell'abilità di valutare tutti questi fattori e di decidere basandosi sullo spirito delle regole se la loro formulazione risulta ambigua. L'esperienza risiede nell'insieme della conoscenza di regole, di politiche e di precedenti acquisita nel corso degli anni catalogando molti materiali. Con queste tre caratteristiche, un catalogatore può produrre registrazioni davvero autorevoli che gioveranno ai catalogatori e agli utenti in tutto il mondo.

### **Metadati e *authority control***

I metadati – letteralmente “dati sui dati” (una definizione che, se presa letteralmente, includerebbe la catalogazione “vera”) – nascono dal desiderio dei non-bibliotecari di migliorare la “ricuperabilità” delle pagine Web e di altri documenti su Internet. Il concetto base dei metadati è che si possono ottenere in quantità sufficiente richiamo e precisione (vedere oltre per una discussione di questi criteri) ricercando nei database senza il processo della catalogazione standardizzata, costoso in termini di tempo e soldi. In altre parole, una via di mezzo tra la ricerca a testo libero dei motori di ricerca (che è poco costosa ma inefficace) e la catalogazione completa (che a volte è lenta, ad alto impiego di manodopera, costosa ma molto efficace). Come tutti gli sforzi di questo tipo per cercare un accordo a metà strada, i metadati finiscono col non essere né una cosa né l'altra e, conseguentemente, non sono riusciti ad avere successo a nessun livello, parametro col quale giudicare ogni sistema di indicizzazione e recupero. Qualsiasi sistema può essere efficace se il database è piccolo. La vera prova è come il sistema gestisce database nell'ordine dei milioni. I cataloghi, anche quelli globali ed enormi come il database di OCLC, hanno dimostrato di essere efficaci. I motori di ricerca, anche i sistemi apparentemente avanzati come Google, sono, e lo si può dimostrare, inefficaci quando hanno a che fare con enormi database.

Dopo molte relazioni e numerosi convegni (un processo nel quale si sono riuniti bibliotecari traditori), un quasi-standard promosso da OCLC e chiamato *Dublin core* è emerso come esempio lampante di metadati e di ciò che potrebbero ottenere. Il *Dublin*

*core* (DC) è costituito da 15 “denotazioni”, ciascuna delle quali ha un equivalente più o meno esatto in una registrazione in formato MARC. Come ogni vero catalogatore sa, il MARC contiene ben più di 15 campi e sottocampi, oltre alle informazioni contenute nei campi codificati a lunghezza fissa. Inoltre, ci sono formati MARC per una molteplicità di tipi diversi di pubblicazioni, dai libri e periodici fino alle risorse elettroniche, che si vanno a sommare alle numerose denotazioni. Quelli che sostengono i metadati e, implicitamente o esplicitamente, credono che l’intera gamma di dati bibliografici possa essere contenuta in 15 categorie, ignorano il fatto che i formati MARC non sono il risultato di ghiribizzi e dell’impulso barocco dei catalogatori, ma si sono evoluti per andare incontro alle caratteristiche reali di documenti complessi di tutti i tipi. Ciò che abbiamo è una semplicistica (in molti modi ingenua) breve lista di categorie che ci si aspetta sostituisca la catalogazione una volta affidata a non-catalogatori.

La letteratura sui metadati è costellata di riferimenti alla “catalogazione MARC”, una espressione ignorante che tradisce la vacuità del concetto stesso di metadati<sup>3</sup>. Il MARC, come ogni catalogatore sa, è uno standard generale che contiene dati bibliografici. Non prescrive il contenuto dei campi – lasciando questo compito agli standard relativi al contenuto, come le AACR2, LCSH, etc. Le persone che parlano di “catalogazione MARC” chiaramente pensano che la catalogazione sia una questione di identificazione degli elementi di una registrazione bibliografica, senza specificare il contenuto di questi elementi. Quindi, è chiaro che queste persone non capiscono che cosa tratta la catalogazione. L’elemento più importante del controllo bibliografico è il *contenuto* e la natura controllata di questo, non le sue denotazioni. Così, quanto tutto il tumulto e le urla sono passati e i capitani e i re dei metadati hanno parlato, veniamo lasciati con l’assurda affermazione che un sottoinsieme di 15 campi della registrazione MARC, senza alcuna specificazione di come questi campi vadano riempiti dai non-catalogatori, è un qualche sostituto della catalogazione reale. Il fatto che i metadati e il *Dublin core* siano stati discussi *ad nauseam* per circa cinque anni con poche persone ad evidenziare nella discussione questo ovvio difetto, ricorda la storia del bambino e del re nudo. In questo caso, tuttavia, il re continua a gironzolare *senza vestiti* (il contenuto controllato), almeno finora.

## **L'*authority control* e il contenuto delle registrazioni bibliografiche**

Anche mettendo da parte il numero limitato e la natura delle categorie proposte per il *Dublin core* e gli altri schemi di metadati, ad essi mancano i concetti di vocabolario controllato e *authority work* – il mezzo mediante il quale i vocabolari controllati sono implementati e mantenuti. Date le strutture complesse delle registrazioni bibliografiche e la necessità di standardizzarne il contenuto, è evidente che il *Dublin core* non può avere successo con database di qualsiasi dimensione. Denotazioni casuali di soggetti, nomi, titoli e serie non soggette ad alcun tipo di standardizzazione – il controllo del vocabolario – produrranno progressivamente al crescere del database risultati rudimentali e, quando un database in *Dublin core* sarà di dimensioni sufficientemente grandi, i risultati non saranno più soddisfacenti di quelli ottenuti con una ricerca libera sul Web.

## **Precisione e richiamo**

Tutti i sistemi di recupero fanno affidamento su due elementi di misura cruciali: precisione e richiamo. In un sistema completamente efficiente, tutte le registrazioni recuperate dovrebbero essere perfettamente attinenti ai termini di ricerca (100% di precisione) e verrebbero recuperate tutte le registrazioni rilevanti (100% di richiamo). Per fare un semplice esempio, entrambi gli elementi di misura sarebbero perfetti se un utente si avvicinasse a un catalogo cercando le opere di Oscar Wilde e trovasse tutte le opere, e nessun'altra, di quell'autore possedute dalla biblioteca. Nella realtà, una biblioteca avrà un certo numero di opere di Oscar Wilde che non vengono recuperate ricercando nel catalogo (poesie contenute in antologie, saggi all'interno di miscellanee), ma ci si potrebbe attendere di trovare tutti i libri, le opere teatrali, le raccolte di lettere e di poesie di Wilde. Si può anche prevedere che la precisione sia alta in una ricerca fatta col termine "Wilde" in un catalogo ben organizzato e non darà alcuno, o pochi, materiali non legati a

---

<sup>3</sup> Si veda, ad esempio, fra i tanti: [4].

questo autore. Paragonate questo ai risultati di una ricerca libera fatta usando un motore di ricerca. Anche una semplice ricerca per autore per qualcuno che abbia un nome relativamente poco comune produrrà risultati aberranti. Ad esempio, per questa relazione ho fatto una ricerca su “Michael Gorman” usando Google. Ha prodotto «circa 7710» risultati. Tre dei primi dieci (i più rilevanti, si suppone) erano relativi a me. Gli altri erano relativi a un filosofo di Whashington, DC, a uno storico a Stanford, a un musicista folk irlandese e a un consulente tecnico di Denver, Colorado. Le restanti 7700 voci sono in ordine sparso e alcune non hanno nessuna relazione con qualcuno chiamato Michael Gorman. Questo è il risultato dell’assenza di *authority control*. Se ogni voce di Google fosse catalogata, avrebbe punti d’accesso per autore con forme standardizzate, per titolo e per soggetto, così le 7700 e più voci apparirebbero in ordine razionale – ogni voce relativa a ciascun Michael Gorman sarebbe raggruppata insieme alle altre relative alla stessa persona ma differenziata dalle voci relative agli altri Michael Gorman. In altre parole, il ricercatore avrebbe una ragionevole possibilità di identificare solo le voci relative al Michael Gorman che sta cercando (precisione) e di recuperarle tutte (richiamo).

Due cose sono ovvie. Il sistema con *authority control* è chiaramente superiore al sistema che ne è privo. In effetti, l’ultimo può a malapena chiamarsi “sistema”, dal momento che i risultati della ricerca sono per la maggior parte completamente inutili. Che cosa fa un ricercatore con centinaia di registrazioni in disordine e non differenziate? La seconda cosa ovvia è che fornire il controllo di vocabolario e l’*authority control* necessari a fare le ricerche con il sistema Google sarebbe proibitivamente costoso in termini di tempo e soldi. Questi due fattori sono il nocciolo del dilemma riguardante la “catalogazione del Web” e il controllo bibliografico del mondo di Internet e del Web. Se vogliamo assicurare precisione e richiamo nelle ricerche, dobbiamo avere vocabolari controllati, ma non possiamo permetterci economicamente di estendere questo controllo all’immenso ammasso di documenti Web marginali, di breve o nessuna utilità. Perché dovremmo farlo?



## Soluzioni

Innanzitutto, credo che dovremmo abbandonare l'idea dei metadati come qualcosa che sarà mai utile nei grandi database usati dai bibliotecari e dagli utenti delle biblioteche, oppure dovremmo arricchire gli schemi di metadati con le caratteristiche delle registrazioni bibliografiche tradizionali. L'idea di rinunciare ai metadati è allettante, anche solo perché è talmente ovvio che tali schemi, così come vengono utilizzati attualmente, non possono avere successo se non in piccoli database di nicchia relativi a materiali specializzati. Tuttavia, l'idea di arricchire i metadati per elevarli agli standard di catalogazione può essere più gradevole psicologicamente e politicamente. Dopo tutto, un buon numero di persone e organizzazioni influenti sono coinvolti o supportano gli schemi e i programmi relativi ai metadati, ed è difficile immaginare che affrontino la realtà e dichiarino morti i metadati applicati alle biblioteche.

Il dilemma che queste organizzazioni devono fronteggiare è precisamente racchiuso nel rapporto della biblioteca della Cornell University relativo alla loro partecipazione a CORC (COoperative Research Cataloging project), uno dei più grandi progetti relativi ai metadati:

«Molti membri dello staff non sono soddisfatti del modulo di selezione cartaceo che stiamo usando ora per far passare le informazioni relative alla selezione, alle acquisizioni e alla catalogazione, ed è dispendioso avere lo staff dei servizi tecnici che inizia il proprio lavoro partendo dagli scarabocchi della descrizione di ogni risorsa Internet. Ma se coloro che selezionano queste risorse e gli addetti al *reference* iniziassero a creare registrazioni preliminari, quanto ci si potrebbe attendere da loro, in termini di contenuto della registrazione? Si dovrebbe insegnare agli studenti e al personale addetto alle acquisizioni a usare CORD e DC per ottenere registrazioni preliminari? Inoltre, per restare sul problema sopra esposto, non siamo sicuri su come implementare l'insieme di elementi *Dublin core*. Ci dovrebbero essere delle direttive? Avrebbe senso essere d'accordo su alcune direttive di base per il contenuto di una registrazione CUL DC [*Cornell Dublin core*] da CORC (come ha fatto l'University of Minnesota)? Siamo sicuri che non vogliamo, usando DC, essere noiosi, complicati o perdere tempo, così, se avessimo delle direttive, queste dovrebbero essere semplici e facili da insegnare e usare [5]».

Il fatto è che l'uso di persone che non hanno la competenza e l'esperienza dei catalogatori per completare i campi dei metadati porterà, inevitabilmente, a database incoerenti e inutilizzabili. Un'altra realtà indiscutibile è che la catalogazione reale è, inevitabilmente, “dispendiosa in termini di tempo” e “complicata”. Il mondo della conoscenza e delle informazioni registrate è complicato, e il numero di complicazioni tende all'infinito. È impossibile immaginare un sistema che permetta il recupero coerente di informazioni rilevanti e contemporaneamente non abbia “direttive” (regole, cioè, che dettino la natura e la forma del contenuto delle registrazioni). Infine, il fatto che non si possa avere una catalogazione di alta qualità a buon mercato farà capolino tra coloro che sono coinvolti negli schemi di metadati, ma non, temo, prima di essere passati attraverso un lungo e costoso processo di educazione e rieducazione e aver sperimentato il fallimento dei database contenenti registrazioni senza standard e *authority control*.

La seconda soluzione sta in un esame rigoroso e completo della natura dei documenti e delle risorse elettroniche. Non possiamo e non dovremmo catalogare la maggior parte dei documenti elettronici, non più di quanto abbiamo catalogato i milioni di sottodocumenti e documenti effimeri appartenenti al mondo della stampa. I problemi sono ora come identificare quei documenti elettronici di valore durevole e, una volta catalogati, come conservarli. Questi sono problemi profondi e complessi, a cui non è facile dare risposta, ma sono vitali per i nostri progressi nel rendere i documenti e le risorse elettroniche disponibili per tutti gli utenti<sup>4</sup>.

## **Conclusioni**

L'*authority control* è centrale e vitale per quell'insieme di attività che chiamiamo catalogazione. La catalogazione – la riunione logica di dati bibliografici in registrazioni recuperabili e usabili – è quell'attività che permette alla biblioteca di svolgere le sue missioni principali, il servizio e l'accesso gratuito e libero a tutta la conoscenza e le informazioni registrate. Non possiamo avere un reale servizio bibliotecario senza

un'architettura bibliografica, e non possiamo avere questa architettura bibliografica senza l'*authority control*. La questione è semplice e profonda allo stesso tempo.

## Note bibliografiche

[1] Helen F. Schmierer. *The relationship of authority control to the library catalog*. «Illinois libraries», 62 (1980), p. 599-603.

[2] Robert H. Burger. *Authority work*. Littleton, Colo.: Libraries Unlimited, 1985.

[3] Michael Gorman. *Authority files in a developed machine system*. In: *What's in a name*, ed. and comp. by Natsuko Y. Furuya. Toronto: University of Toronto Press, 1978, 179-202.

[4] Stuart Weibel. *CORC and the Dublin core*. «OCLC newsletter», (May/June 1999), n. 239.

[5] *CORC at Cornell: final report*. <<http://campusgw.library.cornell.edu/corc/>>

[6] Michael Gorman. *The enduring library*. Chicago: ALA, 2003.

---

<sup>4</sup> Per una trattazione estesa sull'argomento, si veda: [6, cap. 7].