

Aspettando l'Opac nasce una base dati

Dall'information retrieval BIR al Catalogo bibliografico delle università padane

di Maurizio di Girolamo

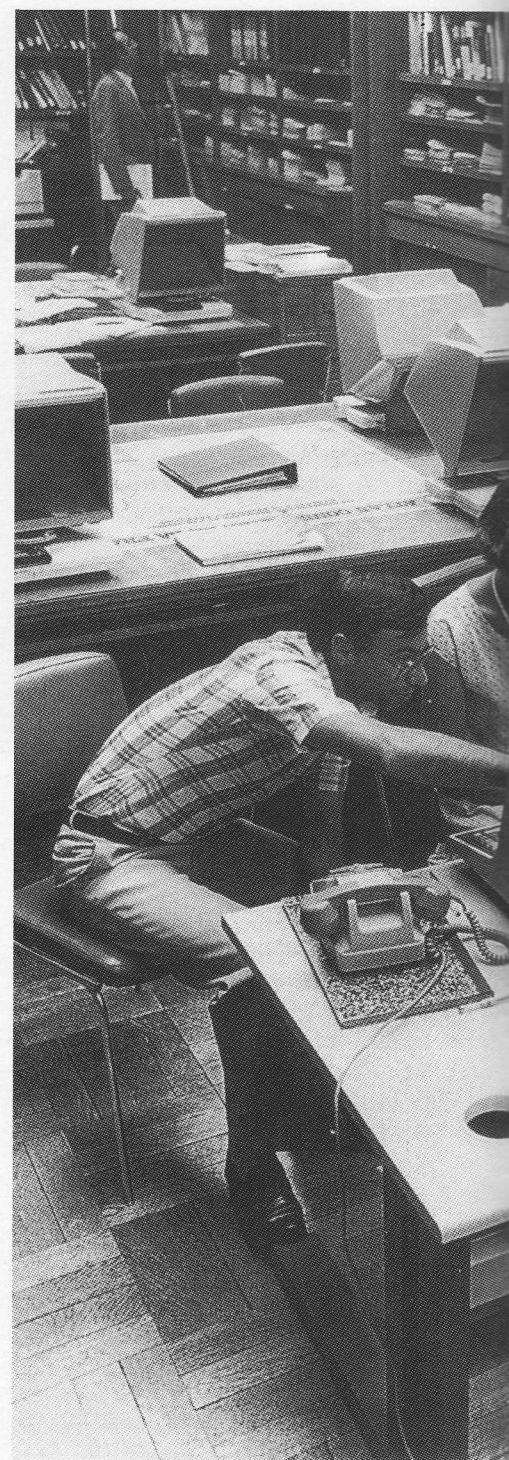
Nato nell'aprile 1990 con lo scopo di offrire all'utente finale uno strumento di accesso al catalogo SBN al tempo stesso più semplice e potente di quello allora messo a disposizione dalle funzioni di interrogazione del software in uso alle biblioteche della base delle università lombarde, in attesa del rilascio dell'OPAC (On line Public Access Catalogue), l'information retrieval BIR sviluppato dal CILEA (Consorzio interuniversitario lombardo per l'elaborazione automatica) con la collaborazione della Unisys (produttrice dell'elaboratore utilizzato per la gestione del prototipo SBN della base delle università lombarde) ha assunto nel tempo una connotazione sempre più autonoma da SBN determinando la nascita di una vera e propria base di dati bibliografici delle università. Cerchiamo di capire dalle parole di Fabio Valenziano, responsabile del progetto, la portata di questa iniziativa che, partita un po' in sordina, incontra un favore crescente da parte dei bibliotecari che vi partecipano andando a volte al di là delle intenzioni dei suoi ideatori.

Iniziamo con il chiedere a Fabio Valenziano quante e quali biblioteche partecipano al catalogo collettivo delle università padane.

Attualmente aderiscono al progetto le 26 biblioteche del polo SBN delle università lombarde (Università degli studi di Milano, Pavia, Brescia, Politecnico di Milano, IULM), quelle delle sedi di Milano, Brescia e Piacenza dell'Università Cattolica, la Biblioteca centrale della Bocconi di Milano, le biblioteche delle facoltà di economia e commercio e di lingue dell'Università di Bergamo. Inoltre le biblioteche delle facoltà di giurisprudenza, lettere e filosofia e scienze politiche, e quelle delle facoltà di ingegneria, di architettura, del CEDAR e di 11 dipartimenti del Politecnico di Milano hanno messo a disposizione della banca dati così creata i loro archivi pre-SBN realizzati con sistemi automatizzati diversi.

Qual è la consistenza quantitativa della banca dati che costituisce il catalogo collettivo delle università padane e quali sono i ritmi del suo incremento?

Circa 79.000 documenti nella banca dati delle biblioteche aderenti a SBN, 174.000 in quella cosiddetta delle università padane [i dati si riferiscono al novembre 1992]: mensilmente le biblioteche forniscono gli aggiornamenti del proprio archivio, anche se in molti casi gli aggiornamenti sono occa-



sionali, in virtù della disponibilità degli archivi di provenienza. Va sottolineato il fatto che, per come funziona il sistema oggi, ogni aggiornamento prevede che l'intero archivio sia ricaricato ex novo, in

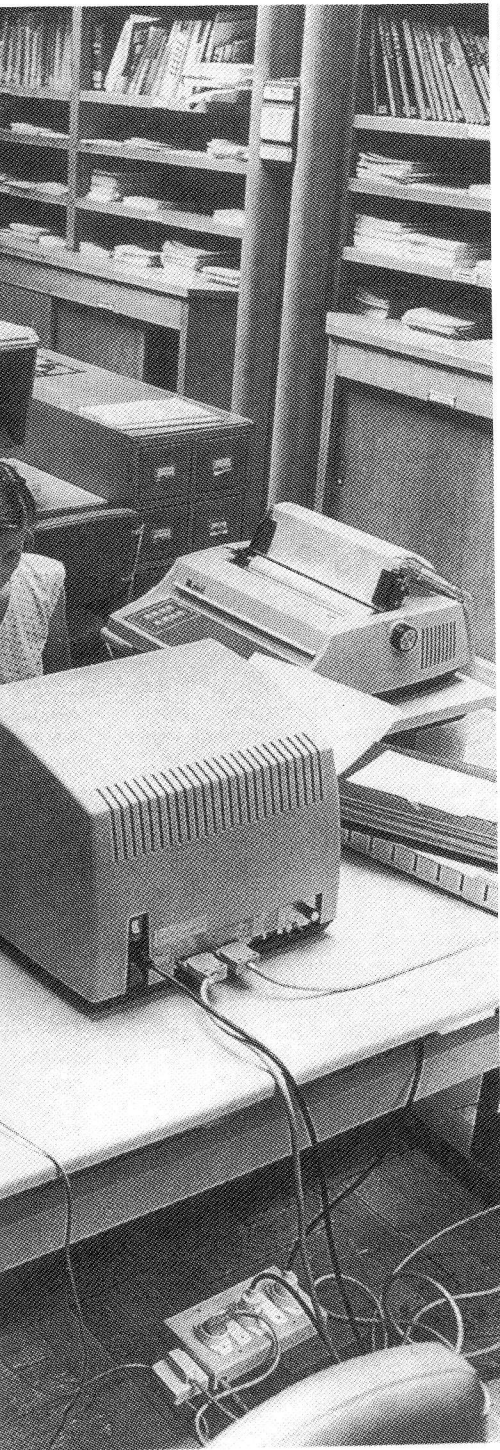


Foto MICHEL BARETTI

ad esempio, quali biblioteche hanno acquisito in tempi successivi un libro già presente in banca dati, perché posseduto da altre biblioteche. Inoltre ripetendo il caricamento complessivo si ottiene una indicizzazione più efficace dell'intero database.

Ci sembra di capire che i due archivi vengono tenuti distinti, e che l'utente finale per essere pressoché sicuro del risultato è costretto a ripetere la medesima ricerca in entrambe le banche dati: qual è il motivo di questa scelta, apparentemente poco vantaggiosa per l'utilizzatore?

Una delle caratteristiche principali del nostro sistema, accanto alla facilità d'uso, alla flessibilità e alla velocità di ricerca, è rappresentata dalla possibilità di raggruppare notizie bibliografiche provenienti da archivi a volte molto diversi tra loro. Ciò può costituire al tempo stesso un pericolo per chi interroga, in quanto non tutte le biblioteche utilizzano gli stessi criteri descrittivi: ad esempio non tutte le biblioteche effettuano la soggettazione dei loro libri, con il risultato che spesso i dati non sono omogenei. Per tali motivi si è ritenuta opportuna la creazione di due banche dati distinte, una delle quali (banca dati polo SBN) presenta una maggiore omogeneità dei dati, facilitandone quindi l'interrogazione.

Può illustrare brevemente ai lettori di "Biblioteche oggi" quali sono i dati caratteristici di ogni record bibliografico?

Il codice dell'archivio di provenienza, il codice identificativo della notizia bibliografica (in qualche caso preesistente nell'archivio di origine, in qualche altro creato ad hoc, che rappresenta comunque un dato univoco di corrispondenza fra record e archivi di provenienza), l'anno di stampa, la de-

scrizione bibliografica (ISBD), gli autori, i soggetti, i dati delle copie possedute (nome della biblioteca, numero di inventario, data di inventariazione, segnatura di collocazione). Di questi l'intera registrazione ISBD, il titolo (ricavato dall'ISBD), l'autore, il soggetto e l'anno di stampa sono utilizzabili come chiavi di ricerca.

Quelli che lei ha descritto sono i dati caratterizzanti i record bibliografici. Può ora descriverci quali sono gli elementi che caratterizzano lo strumento BIR (Bibliographic Information Retrieval) nei confronti di altri sistemi analoghi?

Abbiamo cercato di fornire un prodotto il più possibile user friendly, considerando che non si tratta di un OPAC. Nelle nostre intenzioni, all'utente che interroga il catalogo non è richiesto alcun tipo di addestramento, né l'utilizzo di manuali. L'interrogazione avviene in tre passi successivi. Innanzi tutto la scelta della chiave di ricerca (descrizione, autore o soggetto), quindi la formulazione della domanda, attraverso la semplice digitazione di una o più parole significative, ed infine l'esame dei documenti selezionati, con la visualizzazione dei dati relativi alle copie possedute dalle biblioteche.

Entrando un po' di più nel dettaglio sul funzionamento di questo catalogo bibliografico collettivo, può dirci se è previsto l'uso di operatori booleani, in modo più o meno consapevole da parte dell'utente?

Il sistema effettua automaticamente un *and* fra le parole digitate all'interno di una stessa chiave di ricerca. L'*or* è ottenibile attraverso la scrittura di una barra obliqua (/) tra due parole, mentre il *troncamento* può essere determinato digitando uno o più puntini ➤

quanto la semplice aggiunta dei record nuovi non permetterebbe la sostituzione dei vecchi record con eventuali versioni aggiornate. È facile comprendere come sia invece molto importante conoscere,

(...) finali per specificare una radice di parola (es. Ital... per ricerca-re Italia, Italiano, ecc.). È possibile inoltre affinare la ricerca, in caso di eccessivo rumore (viene sempre visualizzato il numero dei documenti selezionati), operando una ulteriore selezione all'interno dell'insieme di documenti individuato.

Ci sembra di capire che il ricorso al linguaggio naturale dovrebbe permettere all'utente finale di svolgere le sue ricerche in completa autonomia, senza che sia necessario l'intervento di personale specializzato della biblioteca: dall'esperienza sul campo le risulta che siano emerse critiche a questo modo di procedere, da parte dei bibliotecari? È stato quantificato in qualche modo il gradimento degli utenti a questo approccio?

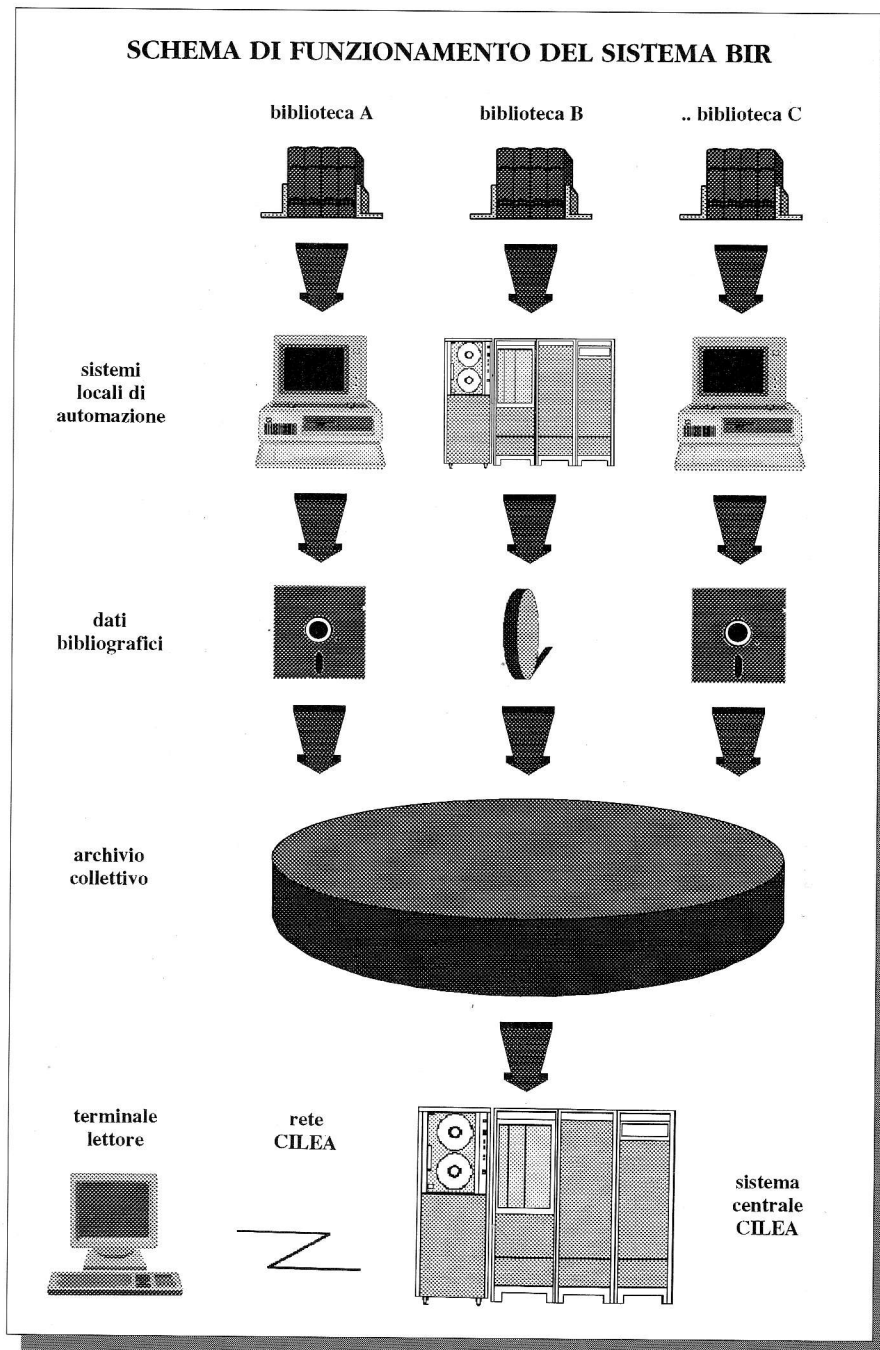
L'esperienza di BOMAS in uso al Politecnico di Milano, che utilizzando il sistema SEARCH richiede di esprimere la domanda utilizzando una certa sintassi (chiave, operatore di confronto, valore, booleano, ecc.), si era rivelata a nostro avviso inadeguata per un utente inesperto, e presupponeva una certa familiarità col linguaggio di interrogazione da parte dell'utilizzatore, o in alternativa, la necessità del supporto di personale specializzato. Di qui l'esigenza di realizzare una interfaccia semplificata, come l'attuale che, almeno a giudicare dalle scarse richieste di spiegazioni che ci provengono da tutti coloro che consultano BIR da tutta Italia, ci sembra riscuotere un certo favore, proprio in virtù della sua semplicità. Certo la migliore risposta a questa domanda la potranno dare i bibliotecari che hanno scelto di mettere lo strumento BIR a disposizione dei loro utenti.

Sono attualmente allo studio modifiche sostanziali al software di

interrogazione del Catalogo bibliografico, anche in relazione all'evoluzione tecnologica dei mezzi hardware e software a disposizione?

Per rispondere a questa domanda è necessario fare un passo indietro che consenta di capire quali sono le caratteristiche di BIR, che

ne fanno uno strumento versatile, ma con alcuni grossi limiti. Innanzitutto si deve considerare che si tratta di un sistema su mainframe, utilizzabile da terminali "stupidi", con caratteristiche di funzionalità base dei terminali (es. non full screen, senza possibilità di utilizzare sottolineature, doppia



intensità, ecc.), che rendono l'interfaccia abbastanza grezzo. D'altro canto proprio scelte tecniche di questo tipo rendono BIR utilizzabile senza grosse limitazioni e con costi molto contenuti da parte di un pubblico estremamente diffuso.

Paradossalmente per migliorare

di creare delle funzionalità specifiche per la creazione di file e la loro cattura su floppy, via file transfer, ma già ora se si dispone di determinati emulatori è possibile effettuare agevolmente lo scambio di dati fra host e personal computer. Ritengo però che questa problematica debba essere at-

al CILEA, sono sicuramente inferiori a quelli legati ai CD-ROM. Stiamo poi studiando delle prospettive di evoluzione che non si concilierebbero con la soluzione CD-ROM, e ci riferiamo alla eventualità di abbinare l'interrogazione del catalogo con un sistema di posta elettronica per la richiesta delle opere o lo scambio di messaggi tra le biblioteche.

Può illustrare brevemente come si configura il Catalogo bibliografico collettivo all'interno della complessa situazione universitaria italiana? È possibile collegarsi, ad esempio, con altre università non padane?

Quasi tutte le sedi universitarie, come è noto, sono collegate tra loro attraverso la rete GARR (Gruppo armonizzazione reti per la ricerca) con dorsale ad alta velocità (2 Mbit/sec), promossa dal Ministero dell'università e della ricerca scientifica e tecnologica. A sua volta la rete GARR è collegata con le principali reti di ricerca internazionali (EARN/BITNET, INTERNET/NSFNET, HEPNET, SPAN). Il CILEA è uno dei nodi di questa dorsale ed è quindi raggiungibile con prestazioni ottimali da qualsiasi sede universitaria collegata. Utilizzando quindi le appropriate modalità di collegamento, è possibile accedere al Catalogo bibliografico collettivo senza spese supplementari, in quanto il sistema BIR non richiede requisiti particolari per quanto riguarda il terminale utilizzato. Tutte le biblioteche dotate di terminali collegati a elaboratori in rete (ad esempio VAX in DECnet, IBM in SNA) sono nelle condizioni di accedere facilmente al servizio: l'elaboratore Unisys 2200 del CILEA è raggiungibile con collegamento SNA alla rete CILEA (es. via rete GARR) con applicazione di nome ITCLDEM2 e via INTERNET (131.175.1.7 comando di apertura sessione: ATTACH DEM, 9). A que- ➤

sostanzialmente BIR si dovrebbe sostituirlo integralmente con un altro prodotto, magari più moderno, ma sicuramente più costoso. Ciò che mi preme sottolineare è che il progetto del Catalogo bibliografico collettivo non va necessariamente messo in relazione con lo strumento di information retrieval utilizzato: l'idea di raccogliere in un unico database dati bibliografici provenienti da archivi preesistenti, per metterli a disposizione di un'utenza sempre più vasta, manterrebbe la sua validità comunque.

E per quanto riguarda il downloading dei dati, si prevede l'attivazione di procedure di stampa (su carta e su floppy), al fine di permettere all'utente finale di ottenere vere e proprie bibliografie?

Stiamo esaminando la possibilità

tentamente considerata dalle singole biblioteche, in quanto introduce nuovi servizi per il pubblico che vanno gestiti nel loro complesso, senza sottovalutare gli aspetti organizzativi, di tariffazione e di sicurezza.

Molte basi di dati on line vengono oggi giorno riprodotte su CD-ROM, supporto che sta ottenendo un crescente successo anche nelle biblioteche italiane: è pensabile una soluzione simile per il futuro del Catalogo bibliografico collettivo?

Non ritengo vantaggiosa, almeno per il momento, una ipotesi di questo tipo, in quanto la nostra banca dati è in continua evoluzione, e sicuramente l'on line ne rispecchia al meglio le caratteristiche. Non si deve dimenticare poi che i costi per il collegamento on line, per le biblioteche collegate



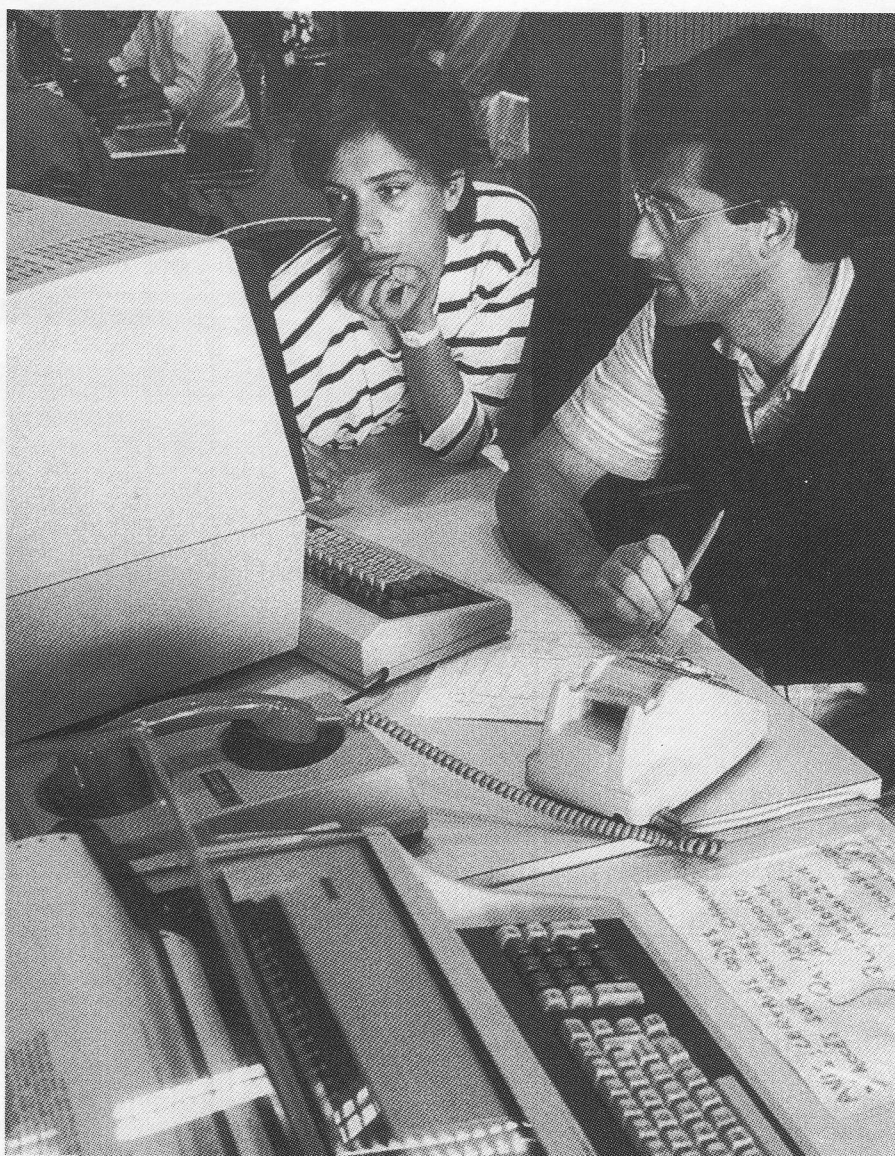
sto proposito ricordo come sia possibile effettuare gratuitamente prove di interrogazione del catalogo usando il codice userid/password: UNIPAD/CILEA.

Passando a concetti più generali, secondo il suo parere, la proliferazione di reti locali, che sta investendo anche le biblioteche, può proporsi come alternativa alla tradizionale fruizione delle informazioni residenti su host remoti, di cui BIR è un esempio, o al contrario facilita il processo di integrazione fra l'anarchia del mondo dei personal computer e la rigida gerarchia dei sistemi centrali?

La rete locale dal punto di vista tecnico può rappresentare una soluzione più funzionale per consentire il dialogo con l'host, ma ritengo che quando ci si misura su ordini di grandezza di un certo livello, la gestione su mainframe offra ancora le maggiori garanzie.

In più di un'occasione BIR è stato presentato come qualche cosa di "provvisorio", una specie di surrogato di alcune funzionalità di SBN nel campo del recupero dell'informazione, non ancora realizzate in modo adeguato. Non crede che l'attuale sviluppo di questo strumento possa dare torto a questa visione "di minima"?

In effetti lo scopo di BIR doveva essere quello di permettere una interrogazione semplificata del catalogo SBN, in attesa di OPAC. Ritengo peraltro che BIR possa continuare a esercitare la sua funzione anche in presenza di OPAC, non fosse altro che per la disponibilità di dati provenienti da archivi non SBN, per la sua accessibilità anche per le biblioteche non aderenti a SBN, e infine per il fatto di non richiedere workstation dalle caratteristiche hardware particolari (non dimentichiamo che OPAC presuppone un software sul personal computer impedendo la fruizione



MAGNUM PHOTO MARTIN FRANCK

del catalogo a utenti occasionali, non dotati di tale software).

A nostro avviso il Catalogo bibliografico collettivo può rappresentare per certi aspetti un superamento della storica contrapposizione, a volte con caratteristiche quasi da "guerra di religione", fra sostenitori ed avversari di SBN, proprio in virtù della messa in comune di dati bibliografici provenienti da biblioteche aderenti a SBN e non: non ritiene che, soprattutto per realtà bibliotecarie limitate sia dal punto di vista del personale, che

da quello delle risorse economiche a disposizione, BIR offra una soluzione più conveniente rispetto a un'adesione formale a SBN, permettendo loro di mantenere in vita sistemi di automazione sviluppati localmente?

Non vorrei essere stato frainteso a questo riguardo: indubbiamente la scelta giusta è quella di SBN, per una serie di ragioni che non sta a me evidenziare. BIR potrebbe essere considerato una sorta di "ultima spiaggia" per chi non può o non vuole aderire a SBN, senza per questo proporsi assolutamente co-

me alternativa a esso. In ogni caso la scelta BIR rappresenta un primo passo verso una logica di cooperazione fra le biblioteche, sotto forma di messa in comune dei propri archivi bibliografici.

Per non limitarci a evidenziare i soli lati positivi del progetto, non è prevedibile a lungo andare una ingestibilità di un database alimentato da archivi così differenti tra loro? Non pensa che questo possa rappresentare, a lungo andare, un problema?

Il rischio è proprio questo: se pensiamo che la stessa opera catalogata in 100 archivi risulterà 100 volte nel database, possiamo facilmente individuare i problemi che ciò comporterà per chi interroga. A questo proposito stiamo esaminando diverse soluzioni al fine di accorpare notizie bibliografiche identiche, anche se ciò non è semplice, anche perché la solu-

zione più corretta sarebbe quella di farlo al momento della nascita della notizia, nell'archivio di partenza, come accade in SBN. Non dimentichiamo che il Catalogo bibliografico collettivo non è altro che il cumulo di diverse banche dati, ma la validità dei dati è nell'archivio a monte. Un'ulteriore lacuna del nostro software è rappresentata dalla totale assenza dei rinvii, sia fra autori che fra soggetti, che limita fortemente la qualità della ricerca: per il discorso fatto prima, è da valutare l'opportunità di investimenti migliorativi in questa direzione, soprattutto in considerazione dell'avvento di OPAC e dell'indice SBN.

Molto spesso iniziative di portata limitata, caratterizzate da un'estrema versatilità, proprio in quanto sviluppate localmente, finiscono col perdere i loro indubbi vantaggi nel momento in cui si in-

seriscono in processi pesantemente burocratici: non crede che il Catalogo bibliografico collettivo possa correre questo rischio?

Fino a ora abbiamo evitato questa eventualità, e penso che se BIR si manterrà nell'ambito in cui è nato, saremo in grado di continuare a lavorare con una certa agilità: è fuori di dubbio però che alcune innovazioni, come ad esempio quella della posta elettronica finalizzata alle richieste di prestito fra biblioteche, potranno avere necessità di un maggiore approfondimento collegiale, che comporterà necessariamente qualche rallentamento. Ritengo comunque che sia un rischio da correre, perché sarebbe un peccato non sfruttare le potenzialità offerte dal collegamento on line, soprattutto potendo contare su una fruttuosa base di lavoro come quella garantita dal CILEA e dalle università lombarde. ■