

Servizio bibliotecario nazionale e indicizzazione semantica: un decollo difficile, una rotta da condividere

Anna Lucarelli

È possibile pensare ad uno sviluppo efficace della catalogazione semantica nel nostro Servizio bibliotecario nazionale (SBN)? A trent'anni dal suo avvio, la situazione di questo versante della cooperazione è tuttora molto problematico, rappresenta un punto dolente che è indispensabile mettere a fuoco per delinearne un quadro realistico, formulare senza reticenze alcune considerazioni, ipotizzare obiettivi ed elaborare proposte realizzabili. L'analisi di quanto è accaduto (e non accaduto) può aiutarci ad individuare una nuova rotta, fatta di maggiore partecipazione, condivisione, coordinamento.

È particolarmente interessante e singolare affrontare questa tematica proprio a Firenze, nel convegno che festeggia i trent'anni di SBN, perché è noto a tutti il ruolo che la Biblioteca nazionale centrale di Firenze (BNCF) ha da sempre rivestito come laboratorio privilegiato nel campo della catalogazione semantica e dei suoi strumenti. Un viaggio iniziato sin dagli anni Venti del Novecento, con precisione nel 1925, quando la BNCF iniziò a corredare di soggetti le notizie del *Bollettino delle pubblicazioni italiane ricevute per diritto di stampa* e che è proseguito con l'esperienza pionieristica del *Soggettario per i cataloghi delle biblioteche italiane*, il cui gruppo di lavoro, coordinato negli anni finali da Emanuele Casamassima, portò alla nota pubblicazione del 1956. Un'avventura ricominciata grazie alla collaborazione con gli esperti del Gruppo di ricerca sull'indicizzazione per soggetto (GRIS), ed infine confluita nella progettazione ed allestimento del *Nuovo*

soggettario. Un lungo e affascinante percorso, affiancato dalla costante applicazione, e negli ultimi anni anche dalla cura scientifica, delle edizioni italiane della Classificazione decimale Dewey. Una storia, quella degli strumenti italiani di indicizzazione semantica, che ha tratto linfa dalla fiorentinità.

Ma se ci volessimo chiedere quale e quanta collaborazione sia stata realizzata su questo fronte nell'ambito di SBN, la più grande rete italiana di biblioteche, ci troveremmo di fronte a dati abbastanza sconfortanti. Contrariamente a quanto avvenuto per la catalogazione descrittiva, dobbiamo parlare di una cooperazione non decollata.

I punti critici e le difficoltà non da oggi lamentate appartengono a varie tipologie. Quelle maggiormente ricordate dagli addetti ai lavori sono legate a fattori indiscutibili: i costi dell'indicizzazione che le biblioteche devono sostenere a fronte di risorse umane sempre più ridotte; la complessità dell'analisi concettuale e del metodo (l'indicizzazione è “una pratica spinosa”, come amava dire Luigi Crocetti); la natura stessa degli strumenti, percepiti a volte come carenti, a volte come ostici; il loro periodico aggiornamento (ad esempio, il passaggio da un'edizione all'altra della DDC e dal vecchio al *Nuovo soggettario*) che impone conoscenza dei cambiamenti ma anche la necessità di una formazione professionale che dovrebbe essere continua e che invece è spesso insufficiente.

E così, in Italia, le biblioteche soggettano poco. E anche quando fanno si sforzano di

farlo per non tradire questo versante della catalogazione, non riescono a valorizzare, rendere visibile, condividere il proprio lavoro.

La stessa Biblioteca nazionale centrale di Firenze, malgrado sia la biblioteca italiana che dà il contributo più considerevole a SBN (sia per l'archivio delle classi che mette a disposizione, sia perché è di sua provenienza circa un quarto dei soggetti inviati all'Indice), ha nel proprio Opac una percentuale molto bassa di opere con soggetti e classi.

Può fare qualcosa il Servizio bibliotecario nazionale per trovare soluzioni a queste difficoltà?

Il contesto bibliografico/documentario in cui operiamo non ha minimamente svilito l'importanza dell'indicizzazione e il web semantico ha confermato il ruolo centrale che le biblioteche possono avere nell'universo dei *linked data* proprio attraverso le "entità" individuabili tramite l'indicizzazione. Sappiamo bene che sono andati in questa direzione modelli concettuali, principi, standard ed indirizzi prodotti e diffusi a livello internazionale. Karen Coyle, di recente, ha dedicato nel suo blog alcune pagine a questo tema¹.

Da tempo si discute dell'importanza di rendere aperti, più visibili e riusabili i nostri dati sui contenuti concettuali, rilasciandoli in formati standard compatibili con il web semantico (lo SKOS/rdf, nel caso dei thesauri). Già in altre sedi si è parlato del *Nuovo soggettario* e delle sue funzionalità per i *linked data*².

Non esiste Paese che abbia abbandonato

1 Karen Coyle, *Catalogs and Content: an Interlude (4 luglio 2016)* <http://kcoyle.blogspot.it/2016/07/catalogs-and-content-interlude.html?m=1>

2 Giovanni Bergamin, Anna Lucarelli, *Il Nuovo soggettario come servizio per il mondo dei linked data*, «JLIS.it», vol. 4, n. 1 (2013) <http://leo.cineca.it/index.php/jlis/article/view/5474>

l'indicizzazione; esistono semmai Paesi che vanno estendendola anche a risorse non bibliografiche, o che di tali strumenti stanno valorizzando il multilinguismo o ancora che hanno messo in cantiere progetti sperimentali, come quello della Deutsche Nationalbibliothek per l'indicizzazione automatica di risorse digitali³.

Alla luce di queste considerazioni, qual è, in concreto, la situazione del Servizio bibliotecario nazionale relativamente alla semantica?

Uno sguardo al passato

Volendo tracciare le linee fondamentali del percorso seguito sinora, dovremmo subito ricordare che, nel 1986, dunque alle origini di SBN, venne fatta la scelta, allora sicuramente molto saggia, di non imporre alle biblioteche una catalogazione semantica "partecipata" e di non obbligarle a condividere dati.

Le ragioni di questa scelta erano di vario tipo: tecniche, legate ai software di polo; politico-gestionali, dovute cioè alla varietà dei livelli di specializzazione (biblioteche con utenze ed esigenze informative diverse); relative agli strumenti allora disponibili: come avrebbe potuto il *Soggettario* del 1956, con il suo invecchiamento, soddisfare esigenze di una rete di biblioteche così diverse fra loro? E come si sarebbero potuti creare "allineamenti" fra notazioni Dewey, dal momento che alcune biblioteche usavano edizioni integrali, altre edizioni ridotte e soltanto pochissime usavano edizioni recenti? A queste, che forse non furono le uniche ragioni ma senz'altro furono le principali, si deve aggiungere il fattore

3La notizia del progetto tedesco è stata data alla Post-Conference IFLA Beyond libraries (Tallinn, 17-18 agosto 2012) e successivamente pubblicata in: Ulrike Junger, *Can indexing be automated? The example of the Deutsche Nationalbibliothek*, «Cataloging & Classification Quarterly», Volume 52, 2014 - Issue 1: Beyond Libraries: Subject Metadata in the Digital Environment and Semantic Web (<http://www.tandfonline.com/doi/abs/10.1080/01639374.2013.854127?journalCode=wccq20>)

della formazione e competenza professionale, estremamente varia e diseguale in quegli anni, fra gli addetti ai lavori.

Insomma, tutto faceva pensare che, sul fronte della semantica, si fosse ben lontani da quel grado di uniformità catalografica che invece i codici di catalogazione per autore e gli standard descrittivi (ISBD) avevano comunque permesso di raggiungere.

Così anche quando, nei primi anni Novanta, i poli iniziarono a inviare le loro notizie nell'Opac dell'Indice SBN, le informazioni semantiche rimasero escluse, nonostante l'80% circa delle biblioteche dichiarasse di usare lo stesso strumento, il *Soggettario*. Venne deciso che solo i soggetti della BNCf sarebbero stati caricati, periodicamente e offline⁴.

Una svolta ci fu nel 1997 quando, per l'Opac di Indice, venne implementata la ricerca per soggetti e classi, inizialmente basata su un archivio alimentato in *batch* da 11 poli SBN (scelti in base alla qualità dei loro dati e alla copertura) e, in una seconda fase, tramite procedure online. Ma cosa è realmente successo dopo quel primo riversamento del 1997, quando i poli, in totale, erano soltanto 33?

Il percorso è stato complesso, accidentato, per varie ragioni non efficace. Il numero di notizie in Indice corredate di soggetti e classi è stato sempre basso e, proporzionalmente, in decrescita.

⁴ Una ricostruzione analitica di queste fasi, è stata fatta nel 2003 da Maria Lucia Di Geso, dell'ICCU: M. L. Di Geso, *L'indicizzazione per soggetto in SBN*, in *Authority control: definizione ed esperienze internazionali : atti del convegno internazionale, Firenze, 10-12 febbraio 2003* / a cura di Mauro Guerrini e Barbara B. Tillett, con la collaborazione di Lucia Sardo. - Firenze : Firenze university press ; Roma : Associazione italiana biblioteche, 2003, p. 289-296 http://www.sba.unifi.it/ac/relazioni/digeso_ita.pdf

Nel 2003, sei anni dopo quella svolta, le notizie in Indice con soggetto erano il 23% del totale (una percentuale che potrebbe essere innalzata fino a circa il 28% se si considerasse che non tutto si soggetta). Le catture con i soggetti erano solo il 50%.

Nel già citato convegno del 2003, Maria Lucia Di Geso, dell'ICCU, si era espressa in modo molto chiaro: “Se si intende avere una maggiore copertura dell'archivio soggetti e favorire ulteriormente la cooperazione [...], occorre rivedere [...] le scelte iniziali, eventualmente estendendo la funzionalità di catalogazione semantica in Indice ad altri poli; [...] 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”.]. Un'analisi lucida e sempre attuale.

Erano gli anni in cui il progetto “Evoluzione Indice SBN” dell'ICCU avrebbe portato funzionalità nuove, con i soggetti diventati entità autonome e non più appendici della notizia bibliografica⁵.

A breve distanza di tempo ci sarebbero state altre novità importanti: nel 2007 il prototipo del *Nuovo soggettario* (con la *Bibliografia nazionale italiana* che iniziò ad usarlo sin da subito, facendo da rompiggiaccio); nel 2009 la pubblicazione della versione italiana della 22. edizione integrale della DDC.

⁵“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 colloquiaranno con il nuovo protocollo SBN MARC” (Maria Lucia Di Geso, cit.)

Sempre nel 2009, un Gruppo di lavoro dell'ICCU produsse un report dal quale risultò che la quantità di documenti soggetti rispetto al totale era scesa al 20%.

Il Gruppo, nel 2010, diffuse fra le biblioteche della rete un questionario ai cui macro-dati è interessante fare riferimento. Risposero 59 poli sui 70 di allora; 57 soggettavano (tanti, dunque), di questi, 13 tramite il *Nuovo soggettario*; 52 classificavano con varie edizioni della DDC; 15 poli inviavano i soggetti in Indice. Le biblioteche che soggettavano erano 1511, allora più del 50% delle biblioteche di SBN, ma ben 35 poli (su 59) non avevano un referente per la catalogazione semantica⁶.

Gli ultimi anni.

Non è semplice oggi fare un raffronto con i dati parametrizzati del 2009 e 2010; sapere esattamente chi soggetta, chi usa il *Nuovo soggettario* o chi ancora il *Soggettario* del 1956; chi classifica e come mai è così rallentata l'adozione della WebDewey italiana. Certamente sarebbe utile un'altra indagine, analoga a quella di qualche anno fa. Quello che si può dire è che, ad oggi (ottobre del 2016), i poli che inviano soggetti in Indice sono 43 (circa il 44% dei 98 totali), ben più dei 15 del 2010.

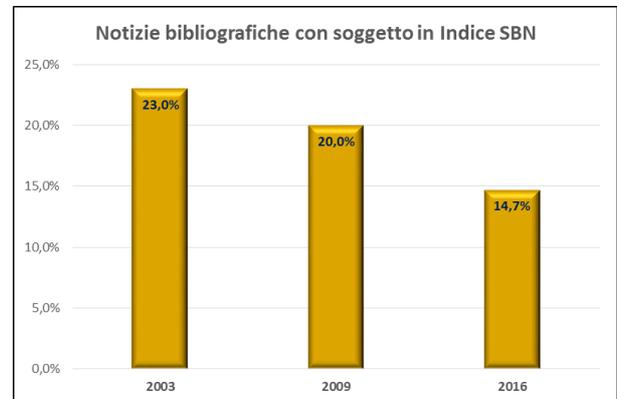
⁶ Il Gruppo di lavoro (il cui scopo era anche di individuare modalità che favorissero l'applicazione del *Nuovo soggettario* in SBN) era composto da: Maria Lucia Di Geso e Gabriella Contardi (ICCU); Giovanni Bergamin e Anna Lucarelli, (Biblioteca nazionale centrale Firenze); Carmela Santucci (Biblioteca nazionale centrale di Roma); Monica Ferrarini (Regione Emilia Romagna); Raffaella Gaddoni (Università degli studi di Bologna); Sara Legnaro (Università degli studi di Padova); Sergio Graffi (Regione Lombardia); Giulio Negretto (Regione Veneto):

http://www.iccu.sbn.it/opencms/opencms/it/main/attivita/gruppilav_commissioni/pagina_385.html.

Per il documento prodotto :

http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/Applicazione_del_Nuovo_Soggettario_in_SBN.pdf

Ma se volessimo calcolare la percentuale di notizie con soggetti, vedremmo che è ulteriormente calata: dal 23% del 2003 e dal 20% del 2009, si scende al 14,7% (13,2% quelle con classe)⁷.

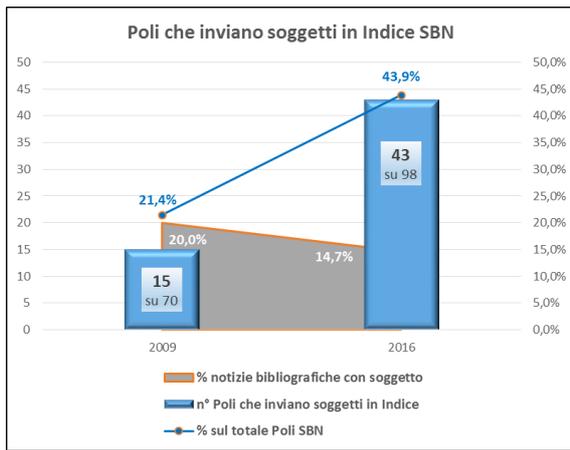


Il sottoinsieme delle opere pubblicate fra il 2007 (anno in cui viene messo a disposizione il prototipo del *Nuovo soggettario*) e il 2016 presenta soltanto il 13,8% di soggetti, dei quali il 19,2% elaborato con il nuovo strumento di indicizzazione. Un dato che deve comunque essere letto sia in relazione al totale indistinto delle notizie (un totale in cui sono comprese tipologie di opere che non verrebbero comunque soggettate), sia al grado di applicazione non immediata ma progressiva, da parte delle biblioteche, del *Nuovo soggettario*.

Delle attuali 5955 biblioteche di SBN, 775 inviano in Indice soggetti elaborati con il *Nuovo soggettario*. Ma il paradosso è che ce ne sono altre che, pur usandolo, non li inviano⁸.

⁷ Ringrazio Patrizia Martina dell'ICCU per avermi fornito dati dell'Indice SBN sulla situazione attuale

⁸ Interessante, a riguardo, la recente tesi di laurea discussa all'Università di Pisa (Corso di laurea in Scienze dei Beni culturali, anno accademico 2015-2016): Dario Mangione, *Nuovo soggettario: una prima indagine su alcune problematiche di applicazione nelle reti bibliotecarie toscane* (relatore prof. Franco Neri)



Da dove [ri]cominciare?

A fronte di queste criticità possiamo ricominciare da cose possibili, sfruttando ciò che la nostra comunità bibliotecaria è in grado di realizzare e puntando su quanto già abbiamo, dunque su procedure e pratiche uniformi, sulla condivisione di strumenti comuni di indicizzazione e di ricerca, sul potenziamento di visibilità e fruizione di dati aperti, su strategie che favoriscano l'accesso ai cataloghi anche da "porte esterne".

A quest'ultimo riguardo non si dimentichi che, da quando qualche anno fa sono stati implementati link reciproci fra il Thesaurus del *Nuovo soggettario* e Wikipedia nella versione in lingua italiana, le visite dell'Opac della BNCF hanno avuto un incremento di quasi il 30%, un dato che la dice lunga sulle opportunità che le biblioteche avrebbero di far conoscere e rendere maggiormente disponibili le proprie risorse, se lavorassero maggiormente su dati aperti e web semantico⁹.

Simonetta Buttò, attuale responsabile dell'ICCU, ha recentemente espresso l'intenzione dell'Istituto di lavorare ad un miglioramento e maggiore coordinamento delle attività legate all'indicizzazione

semantica del patrimonio documentario italiano. E una serie di attività, anche a livello tecnico-informatico, è già stata avviata. Ma sarà importante che anche le singole biblioteche, da parte loro, facciano scelte nella stessa direzione.

Per inaugurare un nuovo corso della catalogazione semantica all'interno del nostro Servizio bibliotecario nazionale, sarebbe utile innanzitutto che tutti i poli nominassero referenti per la soggettazione e classificazione, coordinati dall'ICCU, che si programmassero forme di indicizzazione con modalità anche diversificate (meno costose della pre-coordinazione), ad esempio, rinunciando in qualche caso a stringhe di soggetto e adottando almeno parole chiave, magari per certe tipologie di risorse (i periodici, il materiale grafico, i libri scolastici...).

Ma poco potrà migliorare se, su scala nazionale, le varie istituzioni non decideranno di abbandonare strumenti obsoleti, allestiti ormai molti anni fa con criteri invecchiati; il *Soggettario* del 1956 ha ancora motivo di essere usato (come avviene ancora in alcune biblioteche italiane) quando esiste il *Nuovo soggettario* aggiornato, aderente a standard internazionali, gratuito, disponibile con licenze libere, con dati esposti in rdf, scaricabile, riusabile? Uno strumento che ha collegamenti con Wikidata, che ha migliaia di equivalenti in inglese e in francese che lo fanno colloquiare con la Library of Congress e con la Bibliothèque nationale de France, favorendo peraltro diffusione e conoscenza dell'italiano all'estero, un aspetto che si incrocia singolarmente con l'iniziativa degli Stati Generali della lingua italiana nel mondo, promossa dal MIBACT.

Per quanto riguarda l'invio di soggetti in Indice e la visualizzazione dei dati semantici, un nuovo inizio non potrà darsi se non

⁹ Sul progetto di interoperabilità fra Wikipedia e risorse della BNCF, Anna Lucarelli, «Wikipedia loves libraries»: in *Italia è un amore corrisposto...*, «Aib studi», vol. 54 n. 2/3 (maggio/dicembre 2014), p. 241-259 (<http://aibstudi.aib.it/article/view/10108>)

rivedendo ed attualizzando, finalmente, quella lista dei poli ormai molto invecchiata e i requisiti che la dettavano. E non c'è dubbio che si dovrà dare la precedenza a chi garantisce qualità dei dati. Ci sono biblioteche italiane che negli ultimi anni sono diventate molto competenti nella soggettazione e classificazione e che non sono incluse in quell'elenco di vent'anni fa.

Ugualmente utile sarebbe progettare una spartizione di compiti fra biblioteche che potrebbero specializzarsi nella soggettazione e classificazione di particolari pubblicazioni (ad esempio, tesi di dottorato, periodici; risorse digitali, materiali grafici), di opere di particolari ambiti disciplinari (ad esempio, delle discipline giuridiche, delle “scienze dure”, oppure delle scienze religiose), di particolari generi editoriali (ad esempio, la manualistica, i libri di viaggio; certi tipi di narrativa).

Inoltre, per un Servizio bibliotecario nazionale più ricco, più inclusivo e aperto, dovremmo pensare - come avviene in altri Paesi - ad un ampliamento delle tipologie di risorse con soggetto e classe, includendo quelle digitali (la BNCF sta avviando una sperimentazione, in collaborazione con alcune università italiane), le fotografie (facendo tesoro delle esperienze importanti realizzate soprattutto in Emilia Romagna), i documenti archivistici (in BNCF, è stato recentemente soggettato l'archivio del Laboratorio di Restauro, nato a seguito dell'alluvione del 1966)¹⁰. In questo quadro non si dovrebbe

10 Questa problematica è al centro del lavoro condotto da un gruppo del MAB Toscana che sta sperimentando il *Nuovo soggettario* per la descrizione di oggetti di musei e di documenti archivistici. Sull'esperienza del MAB Toscana, sono stati pubblicati recentemente, a cura di Francesca Capetta, gli Atti del Convegno tenuto il giorno 8 ottobre 2015 presso la Soprintendenza archivistica per la Toscana, organizzato da ANAI Toscana: *Il nome delle cose, il linguaggio come punto d'incontro tra archivi, biblioteche e musei: l'esperienza del gruppo linguaggi di MAB Toscana*: http://www.ilmondodegliarchivi.org/images/Quaderni/MdA_Quaderni_n1.pdf

ignorare l'importanza della soggettazione di fondi antichi, un'attività consolidata in altri Paesi e sulla quale la BNCF, in collaborazione con la Biblioteca dell'Accademia della Crusca, sta predisponendo Linee guida che pubblicherà nel corso del 2017.

Un rinnovamento di questo tipo, richiede ovviamente investimenti sul versante tecnologico del catalogo di SBN che potrebbe implementare ulteriori campi di ricerca per vari livelli di informazioni semantiche dedicati, ad esempio, a sommari e a parole chiave, come già da tempo realizzato da WorldCat¹¹ o dal catalogo della Bibliothèque nationale de France¹².

Karen Coyle, nell'intervento già indicato, si pone la domanda, da lei ritenuta di grande importanza, su quali entità e relazioni il ricercatore sia interessato a incontrare nel suo percorso di ricerca e quali possano essere messe a disposizione dal catalogo (“*What should happen between the search box and the bibliographic display?*”). Una domanda del tutto coerente con l'enfasi che legittimamente si pone sui linked data.

Potenziare link fra l'Opac dell'Indice SBN e il Thesaurus del *Nuovo soggettario* e, viceversa, fra il Thesaurus e l'Opac SBN (così come già avviene per il catalogo BNCF), porterebbe, ad esempio, notevoli vantaggi agli utenti in fase di ricerca.

E non ci sono dubbi che le nuove modalità di produzione della *Bibliografia nazionale italiana* (BNI), in via di inaugurazione proprio in questi mesi, potranno aprire prospettive nuove anche per la cooperazione sui soggetti.

11 <https://www.worldcat.org/>

12 <http://catalogue.bnf.fr/index.do>

Nell'immediato, si potrà partire semplicemente dal potenziamento degli invii di soggetti e classi in Indice da parte di un numero maggiore di biblioteche. Durante recenti incontri di lavoro con colleghi dell'ICCU e della Biblioteca nazionale centrale di Roma, si è proprio stabilito di avviare il completamento dei riversamenti dei soggetti dei poli (la BNCF nel 2013 ha già inviato buona parte della sua base in Indice e da allora lo fa in tempo reale; la BNCR dovrà

farlo in toto) e parallelamente alla soluzione dei problemi tecnici di allineamento con soggetti già presenti, pregiudiziale anche per il corretto funzionamento di SBN web.

Con un percorso graduale e collaborativo ce la possiamo fare.

[Dati calcolati a dicembre 2016]