

Un catalogo “ricco”

Alcune proposte per accrescere le potenzialità informative dello strumento principe di relazione tra patrimonio e utente

di Meris Bellei

In una recente conversazione con Luigi Crocetti, si discorreva di come la catalogazione sia passata da una antica centralità all'interno della professione, senza dubbio eccessiva, ad una sottovalutazione rispetto a mansioni a diretto rapporto con l'utente. Sarà anche stato un processo salutare ma, come spesso accade nei momenti di cambiamento, si è forse travalicata la misura. Se l'automazione e la cooperazione hanno liberato energie prima dedicate alla catalogazione individuale, è stato giusto riversarle nei servizi di *reference* per ridurne le carenze; ma il catalogo resta lo strumento principe di relazione tra patrimonio e utente, pur attraverso una migliore mediazione del bibliotecario, e quindi dipenderà in buona parte dalla sua qualità l'esito di ogni servizio di consulenza e in definitiva di ogni buona relazione con l'utente. Qualche energia dovrà forse allora tornare al catalogo.

Nelle riviste di biblioteconomia di area anglosassone il dibattito sulle caratteristiche della catalogazione intesa in termini tradizionali e sulla catalogazione del futuro è ampiamente sviluppato, ed è avvertita

l'urgenza di porsi due obiettivi: accrescere le possibilità di esito positivo della ricerca a catalogo, e dare maggiori opportunità di comprendere se il documento identificato è utile per la ricerca che si sta facendo [1, p. 74-75].

Limitatamente al primo punto, si è riflettuto anche da noi in questi anni sulle caratteristiche della catalogazione descrittiva, sull'importanza delle liste di autorità, sui criteri per la formulazione dei soggetti; e in questa direzione molto ancora resta da riflettere e da fare. È tuttavia forse giunto il momento per una riflessione più ampia, che prospetti una sorta di rifondazione del catalogo attraverso l'utilizzo dei nuovi strumenti disponibili: un catalogo, quindi, “arricchito”,¹ che tenga presente tutti e due gli obiettivi e non solo il primo.

Un fatto è evidente: i cataloghi ottocenteschi e di inizio secolo, pur in genere privi di sistematicità e omogeneità di scelte, offrono spesso informazioni sul contenuto dei documenti più di quanto non facciano i cataloghi odierni, e ciò grazie all'inclusione di indicazioni analitiche sulle opere racchiuse nelle pubblicazioni; negli ultimi

decenni l'attenzione si è invece spostata in modo netto su altri temi: la costruzione di basi di dati nazionali, lo sviluppo degli standard, i formati di scambio dei dati; impegno lodevole, ma col triste risultato di escludere ogni attenzione al contenuto informativo delle pubblicazioni [2; 3, p. 7; 4, p. 17].

Ora si propone di andare oltre la catalogazione analitica, e il discorso ha difficoltà a svilupparsi in una realtà come la nostra, in cui sull'analisi la riflessione teorica è a livello iniziale, e la pratica di lavoro tende a ignorare la questione.²

Vediamo allora alcune proposte rivolte ad ottenere il duplice risultato descritto sopra.

Arricchire la descrizione

In aggiunta all'eventuale aumento dei punti di accesso, la registrazione bibliografica si arricchisce se si espande il suo contenuto descrittivo, tanto da cambiarne la natura [1, p. 74]. In che senso?

Ora la descrizione è composta di una successione di aree corrispondenti ognuna ad un elemento “estratto” dal volume; di questi nessuno ha specificità, tanto è vero che in genere appartiene a liste di autorità dense di legami con molte descrizioni. Ciò che caratterizza la registrazione nel suo insieme è allora — a parte i dati di collazione — l'assemblaggio, la particolare combinazione degli elementi e dei loro ruoli all'interno della registrazione: in questo specifico record, un nome ha relazione di autore, un titolo funge da collana, si aggiunge un editore trovato nella sua lista, un soggetto dal soggettoario, un codice di classificazione dall'elenco dei codici.

Si potrebbe arricchire il numero e il carattere degli elementi di descrizione *distintivi*, cambiando così la natura stessa del record e più in



generale dell'opera di catalogazione. Vediamo alcuni esempi.

Il titolo è elemento della descrizione, se lo si guarda dal punto di vista del catalogatore: richiede infatti un'opera di trascrizione e non di interpretazione. Ma dal punto di vista della ricerca è uso riferirsi al titolo, oltre che per la identificazione, anche per informazioni sul contenuto delle pubblicazioni, soprattutto nei cataloghi in linea che consentono la ricerca attraverso le singole parole. D'altra parte il titolo è spesso privo di informazioni e fuorviante, in particolare nelle scienze umane. Si potrebbe allora aggiungere al titolo parole o frasi che indichino il reale contenuto (*enriching word or phrases*), facendo affidamento sulla potenza degli OPAC per il recupero dell'informazione. Al di fuori del contesto della soggettazione, e quindi con libertà

di linguaggio, questo metodo può consentire flessibilità per usare, ad esempio, termini in altra lingua entrati nell'uso di particolari discipline [3, p. 9].

I rapporti di indagine rivelano un aumento del grado di richiamo nella ricerca per argomento quando si aggiungono informazioni sul contenuto recuperabili per via automatica (*content enriched access*) [5, p. 134]. Il metodo può prevedere l'estrazione di parole significative dall'indice analitico del volume, dal sommario, da altre parti di rilievo come i titoli dei capitoli: ma anche la completa o parziale trascrizione di questi elementi, da associare alla descrizione tradizionale per aumentarne la capacità informativa. Nel caso di pubblicazioni composite come antologie, raccolte, atti di congressi, l'inserimento nel catalogo del sommario aggiun-

ge informazioni su nomi e titoli altrimenti esclusi.

Un esperimento in atto all'Università di Toronto chiarisce quali possono essere le condizioni di un metodo come questo: il riferimento principale è al sommario, a condizione che contenga una voce ogni 25 pagine di testo in media, che sia formulato con termini significativi e non allusivi o enfatici, e che non contenga troppe ripetizioni; in caso contrario, l'accento si sposta su altre parti o sulla pubblicazione nel suo insieme. L'esito è, al momento, la individuazione di 300 parole per titolo, raggruppate in 30 punti di accesso sotto forma di frasi o gruppi di termini (come titoli di capitoli, linee di sommario) [1, p. 78].

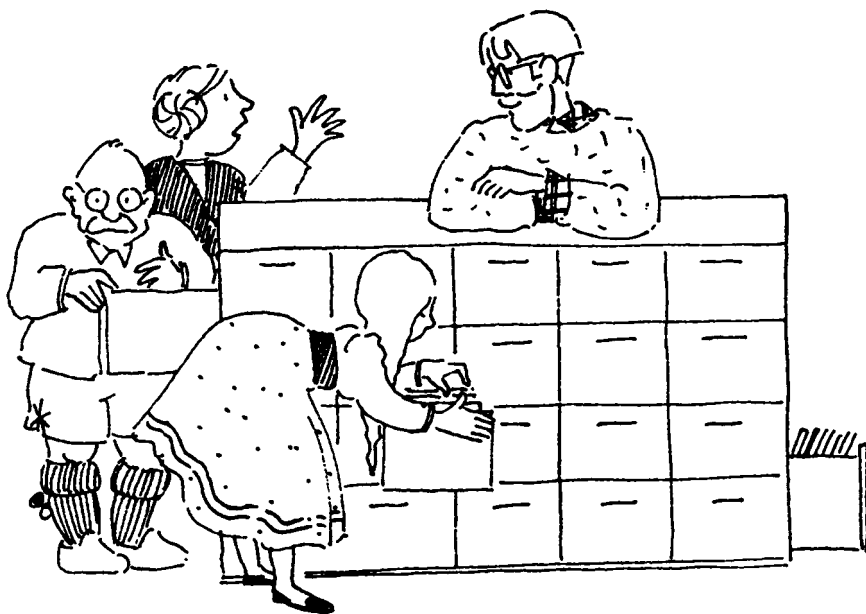
È evidente come l'*abstract* sia il principale candidato alla produzione di parole significative: ma la ►

maggior parte dei libri non lo contiene, a vantaggio di presentazioni editoriali in quarta di copertina non affidabili ai fini della ricerca. Dovrebbe essere l'autore stesso, su sollecitazione dell'editore, a formulare la sintesi del contenuto; in assenza, la sostituzione da parte del catalogatore è troppo costosa per essere presa in considerazione.

Arricchire l'indicizzazione

Come per la descrizione bibliografica gli obiettivi avanzati presuppongono il raggiungimento di livelli di qualità nella pratica tradizionale — applicazione degli standard, omogeneizzazione delle basi di dati, completezza delle descrizioni — così per la indicizzazione ogni riflessione sul futuro deve considerare conquistati gli obiettivi della uniformità nell'uso dei descrittori di soggetto e dei codici di classificazione, e del rispetto dei criteri di coestensione e specificità [6, p. 10]. Su questi temi si sono fatti molti passi avanti, e la catalogazione partecipata ha decisamente agevolato i risultati. Molta strada resta invece da percorrere rispetto alla costruzione delle reti dei richiami, ed è una via impervia perché solitaria, riferita al patrimonio e al pubblico di ogni biblioteca.

Per arricchire la indicizzazione forse il punto da cui cominciare è l'aumento del numero degli accessi per soggetto (*additional subject headings*), che da ricerche risulta assestato su una media di 1,3 per documento, senza aumento negli ultimi cinquant'anni [7, p. 10]. In nome della specificità, ad un documento potranno essere assegnati numerosi soggetti, senza più il vincolo dato dai cataloghi cartacei: nei cataloghi elettronici, risulta in effetti poco giustificata la pratica di ricondurre ad un concetto più ampio gli argomenti che si presentino in numero superiore ad un tetto



molto basso, in genere pari a tre [1, p. 75-76]. L'osservazione delle prassi correnti dimostra una palese ambiguità: quando si tratta di definire metodi e criteri, l'accento è sulla specificità (il soggetto deve avere lo stesso grado di copertura semantica dell'argomento del documento), mentre quando si lavora in concreto, le direttive tendono al risparmio nel numero delle intestazioni, col risultato di sostituire a numerosi soggetti precisi uno più generale che li comprende.

Alle intestazioni per soggetto espresse in linguaggio controllato, si potrebbero affiancare soggetti in linguaggio naturale (*naturale language subject headings*), vicini al lessico usato nei documenti e perciò in grado di riflettere l'evoluzione della terminologia delle diverse discipline. La *sostituzione* del linguaggio naturale a quello controllato sarebbe rovinosa rispetto allo sforzo di standardizzazione e soprattutto rispetto ai risultati della ricerca dell'utente [8, p. 332]; ma l'aggiunta porta solo arricchimento, tanto più se il sistema prevede connessioni tra i termini "liberi" e il soggettario a linguaggio controllato³ [1, p. 81].

In aggiunta e sulla stessa lunghezza d'onda, si potrebbe estrarre una serie di descrittori di soggetto dalla Classificazione decimale Dewey, utilizzando sia le liste di codici sia gli indici; naturalmente dovrebbe provvedere l'elaboratore ad assegnare per via automatica i termini di soggetto ogni volta che il catalogatore seleziona un codice di classificazione: non si aggiunge perciò lavoro di catalogazione, ma solamente un'opera preliminare di caricamento di termini e di connessioni. L'effetto per l'utente finale è un consistente arricchimento del vocabolario di interrogazione per soggetto. I risultati di un progetto sperimentale, il *DDC Online Project*, dichiarano un contributo medio per documento di 10,19 termini dalle liste della CCD, e di 4,23 termini dagli indici [1, p. 79].

Arricchire la catalogazione

Oltre ad agire sulle caratteristiche del record bibliografico come proposto sopra, si può modificare il numero delle unità di informazione prese in esame a parità di patri-

monio bibliografico, attraverso una catalogazione analitica. La suddivisione delle pubblicazioni complesse — libri a più mani, fascicoli di periodico, audiovisivi contenenti più opere autonome — in “parti” da catalogare individualmente fa emergere l’informazione “nascosta” nel patrimonio catalogato con i metodi tradizionali, e porta ad un aumento esponenziale delle informazioni disponibili per l’utente [9, p. 633; 10, p. 175-177].

Ma c’è chi approfondisce il tema con la proposta di definire in termini davvero molto ampi il concetto di unità di informazione: si potrebbe ad esempio considerare che un catalogo di mostra sia a questo riguardo una vera “miniera d’oro bibliografica”, composta di tante unità da catalogare separatamente quante sono le coppie di due pagine che contengono l’una la riproduzione di un’opera d’arte e l’altra il breve saggio a commento [3, p. 8]. È evidente che si tratta di una proposta impegnativa, che parte comunque da un dato tangibile: il catalogo come tale è cieco rispetto a molte chiavi di ricerca, quindi in buona parte perduto: la sua analisi di dettaglio massimizza allora le possibilità di recupero e consultazione. Un tale concetto di unità di informazione non potrebbe certamente mai essere codificato, legato com’è a stretto nodo con lo specifico contesto; ma il punto è proprio quello di condividere una impostazione, e successivamente trovare le giuste risposte alle esigenze locali.

In alternativa o in aggiunta, il concetto di unità di informazione può essere mantenuto flessibile, in modo che possano coesistere più livelli di analisi del documento, riferiti all’insieme e contemporaneamente a una o più parti che lo compongano (*double indexing*). Il catalogatore assume così ampia discrezionalità, e potrà valutare opportuno catalogare un libro, un ca-

pitolo nel libro, una sezione o paragrafo nel capitolo [3, p. 9].

Per favorire un percorso come questo una commissione ha discusso le possibilità offerte dal formato MARC, per scegliere la migliore opportunità per l’inserimento di registrazioni analitiche di capitoli di libro e il tipo di connessione tra queste registrazioni e la descrizione madre cui debbono essere collegate.⁴

L’accento sul capitolo di libro deriva dalla considerazione che l’articolo di periodico sia almeno in parte reso disponibile per la ricerca dalle basi di dati commerciali. In realtà è noto quanti siano i problemi di copertura delle testate e di qualità della indicizzazione, e allora una proposta economica di spoglio di periodici potrebbe consistere nell’acquisto sul mercato dei

sommari dei fascicoli trasferiti su formato elettronico, per rendere disponibili quanto meno i dati descrittivi; lo sviluppo dei sistemi di lettura ottica dei caratteri può rendere conveniente a un certo livello la digitalizzazione effettuata in proprio [5, p. 133].

Arricchire gli strumenti di ricerca

L’offerta di un catalogo più ricco non può passare solo per la via dell’aggiunta di dati alle registrazioni, molto costosa in termini di lavoro. Anche l’evoluzione della tecnologia dovrà corrispondere a questa esigenza, e andare oltre l’opportunità di interrogazione per parole chiave e con operatori booleani, da considerare il livello ►



minimo anche se non ancora prevista da tutti i sistemi. Una possibilità di ricerca avanzata non sopprime certamente alla povertà dei dati immessi, ma rende invece più produttivo l'impegno speso nel miglioramento della qualità della catalogazione: è quindi non il sostituto, ma il giusto complemento di un catalogo in linea più ricco di dati.

Uno dei punti critici da affrontare è certamente la ricerca per soggetto, che rischia di registrare un basso livello di richiamo per la difficoltà di approccio dell'utente a un linguaggio controllato, e per l'assenza di aiuto al riguardo; allora un sistema di relazioni automatiche tra descrittori, e di rinvio dalle voci scartate alle voci ammesse, porterebbe a un consistente miglioramento perché consentirebbe una ricerca libera, tradotta poi per via automatica nel linguaggio usato dal catalogatore e riconosciuto dal sistema [1, p. 82].

È ovvio che la predisposizione di una rete come quella ipotizzata richiede un lavoro lungo e qualificato; la biblioteca può agevolarlo e prepararlo, per esempio monitorando le ricerche degli utenti che non hanno avuto esito, e incaricando gli addetti al *reference* di prendere nota dei termini usati dagli utenti che chiedono il loro aiuto: questo vocabolario inconsapevolmente redatto dagli utenti può costituire il nucleo del sistema di traduzione dal linguaggio libero a quello standard.

Si può inoltre pensare di modificare la risposta negativa alla ricerca dell'utente, abilitando il sistema a non fare semplicemente un raffronto tra stringhe di caratteri il cui esito può solo essere del tipo "sì-no", ma a presentare termini vicini a quello ricercato e suggerimenti di variazione che possano portare a risposte positive. Una sperimentazione in questo senso condotta a Londra dimostra che un sistema

sofisticato è in grado di "suggerire" una giusta correzione in circa la metà degli errori materiali di scrittura e di errata codifica dei termini di ricerca [1, p. 88]; lo sviluppo naturale porta a prevedere il riconoscimento di forme diverse della stessa parola, la visualizzazione di proposte di modifica dei termini immessi, la previsione di controlli ai diversi stadi della ricerca per *risparmiare il tempo dell'utente*. Si tratta, nella sostanza, di poter fare affidamento su strumenti che instaurano un rapporto "intelligente" con l'utente, dialogano con lui non solo amichevolmente ma anche con costrutto, e davvero producono miglioramento degli esiti della ricerca: per fare questo, deve trattarsi di sistemi davvero "esperti".

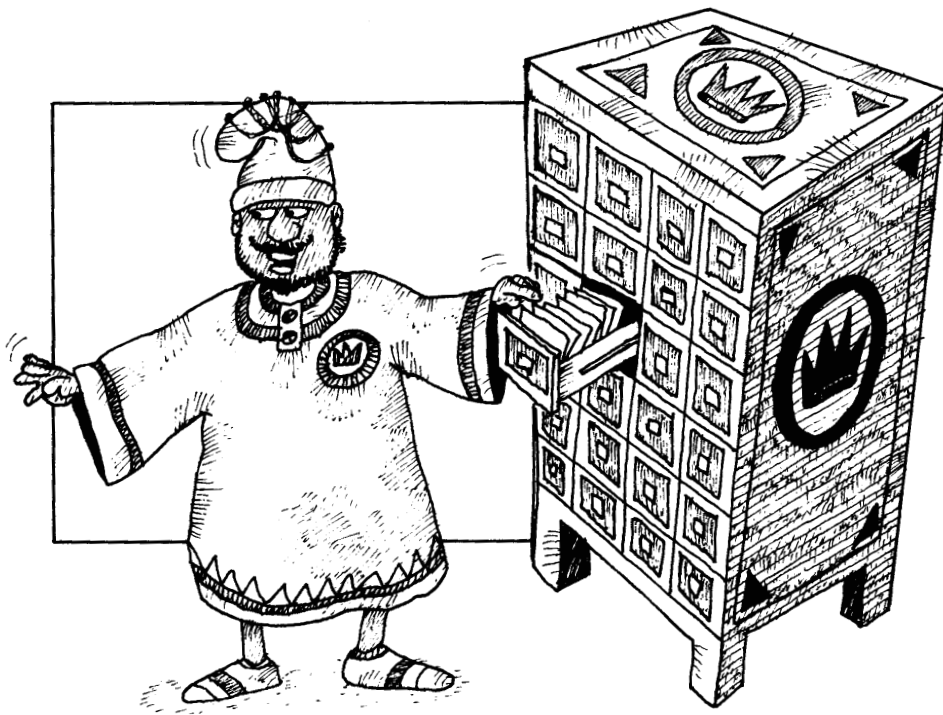
Una leva già disponibile per arricchire le prestazioni del catalogo in linea è l'informazione all'utente sulle sue caratteristiche, e la istruzione all'uso di tutte le sue opportunità: spesso infatti si assiste a un sottoutilizzo, circoscritto alle funzioni di base senza ricorso alle possibilità di limitare le ricerche troppo estese, di usare gli operatori booleani, di modificare le stringhe di ricerca attingendo alle liste di autorità. Quasi sempre in biblioteca il passaggio dal catalogo cartaceo a quello in linea non è accompagnato dalla consapevolezza che il secondo richiede un maggior numero di istruzioni rispetto al primo: e invece sono necessari manuali d'uso collocati vicino ai terminali, sistemi di aiuto consultabili a video, istruzioni più formali trasmesse a voce o per scritto. Un elemento indispensabile comunque è la presenza di bibliotecari ben addestrati a conoscere tutte le opportunità e a risolvere tutti i problemi. La misura finale del risultato sarà data, più che dal grado di soddisfazione dell'utente al terminale, dall'aumento nella circolazione dei documenti [1, p. 84-85].

Considerazioni a margine

Arricchire il catalogo sottintende sapere assumere decisioni: quale livello di complessità scegliere, in primo luogo, e secondariamente a quali parti del patrimonio rivolgere l'attenzione. Sarà infatti improbabile che tutte le proposte riassunte sopra possano trovare applicazione, e che i catalogatori siano in grado di estenderle a tutte le registrazioni bibliografiche. Un buon risultato già sarebbe selezionare una parte anche molto ristretta sulla quale sperimentare una catalogazione diversa, andando a costituire di fatto un subcatalogo approfondito dentro al catalogo generale [5, p. 134].

Ma come scegliere? Pare che si possa estendere a queste considerazioni quanto dice Carlo Revelli a proposito di catalogazione analitica: "La valutazione dei documenti da segnalare [*nel nostro caso, da catalogare in maniera più ricca*] non può essere uguale per tutti ed ancora una volta ci troviamo di fronte alle domande sul perché questa biblioteca esiste, che cosa desidera il suo pubblico" [11, p. 40]. Ogni biblioteca è così sollecitata a "impostare una politica catalogografica conveniente alla propria natura e al proprio pubblico"; le scelte conseguenti relative al catalogo possono indurre a prendere in esame pubblicazioni omogenee sul piano formale — atti di congresso, raccolte di saggi, antologie — oppure appartenenti a una stessa area concettuale, scelta perché contraddistingue la biblioteca e contribuisce a delinearne l'identità specifica.

Da un punto di vista organizzativo, l'impatto sull'ufficio catalogazione è forte: e questo sia se il nuovo lavoro viene distribuito su tutti i catalogatori, sia che si vada a identificare una figura professionale autonoma che si collocherebbe a metà tra il catalogatore e il docu-



mentalista, sia che si pensi a soluzioni innovative prevedendo l'intervento in catalogazione dell'addetto al *reference* in veste di esperto della ricerca e delle sue esigenze [3, p. 9]. Nessuna infatti delle operazioni proposte è di tipo esecutivo, anzi ognuna presuppone un alto grado di professionalità, che metta in grado di selezionare le giuste stringhe di caratteri per un'intestazione di capitolo, o di indicizzare parti di pubblicazioni di argomento minuto. Una mancanza di precisione invaliderebbe buona parte dello sforzo: errori banali di scrittura o di codifica possono essere la causa di mancata concordanza tra termini immessi per la ricerca e descrittori usati dai catalogatori e, più in generale, una bassa qualità nel lavoro produce danni seri, contenibili solo attraverso una verifica costante di tutto il flusso di attività.

I costi, è ovvio, non sono quantificabili facilmente, perché legati alle scelte specifiche. Solo come esempio, si citano le deduzioni di un

recente progetto statunitense rivolto a estrarre termini dalle parti significative delle pubblicazioni: a ogni descrizione bibliografica sono stati aggiunti in media 20,7 intestazioni e 53,6 termini di ricerca; questo ha richiesto un tempo di lavoro che va dai dodici ai quindici minuti [1, p. 86].

Questi numeri portano a considerare un altro aspetto della questione, gli effetti provocati nel catalogo dalla immissione di tante nuove informazioni. Di certo si avrà un risultato migliore di richiamo (o recupero di informazioni rispetto a quelle potenzialmente esistenti nella base di dati), e di contro si perderà in precisione, intesa come percentuale di notizie utili tra tutte le notizie recuperate. Il *rumore* che si genera nel catalogo è da considerare come punto critico, elemento di valutazione per determinare il punto di equilibrio o "limite di utilità" tra la quantità di informazioni che si immettono e il vantaggio prevedibile per gli utenti [11, p. 41]. Da una parte sta la mole di

notizie nascoste dentro le pubblicazioni e ora in gran parte perdute perché non rispecchiate nel catalogo; dall'altra l'insieme delle registrazioni bibliografiche potenziali, riferite sia alle pubblicazioni come tali che a loro porzioni anche molto minute; nel mezzo deve esprimersi la capacità del bibliotecario di individuare la giusta strada che valorizzi al massimo quel particolare patrimonio senza sovraccaricare di troppe risposte la ricerca degli utenti. Per ora nel nostro paese non pare che quest'ultimo rischio sia all'ordine del giorno.

Selezionare l'ambito da catalogare in maniera approfondita sdrammatizza il punto dell'assetto del catalogo. Tante informazioni e il conseguente pericolo di rumore potrebbero far valutare positivamente la formazione di un catalogo speciale inserito dentro o connesso al catalogo complessivo, così da permettere una ricerca tradizionale sul secondo e — solo a chi lo scelga — indagini sofisticate sul primo [5, p. 134]. Ma se il catalogatore decide opportunamente e non genera quantità poco utili di notizie, la soluzione dell'unico archivio di dati in cui fare ricerche di ogni tipo è senz'altro preferibile, per non rischiare di "nascondere" ancora una volta informazioni.

Nella sostanza, la proposta prevede di mantenere un costante investimento sulla catalogazione, non più per esaminare in mille sedi la stessa pubblicazione e registrarne le caratteristiche di base: su questa parte di lavoro la catalogazione partecipata in linea ha già portato a consistenti risparmi, e ancora si potrà procedere nella riduzione delle risorse umane necessarie: Una volta che il nucleo del catalogo possa essere quasi sempre recuperato, restano le scelte locali di integrazione per offrire il massimo supporto alla ricerca: l'operazione non sarà né facile né economica, ma consentirà di mettere a ►

frutto le potenzialità delle nuove tecnologie e di mantenere così il passo con le altre agenzie di informazione. ■

Note

¹ La definizione *enhanced catalog*, qui resa con *catalogo arricchito*, è all'ordine del giorno nelle pubblicazioni scientifiche in lingua inglese degli anni Novanta. Nel termine inglese è incluso un concetto di "miglioramento della qualità" che in parte si perde nella traduzione italiana.

² Recenti segnali nella direzione di una valutazione dell'importanza della catalogazione analitica sono l'inserimento del tema "spoglio" nel programma del corso di aggiornamento professionale sui periodici organizzato da Biblionova, e la previsione di lezioni sulla catalogazione analitica all'interno del corso per bibliotecari addetti alla gestione di materiali musicali ideato dalla Scuola di paleografia e fi-

lologia musicale di Cremona (Università di Pavia), con la direzione scientifica di Mauro Guerrini.

³ Per il problema del *rumore*, vedi l'ultima parte di questo saggio.

⁴ Documenti non pubblicati, citati e parzialmente riprodotti in [1], p. 77.

Riferimenti bibliografici

[1] S.A. WITTENBACH, *Building a better mousetrap: enhanced cataloging and access for the online catalog*, in *Advances in online public access catalogs*, Westport, Meckler, 1992, p. 74-92.

[2] M. BELLEI, *Analisi*, Roma, Associazione italiana biblioteche, 1998.

[3] J. STANLEY, *Reference librarian as a cataloger: analytical indexing as front-end reference*, "Art documentation", 14 (1995), 4, p. 7-9.

[4] K.H. WEIMER, *The nexus of subject analysis and bibliographic description: the case of multipart videos*, "Cataloging and classification quarterly", 22 (1996), 2, p. 5-19.

[5] C. POULSEN, *Tables of contents in library catalogs: a quantitative examination of analytic catalogs*, "Library resources and technical services", 40 (1996), 2, p. 133-138.

[6] ASSOCIAZIONE ITALIANA BIBLIOTECHE, GRUPPO DI RICERCA SULL'INDICIZZAZIONE PER SOGGETTO, *Guida all'indicizzazione per soggetto*, Roma, Associazione italiana biblioteche, 1996.

[7] C.R. MCCLURE, *Subject and added entries as access to information*, "Journal of academic librarianship", 2 (1976), 1, p. 9-14.

[8] A. CHETI, *L'indicizzazione per soggetto negli anni Novanta*, "Bollettino AIB", 36 (1996), 3, p. 329-333.

[9] P. FRANTZ, *A gaping black hole in the bibliographic universe*, "American libraries", 21 (1990), 7, p. 632-633.

[10] M. BELLEI, *L'informazione nascosta*, "Biblioteche oggi", 10 (1992), 2, p. 175-185.

[11] C. REVELLI, *Il catalogo*, Milano: Editrice Biblio grafica, 1996.

[12] S. INTNER, *Functional inaccessibility in libraries*, "Technicalities" 9 (1989), 12, p. 3-5.