

Indicizzazione Semantica in Pillole

IFLA LRM: può contenere tracce di semantica

Pino Buizza

Grande risonanza ha avuto la pubblicazione da parte dell'IFLA del modello di riferimento per le biblioteche, *IFLA Library Reference Model (LRM)*¹, che unifica e sostituisce i precedenti modelli per i requisiti funzionali della famiglia FRBR (FRBR, FRAD, FRSAD). L'attenzione di chi se ne occupa in pubblicazioni, corsi e presentazioni è sempre rivolta agli aspetti descrittivi e nominali (semiotici) dell'informazione bibliografica, come è fin dall'inizio l'impostazione dei modelli concettuali dell'IFLA, mentre con questa nota si vogliono avvertire i consumatori eventualmente allergici che LRM può contenere tracce di semantica.

I modelli sono nati nell'ambito della Cataloguing Section (a parte FRSAD) e ne rispecchiano l'interesse e le competenze prevalenti. In FRBR² all'approfondita analisi dell'oggetto bibliografico visto nella sua quadruplica caratteristica (le entità del gruppo 1) e nella ricca articolazione di relazioni (fra entità e fra istanze della stessa entità) non corrispondeva una benché minima rappresentazione della complessa trama di relazioni e significati tipica del versante semantico. Semplicemente era registrata l'esistenza di entità rappresentate nei cataloghi come accessi semantici in relazione di soggetto con l'entità *opera*, categorizzate come: *concetto*, *oggetto*, *evento* e *luogo*. La relazione di soggetto era riconosciuta anche per ogni altra entità dei gruppi 1 e 2, trascurando peraltro il crearsi di sovrapposizioni come quella dell'*esemplare* che è un *oggetto*. In sintesi, al modello sviluppato per il versante descrittivo era appiccicata in un modo che direi posticcio una inevitabile 'citazione' del versante semantico, che però non veniva modellizzato in quanto tale. L'esistenza di relazioni fra le entità che fungono da soggetto, vuoi paradigmatiche, cioè insite nella natura delle entità stesse e comunemente percepite nei nostri parametri conoscitivi, vuoi sintagmatiche, cioè attivate all'interno di specifici discorsi svolti nelle singole *opere*, non veniva rilevata, analizzata, modellizzata, nonostante siano proprio tali relazioni a costituire il contesto, la rete di significati e la leggibilità stessa dei sistemi di indicizzazione che non vogliono essere soltanto insiemi parcellizzati e sconnessi di singoli accessi.

La consapevolezza di un lavoro ancora da compiere era espressa così: "further analysis is needed of the entities that are the centre of focus for subject authorities, thesauri, and classification schemes, and of the relationships between those entities" (FRBR, p. 5).

Da parte italiana la carenza venne evidenziata con osservazioni critiche e proposte di integrazione secondo l'approccio analitico-sintetico, sulla linea tracciata da PRECIS e GRIS e poi dal Nuovo soggettario, di fatto

1 https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

2 https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf

poco ascoltate in ambito internazionale, mentre risultò incisivo un articolo di Tom Delsey, già consulente e principale ideatore di FRBR, apparso nel 2005 e che richiedeva una piena considerazione delle relazioni semantiche e sintattiche all'interno del modello, pur prevedendo di limitarsi a un livello alto, per lasciare definizioni precise ai singoli linguaggi di indicizzazione³.

La risposta venne con il rapporto FRSAD⁴, inteso anche a integrare FRAD per i dati d'autorità di soggetto là lasciati in sospenso⁵.

In FRSAD veniva superata la categorizzazione delle entità in relazione di soggetto, ponendo l'entità *thema* come unica relazione di soggetto dell'*opera*. Una sorta di superclasse, adottata una volta riconosciuta l'impossibilità di definire categorie senza sovrapposizioni e senza residui e valide per tutti. La designazione scelta, *thema*, riconosceva la semanticità delle entità assunte come soggetto; le parti del rapporto dedicate ad *aboutness*, *ofness*, *isness* e alle diverse posizioni filosofiche (nominalismo vs realismo) in discussione intorno all'analisi del contenuto concettuale e all'impostazione dell'indicizzazione, così come i riferimenti esemplificativi ai sistemi più diffusi, collocavano questo modello pienamente in ambito semantico. Venivano analizzate le relazioni semantiche fra diverse istanze di *thema* secondo le tipologie comuni in letteratura: gerarchiche, associative ... (5.3). Non venivano presentate le relazioni sintattiche fra concetti che insieme costituiscono un soggetto complesso, ma se ne poteva scorge la presenza dove era riconosciuto che un *nomen* può essere costituito da parti componenti e che una parte può essere a sua volta un *nomen*, con una relazione intero-parte (5.4.2), per esempio un termine come parte di una stringa in cui è presente. Questa relazione però non veniva sviluppata nel suo valore; restando a livello di *nomen*, quindi di termini, notazioni, simboli, non di *thema*, quindi di concetti, non si incrociava con la rete semantica, e come relazione intero-parte dei *nomens* non rappresentava i ruoli dei singoli concetti che costituiscono il *thema* d'insieme. Toccava la composizione dei descrittori, trascurando i significati sottesi.

Il rapporto FRSAD manteneva posizioni neutre rispetto a specifici sistemi di indicizzazione, non volendo adottarne qualcuno né escluderne alcuno, doveva quindi ammettere diverse applicazioni dell'analisi concettuale, rispettare le differenze fra i linguaggi di indicizzazione in uso. Si giustificava così la sospensione di analisi e definizioni più approfondite che avrebbero seguito inevitabilmente l'orientamento di qualche sistema senza trovare corrispondenza in tutti. Modello che fotografava gli elementi comuni della variegata realtà dell'indicizzazione, una grande panoramica di masse un po' sfuocate, senza i dettagli, lasciati alle specifiche implementazioni.

3 Modeling subject access : extending the FRBR and FRANAR conceptual models / by Tom Delsey // *Cataloging & classification quarterly*, 39 (2005), no. 3/4, p. 49-61

4 <http://www.ifla.org/files/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>

5 www.ifla.org/files/assets/cataloguing/frad/frad_2009-it.pdf; ugualmente nella successiva versione corretta del 2013, https://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf

L'armonizzazione dei tre modelli ha prodotto l'IFLA LRM che ha ristrutturato le entità abolendo i gruppi di FRBR (restano le entità dei gruppi 1 e 2, solo il *luogo* fra quelle del gruppo 3) e introducendo una nuova entità generale, con la trasmissione delle sue proprietà alle altre entità, che le sono subordinate in relazione gerarchica (4.1.2). La nuova super-entità è denominata *res* ed è in relazione di soggetto con l'*opera* (LRM-R12). Quindi soggetto di un'*opera* è sempre genericamente una *res* e può essere più specificamente una qualsiasi delle entità bibliografiche definite (*opera, espressione ... persona, ente*).

L'attributo *categoria* di *res* offre la possibilità di marcare le istanze di *res* secondo le categorizzazioni eventualmente previste dai sistemi di indicizzazione adottati, che possono essere implementati nel modello (LRM-E1-A1; sono qui recuperate le entità dismesse dall'ex gruppo 3, *concetto, oggetto, evento*, declassate a esempi di categorie di *res*).

L'entità *thema* di FRSAD è disciolta nella *res* (e la parola non compare nel nuovo documento), con un'equivalente possibilità di copertura universale, ma perdendo la caratterizzazione propriamente semantica che aveva in FRSAD. Stessa sorte per le relazioni semantiche fra *thema* e *thema*, scomparse ed eventualmente lasciate all'implementazione locale, non sostituite dalla genericissima e indifferenziata relazione '*res* è in relazione con *res*' (LRM-R1). Qui la nota d'ambito precisa: "This is a general relationship valid for all entities in the bibliographic universe. In general, specific refinements would be defined to carry more precise semantics" e fra gli esempi compare una relazione associativa in senso paradigmatico ('*Quantum theory* è associata a *Thermodynamics*', termini che troviamo in relazione reciproca in LCSH). Secondo il carattere gerarchico attribuito anche alle relazioni (4.3.2), le relazioni paradigmatiche tipiche dei thesauri potrebbero essere implementate sul secondo livello come LRM-Rx in aggiunta alle 35 già definite come LRM-R2/R36: per esempio, le relazioni gerarchiche potrebbero essere definite nella forma '*res* is broader than/is narrower than *res*'.

È confermata la relazione partitiva, già in FRSAD, fra un *nomen* e una sua parte che sia pure un *nomen* (LRM-R16): valgono le osservazioni svolte sopra, con l'avvertenza che qui sono coinvolti i più disparati generi di *nomens*, non solo gli indici semantici. Troviamo un esempio d'indice semantico: '*Measles* [Morbillo] è parte di *Measles/epidemiology*', che è solo una segmentazione della sequenza del *nomen* e giustamente non significa che il morbillo sia una parte dell'epidemiologia del morbillo, ma si avverte la mancanza di una relazione che rilevi che il morbillo (una malattia esantematica) può essere visto come oggetto di studio della disciplina epidemiologia (una specialità medica) e di fatto lo è almeno nel caso reale di questa opera.

Non sono da confondere con relazioni sintattiche le relazioni associative fra *res* e *luogo* e fra *res* e *intervallo di tempo* (LRM-R33 e LRM-R35): si tratta della trasformazione in entità-in-relazione degli attributi luogo e tempo di FRBR, ancora funzionali all'identificazione dell'entità correlata (per esempio luogo e data di nascita di una persona, di composizione di un'*opera*), non delle classiche suddivisioni (o ruoli o faccette)

spazio/temporali applicate a un soggetto così delimitato; queste ultime rientrano nella *res-thema* di cui fanno parte senza tenervi un posto distinto. Seguendo il modello, possiamo dire che la *persona* Mozart è in relazione con il *luogo* Salisburgo e con il *tempo* dell'anno 1756, suoi dati anagrafici, mentre per la 'permanenza e attività di Mozart a Napoli nel 1770', che è il tema di un'*opera*, ci dobbiamo limitare a registrarla nel suo insieme come una *res*, senza altre relazioni, a parte quella con il *nomen* che la rappresenta nel nostro linguaggio, 'Mozart, Wolfgang Amadeus - Napoli - 1770'; stringa di cui potremmo considerare le relazioni con i singoli *nomens* che ne fanno parte, relazioni qualificate partitive secondo LRM-R16, ma non c'è modo di registrare la connessione tra persona e luogo e tempo, dichiarando questi ultimi delimitazioni spaziale e temporale del discorso svolto nell'opera su quella.

Se scorriamo poi le funzioni per l'utente che stanno alla base dell'elaborazione del modello LRM (3.2 e 6.1) vi troviamo alcune funzioni normalmente svolte dall'indicizzazione semantica, ma se quelle ovvie per i principi di uniformità e univocità (come trovare e identificare) funzionano col modello sulla base della relazione fondamentale '*res* has appellation/is appellation of *nomen*' (LRM-R13), quelle tipiche della semantica non vi hanno riscontro, se non attraverso l'implementazione di sistemi d'indicizzazione specifici (come esplorare, quando richiede le relazioni semantiche, qui assenti).

Come dichiarato in 2.2 il modello concettuale intende mantenersi a un livello alto, è pensato come base per l'implementazione di sistemi bibliografici, che hanno bisogno di maggior precisione e di espansione, e possono permetterselo. E così è: solo attraverso l'implementazione di altre entità, attributi, relazioni corrispondenti ai sistemi reali si potrà dire che il modello concettuale rappresenta in modo sufficientemente articolato e preciso l'universo bibliografico nella sua totalità.

Queste le tracce di semantica che trovo in LRM. Ripudiata la semantica di FRASAD siamo ritornati a livello di doverosa citazione in un documento volto alla catalogazione descrittiva (se si può ancora dir così), nulla di più. D'altra parte è quanto praticato a suo tempo da FRAD e anche nella *Statement of international cataloguing principles* (ICP) e sembra il destino dell'innovativa RDA che ritarda e poi rinuncia a compilare i capitoli predisposti per l'indicizzazione semantica. S'è già detta la motivazione principale di questo vuoto, l'estrema differenziazione dei sistemi in uso che rende impossibile individuare anche solo pochi elementi comuni, appropriati per ogni sistema. Né è obiettivo dell'IFLA produrre e promuovere standard o porre soglie minime di tolleranza in questo campo, nel quale manca quella tradizione di confronto e costruzione in comune di principi e criteri normativi condivisi che caratterizza l'ambito della catalogazione descrittiva. Diversamente da altri organismi, come l'ISO, che istituzionalmente producono norme internazionali e puntano a prescrivere le migliori pratiche secondo i principi più corretti, l'IFLA si limita a riordinare concettualmente l'esistente. L'intento dichiarato, da quando la parola *record* è stata sistematicamente sostituita da *data*, è di rendere i dati registrati adatti al nuovo ambiente della comunicazione che ormai si sta delineando con le tecniche dei dati collegati aperti (*linked open data*, LOD): avanzamento nelle tecnologie, non nella teoria e pratica della

catalogazione e indicizzazione, ampliamento dell'area informativa d'interesse, non approfondimento dei meccanismi di rappresentazione e recupero dell'informazione bibliografica. Questo resta all'elaborazione che vorranno farne i gestori dei singoli sistemi, in proprio e in spazi di confronto che non mancano; godendo della massima libertà che un modello così scarno ed essenziale lascia.

A partire da queste considerazioni forse non ci si doveva aspettare di più dal consolidamento dei modelli concettuali. Tuttavia, per spezzare un'ultima spuntata lancia in favore del modello analitico-sintetico, mi chiedo se non ci sia un difetto di fondo nel punto di vista che, già da FRISAD, analizza l'indicizzazione semantica attraverso gli indici che vengono prodotti e utilizzati, considerandoli come unità rigide, siano essi termini, stringhe, notazioni o altro, quindi senza praticarne la scomposizione (e ricomposizione) in funzione, per esempio, dell'esplorazione. Mentre considerarne il riferimento o i riferimenti, la o le entità reali che essi denotano apre immediatamente alla possibilità di modellizzare le singole componenti e le loro associazioni, di esplorarne tutti i collegamenti, di trovare in modo preciso e di selezionare in modo consapevole le entità effettivamente interessanti, sfruttando davvero la tecnica dei dati collegati. Realtà riconoscibile anche sotto quei sistemi che non la esplicitano e nelle logiche di ricerca post-coordinate, che non risulterebbero quindi escluse dall'assunzione di elementi analitico-sintetici nel modello. Che resta, al di là delle dichiarazioni di intenti e con tutte le giustificazioni del caso, monco dell'altra metà della catalogazione.