

L'indicizzazione per soggetto in SBN

Maria Lucia Di Geso, ICCU

Introduzione

Nel corso degli ultimi dieci anni il Servizio bibliotecario nazionale (SBN) si è venuto configurando come la più grande rete italiana di servizi di informazione, localizzazione e disponibilità dei documenti, basata sulla cooperazione e la condivisione delle risorse bibliografiche fra migliaia di biblioteche sparse in tutto il paese: obiettivi presenti fin dall'inizio nella mente di Angela Vinay, primo direttore dell'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU). Nel processo che ha portato alla realizzazione di tali obiettivi, che tuttora non possiamo considerare pienamente conclusi, vi è stata una lunga fase nella quale SBN si è identificato, almeno a livello di Indice, quasi esclusivamente con i problemi di catalogazione per autore e descrittiva. Presupposto allo sviluppo dei servizi e della cooperazione è stato, infatti, la realizzazione della base dati Indice e il suo incremento mediante la catalogazione partecipata, aspetti sui quali necessariamente, in una prima fase, si è dovuta focalizzare l'attenzione. Nel corso del tempo, all'obiettivo della realizzazione della rete si è aggiunto quello di offrire al mondo bibliotecario e all'utenza una base dati non solo quantitativamente, ma anche qualitativamente significativa, prevedendo lo sviluppo in Indice di funzionalità di monitoraggio e "pulizia" dei dati, eliminazione di duplicazioni e, relativamente all'archivio autori, la creazione e la gestione di un *authority file*.

Per anni, quindi, il discorso, almeno a livello di Indice, si è limitato ai problemi della catalogazione per autore, essendo la catalogazione semantica esclusa dalle attività della cooperazione. In una fase nella quale il problema della creazione di una base dati coerente, di qualità e senza eccessive duplicazioni aveva la prevalenza, sembrò prematuro affrontare il discorso della catalogazione semantica, considerando anche le difficoltà di perseguire nel campo della soggettazione, quella uniformità e standardizzazione che si può ragionevolmente, anche se con difficoltà, ottenere nella catalogazione per autore.

L'Indice SBN

La catalogazione partecipata in Indice non contemplava le funzionalità relative all'aspetto semantico, né prevedeva la condivisione logica delle informazioni relative al contenuto concettuale dei documenti: si decise di non vincolare le biblioteche aderenti alla rete all'adozione di strumenti condivisi, in base al principio della specificità e della diversa tipologia di istituti. In fase di analisi dell'Indice si definirono pertanto come funzioni "protette" solo quelle relative alle procedure di catalogazione per autore/titolo (gestione bibliografica) e di prestito, mentre per l'area semantica fu lasciata ampia autonomia alle implementazioni dei *software* di Polo. Questa impostazione ha fatto sì che le procedure di soggettazione e classificazione dei documenti si presentassero difformi nei *software* di Polo sia rispetto ai dati che alle funzioni, soprattutto nella parte relativa alla costruzione dell'archivio dei termini e alle relazioni tra i termini stessi.

Da tali presupposti derivò che, quando agli inizi degli anni Novanta, con le attività di "migrazione" dei diversi Poli si realizzò il Sistema Rete - Indice SBN, si fece la scelta di riversare nell'Indice solo i dati relativi alla catalogazione descrittiva (autori e relative descrizioni bibliografiche), escludendo le informazioni di tipo semantico (soggetti e simboli di classificazione).

Il Gruppo di lavoro sui soggetti e le classi in SBN

L'assenza di dati semantici nell'Indice e di direttive comuni in ambito SBN non impedì che si realizzasse comunque un processo di omogeneizzazione anche nel settore della indicizzazione per soggetto, sia mediante una diffusione sempre maggiore del *Soggettario delle biblioteche italiane*, sia attraverso iniziative volte a creare momenti comuni di confronto fra i bibliotecari che lavoravano nel settore della soggettazione/classificazione all'interno dei Poli.

Nacque così, nel febbraio del 1990, il Gruppo di lavoro sui soggetti e le classi, istituito dall'Istituto, su indicazione della Commissione paritetica di esperti per SBN, con il compito di mettere a confronto e stabilire la validità delle funzioni *software* offerte dai vari applicativi presenti relativamente alla gestione dei soggetti, nonché l'adeguatezza dei servizi offerti rispetto alle aspettative degli utenti. Il gruppo di lavoro fu costituito tenendo conto, nella sua composizione, della rappresentatività di tutte le applicazioni *software* e delle varie tipologie di biblioteche operanti. In base alle linee di lavoro suggerite dalla Commissione, il Gruppo predispose dapprima un questionario volto a raccogliere informazioni sull'uso degli strumenti di indicizzazione diffusi fra i Poli SBN e da questo risultò che l'80% delle biblioteche utilizzavano il *Soggettario* di Firenze come principale strumento di indicizzazione.

Successivamente, mediante lo studio delle funzioni di gestione soggetti e classi negli applicativi allora esistenti, BULL, UNISYS, IBM/ADABAS e IBM/SQL (che si avvale anche di un'analisi comparata dei dati relativi all'Archivio soggetti presenti nei quattro prototipi), si giunse al tentativo di definire uno "standard SBN", a livello di funzioni e di dati, per la soggettazione e la classificazione che, da un lato, fosse la *summa* delle realizzazioni ottimali presenti nelle ultime *release* dei vari *software*, e dall'altro, tenesse conto delle procedure e dei metodi raccomandati per il controllo terminologico e per stabilire i tipi di relazioni tra i termini di indicizzazione suggeriti dagli standard internazionali, quali la Norma UNI/ISO 2788.

Sulla base del confronto e delle esperienze maturate, il Gruppo di lavoro sui soggetti raccomandò che a livello di Polo fosse utilizzato un unico strumento di riferimento per l'indicizzazione, senza precludere la possibilità di gestire *a latere* altri strumenti di natura più specialistica. La proposta di una condivisione dei soggetti fra i vari Poli, pur favorendo indubbiamente una razionalizzazione ed una normalizzazione nella pratica di catalogazione semantica in SBN, fu ritenuta invece prematura in quella fase, coincidente con l'avvio della rete: infatti avrebbe dovuto prevedere anche la creazione di liste di controllo centralizzate e una complessa organizzazione di *authority control* per il mantenimento e la coerenza dell'intero processo. Il Gruppo si espresse pertanto per un'ipotesi più ridotta, che prevedeva da un lato la possibilità di un caricamento *offline* in Indice dei soggetti e delle classi della sola Biblioteca nazionale centrale di Firenze (BNCF), da effettuarsi con aggiornamento periodico; dall'altro, funzioni di visualizzazione di soggetti/classi legati a notizie della BNCF e possibilità di memorizzare in un file di appoggio i dati da utilizzare successivamente a livello locale, con la creazione del relativo legame con la descrizione del documento.

Tale soluzione, anche se parziale, intendeva dare una risposta alle continue richieste di utilizzare le funzionalità di Indice per catturare, insieme alle singole notizie bibliografiche, anche i dati semantici ad esse collegati, con notevole risparmio di tempi e di costi per le biblioteche che intendessero utilizzare i soggetti attribuiti dalla Biblioteca nazionale centrale di Firenze; inoltre, la presenza di tali informazioni in Indice avrebbe offerto ai catalogatori una forma di "riscontro" e una guida per la formulazione delle voci di soggetto.

Lo sviluppo dell'OPAC e la creazione dell'archivio soggetti e classi nell'Indice SBN

L'ipotesi del Gruppo di lavoro sui soggetti si poté concretizzare solo alcuni anni dopo, nel 1996/97, grazie al progetto Accessibilità delle basi dati residenti sull'Indice SBN, finalizzato alla creazione dell'OPAC di Indice; la realizzazione di quest'ultimo rappresenta un'importante tappa

nell'evoluzione di SBN in direzione dello sviluppo di quella rete di servizi che costituisce l'obiettivo principale di tutto il progetto. La decisione di arricchire dei dati semantici le notizie residenti nell'Indice fu presa dal Comitato di gestione SBN su proposta dell'ICCU: relativamente alla base dati Libro moderno si pensò di sviluppare nell'OPAC delle funzionalità che permettessero la ricerca semantica dei documenti mediante l'interrogazione per soggetto e/o per classe. A tale obiettivo si aggiunse quello di offrire un supporto nella catalogazione semantica all'utente bibliotecario mediante il riversamento di queste informazioni anche nel database gestionale dell'Indice.

Nell'ambito del progetto furono individuate due fasi:

- una prima fase di costituzione della base dati dei soggetti e delle classi di Indice e una sua alimentazione mediante riversamenti batch dei soggetti/classi di 11 Poli;
- una fase successiva, in cui l'alimentazione dell'Indice sarebbe stata effettuata mediante l'ausilio di procedure *online*, da sviluppare presso gli stessi Poli individuati nella prima fase.

Per garantire omogeneità e coerenza alla costituenda base dati si stabilì di inserire solo soggetti creati dai Poli che utilizzavano il *Soggettario* delle biblioteche italiane; una volta data la priorità al Polo della BNCF, la scelta degli altri 10 Poli si basò su due principi: la qualità dei dati e il livello di "copertura" dell'Indice, intesa come percentuale di notizie possedute da un Polo rispetto all'intera base dati nazionale.

Questa prima fase si concluse nel febbraio del 1997, con il riversamento in Indice di circa un milione di soggetti e con circa 860.000 titoli soggettati; successivamente venne effettuato il riversamento dei dati nell'OPAC SBN.

La seconda fase è stata quella che ha interessato lo sviluppo di funzionalità di Polo sia per l'alimentazione che per la cattura dei dati semantici di Indice; dopo l'applicativo BULL, progressivamente anche altri *software* si sono dotati di tali funzionalità, come ad esempio gli applicativi IBM/ADABAS e UNIX in architettura *client server*.

La struttura dei dati e le funzioni previste in Indice presentavano, e presentano tuttora, caratteristiche diverse rispetto a quelle di Polo: il soggetto è registrato come un legame alla notizia bibliografica, come una sua "appendice"; ciò determina una duplicazione di stringhe di soggetto uguali, quando queste vengono inviate all'Indice da Poli diversi e quindi con identificativi (CID) diversi; dal punto di vista delle funzionalità, la scelta è stata quella di rinunciare a gestire l'archivio dei soggetti secondo il principio della catalogazione partecipata e pertanto:

1. non esiste una funzione di interrogazione soggetto, ma il soggetto si vede solo in fase di interrogazione/cattura a partire dalla notizia alla quale è collegato;
2. in fase di cattura di una notizia è lasciata ai Poli la massima libertà di catturare o meno il soggetto ad essa collegato;
3. i Poli che alimentano l'Indice con i loro soggetti hanno a disposizione, in relazione al *software* sviluppato, funzioni *batch-online* di invio cumulativo di dati alla fine della giornata di lavoro, oppure funzioni direttamente *online* che permettono anche una selezione dei soggetti da inviare;
4. si è esclusa la possibilità di legare ad una stessa notizia soggetti o classi provenienti da più Poli: una notizia presenta sempre i soggetti e la classe di un unico Polo;
5. nel caso di "concorrenza" di dati provenienti da più Poli, l'Indice sceglie secondo una scala di priorità dei Poli nella catalogazione semantica, che vede la BNCF al primo posto;
6. a fronte di una correzione o di una sostituzione di una stringa di soggetto i Poli non ricevono allineamenti; può accadere, pertanto, che un Polo che abbia catturato un soggetto successivamente modificato in Indice rimanga disallineato.

Situazione attuale

L'ultimo caricamento totale di soggetti e classi nell'Indice è stato eseguito nell'ottobre 2001; probabilmente esso sarà l'ultimo riversamento *batch*, poiché ormai quasi tutti i *software* SBN sono dotati di procedure che permettono l'alimentazione *online*.

Attualmente in Indice sono presenti circa 5.585.000 descrizioni di documenti¹, di cui oltre 5.124.000 sono relativi a monografie, 3.150.000 delle quali edite in Italia.

Sul totale delle notizie presenti 1.280.000 sono soggettate e alcune di queste presentano più di un soggetto, come dimostra la presenza di 1.610.000 legami notizia/soggetto.

In termini assoluti, la percentuale di notizie soggettate in Indice non arriva al 23%; se tuttavia teniamo presente che molte pubblicazioni non vengono soggettate perché di fatto non hanno un soggetto (come ad esempio la narrativa, i classici, le enciclopedie generali, ecc.) e che molte notizie provenienti da recuperi probabilmente non saranno mai soggettate, potremmo forse ridurre le notizie di Indice effettivamente soggettabili ad una cifra che si aggira fra i 4.500.000 - 5.000.000; in tal caso la percentuale di notizie soggettate oscillerebbe dal 28,4 al 25,6% dell'archivio.

Di contro, da indagini effettuate, si è visto che, man mano che gli applicativi di Polo si dotavano delle funzionalità di cattura dei dati semantici di Indice, vi è stato un progressivo aumento delle attività di cattura di soggetti: attualmente, a fronte di circa 6.500 notizie catturate al giorno, si effettuano dalle 2.700 alle 3.000 transazioni di interrogazione soggetto per cattura e si è passati dalle circa 1.000 catture di soggetti al giorno dei primi anni alle attuali cifre, con un incremento del 270-300%.

Dall'analisi, anche se sommaria, dei dati riportati emergono alcune considerazioni: dal primo riversamento dei soggetti in Indice, avvenuto all'inizio del 1997, sono ormai passati sei anni; allora le pubblicazioni presenti ammontavano a circa 2.800.000 ed ora, come si è detto, sono 5.585.000; i Poli che allora erano 33, attualmente sono 51: abbiamo dunque un Indice che è quasi raddoppiato nell'arco di tale tempo. Appare quindi evidente che la quantità di notizie con soggetto è esigua rispetto ad un archivio così consistente in quanto circa un quarto della base dati si presenta soggettata; d'altro canto si è visto che la cooperazione si è avvalsa abbondantemente di queste nuove funzionalità per facilitare e velocizzare le attività di catalogazione dei Poli, che hanno dimostrato un grande interesse a catturare i soggetti: oggi, per quanto riguarda il materiale corrente, circa il 50% delle notizie catturate vengono catturate con il soggetto e sicuramente tale percentuale salirebbe se in Indice vi fosse una maggiore quantità di soggetti.

Se si intende avere una maggiore copertura dell'archivio soggetti e favorire ulteriormente la cooperazione in tale settore, occorre rivedere in parte le scelte iniziali, eventualmente estendendo la funzionalità di catalogazione semantica in Indice ad altri Poli; i criteri di selezione potrebbero essere gli stessi che hanno improntato la scelta dei precedenti undici: l'utilizzo del *Soggettario* e una consolidata e corretta pratica nell'indicizzazione dei documenti; i nuovi Poli potrebbero essere identificati in base al loro contributo nella soggettazione di documenti che non sono presenti in BNI, come ad esempio il materiale straniero, quello estremamente specialistico o quello locale; in riferimento a quest'ultimo tipo di materiale va ricordato che alcune pubblicazioni in Indice non avranno mai soggetto se non le soggetta l'unico Polo che le possiede.

Inoltre, soprattutto fra i Poli che presentano una maggiore omogeneità nella soggettazione, come ad esempio quelli delle due Biblioteche nazionali centrali e sicuramente anche altri, si potrebbe pensare ad una ripartizione delle pubblicazioni da soggettare, dividendo il lavoro per tipologia di materiale o in base ad altri criteri che possono emergere solo da un confronto fra le realtà bibliotecarie.

Il progetto di Evoluzione dell'Indice SBN

¹ I dati di seguito riportati si riferiscono ad un'indagine effettuata il 15 Dicembre 2002

Il progetto dell'Evolutione Indice, nel quale l'ICCU è impegnato da circa due anni, è nato dall'esigenza di far evolvere SBN verso un nuovo sistema, che diventi riferimento esteso per l'erogazione di servizi diffusi.

A tale scopo il progetto prevede un arricchimento dei servizi, integrazioni ed estensioni del database e l'apertura a nuove tipologie di utenza con profili diversificati.

Oltre a garantire – per i sistemi attualmente in essere – il mantenimento delle regole già attive nell'attuale SBN, il futuro sistema consentirà:

- un'estensione dei servizi forniti dal catalogo centrale;
- un'estensione della cooperazione SBN ad altre tipologie di materiale;
- una semplificazione delle interazioni fra sistema centrale e Polo nella nuova catalogazione partecipata (anche tramite l'introduzione di meccanismi di catalogazione derivata);
- l'apertura verso altri LMS tramite formati di scambio internazionali (*import-export di descrizioni bibliografiche*).

Nell'ambito del progetto sono state previste anche nuove funzionalità e nuovi dati relativi all'area della catalogazione semantica.

Nella nuova struttura della base dati verranno implementati un archivio dei soggetti e un archivio delle classi che andranno ad affiancare gli altri due archivi dei titoli e degli autori: il soggetto non sarà più una "appendice" della notizia bibliografica, cioè un'informazione visibile soltanto a partire dalla visualizzazione analitica di un titolo, come si verifica attualmente, ma un'entità autonoma, interrogabile separatamente.

Tale struttura, oltre alle funzionalità già esistenti di creazione/correzione e cattura dei soggetti e delle classi in Indice, attivabili già con l'attuale protocollo SBN, dovrà permettere la realizzazione delle seguenti nuove funzionalità, a disposizione dei Poli che colloquieranno con il nuovo protocollo SBN MARC:

1. interrogazione Indice per dati semantici, soggetti e simboli di classificazione, utilizzati sia come canali di accesso diretto agli archivi dei soggetti e delle classi, sia come canali di accesso alle notizie;
2. "navigazione" dall'archivio dei soggetti a quello dei titoli e viceversa;
3. attività di "monitoraggio" dell'archivio dei soggetti, volte a rilevare la presenza di eventuali "duplicati", squadrature, errori di digitazione, ecc.;
4. funzionalità di "pulizia" dell'archivio, le cui attività dipenderanno dall'esito e dai risultati ottenuti con l'attività di monitoraggio e finalizzate alla creazione di una base dati coerente;
5. statistiche sulle attività e la produzione dei Poli relativamente all'area semantica;
6. estensione delle funzioni di catalogazione semantica alla nuova tipologia di materiale, diverso da quello librario, che l'Indice si predispone a gestire (ad es. stampe, foto, carte, ecc.).

Lo Studio di fattibilità del nuovo *Soggettario* delle biblioteche italiane

Quasi contemporaneamente al progetto di Evoluzione dell'Indice SBN, si è sviluppato il progetto della Biblioteca nazionale centrale di Firenze relativo allo Studio di fattibilità sul rinnovamento del *Soggettario* per i cataloghi delle biblioteche italiane. L'ICCU ha accolto con interesse l'invito della BNCf a partecipare alle riunioni di lavoro del Gruppo che ha curato lo studio e ne ha condiviso le finalità: la scelta di giungere alla costruzione di un nuovo linguaggio documentario coerente, organico e costruito sulla base di poche regole chiare, semplici e di facile applicabilità risponde pienamente alle esigenze di rinnovamento espresse più volte e da più parti nel mondo bibliotecario e in ambito SBN; i tre principi ai quali si ispira il nuovo linguaggio, quello dell'unità, della

predittività e della specificità, sono entrati ormai a far parte del bagaglio culturale di chi si occupa di indicizzazione; ipotesi valide ci sono sembrate quelle relative al principio della coestensione della stringa, all'impostazione di un ordine di citazione non più fondato sulla struttura voce principale/suddivisione, ma sull'analisi dei ruoli sintattici. Inoltre, il richiamo all'applicazione della norma UNI/ISO 5963 per la fase dell'analisi concettuale dei documenti e, per l'aspetto relativo al controllo del vocabolario, il richiamo all'altra norma ISO 2788-1986 vanno in direzione di quell'attività di diffusione degli standard internazionali che l'ICCU promuove da anni. La realizzazione di tale progetto sembra andare dunque verso la creazione di un "archivio di autorità" nel campo della catalogazione semantica che potrà essere un ottimo strumento di ausilio per i Poli SBN.

Occorre tuttavia, a mio avviso, fare un ulteriore sforzo per poter capire se sia possibile arrivare ad un coordinamento fra i due progetti, individuando gli strumenti, le modalità e i livelli di integrazione del sistema "nuovo *Soggettario*" nell'ambito della cooperazione e del Sistema SBN.

Molto è il lavoro che ancora ci aspetta, in vista della realizzazione dei due progetti; ma l'ottimismo ci suggerisce che le due strade intraprese possano un giorno incontrarsi, in modo che anche l'Italia possa offrire in futuro alla comunità degli utenti e dei bibliotecari non solo una grande base dati bibliografica, ma un sistema articolato e complesso, aperto e dotato di strumenti qualitativamente significativi, come ad esempio un archivio di autorità per soggetto, come accade già nei paesi più evoluti.

L'Istituto centrale per il catalogo unico delle biblioteche italiane sta lavorando in questa direzione.