

## **Modeling knowledge, archival and bibliographic standards in comparison.**

**1<sup>st</sup> JLIS.it Seminar - May 19<sup>th</sup> 2022**

### **Editorial board**

---

The first JLIS.it Seminar, “Modelling Knowledge”, held in Florence last May 19th was an occasion of cultural and scientific growth for the journal and, we hope, also for LIS community.

Try to contribute to modeling Knowledge is undoubtedly an ambitious task requiring, firstly and foremost, humility and complete intellectual openness. We have seen this constructive attitude in all those who contributed to the Seminar through their lectures and interventions in the debate. Also for this reason - and as was hoped - more problems than solutions emerged from the discussion and a strongly shared need to deepen conceptual analysis in the direction of a sustainable application of the various archival and library science methodologies was manifested.

Modeling Knowledge means taking note of the inevitable integrated dimension in which we must now place the LIS disciplines: Archival Science, Library Science, Digital Humanities and Documentation. The sense of this integration lies in the collaborative and concrete comparison between structured disciplines intrinsically destined to act together, especially in the future.

The descriptive paradigms of Archival Studies and Library Science have long been urged to open up to multidimensionality in order to go beyond the sound methodological background acquired over decades of reflections on these themes.

The issue is broad and complex, and many possible approaches are achievable, just as the applicative contexts in which each reflection must necessarily be dropped seem to be diversified.

In the awareness of these complexities, JLIS.it seeks to interpret these widespread instances and propose them for a methodological reflection, in the conviction that integration must first be declined on a theoretical level. The models referred to are essentially an attempt to imagine not new but renewed arrangements considering the vaguely paligenetic need running through much of the vision within our disciplines. Theoretical speculation, however, must inevitably be stressed inside the applicative dimension, description, that represents the operational arm and litmus test of any theoretical consideration.

For all these reasons, JLIS.it intends and will increasingly pay attention to concrete knowledge modeling tools, from shared metadata to intercultural and multi-contextual solutions, which are less and less futuristic, indeed already present in various professional contexts.

The technical aspects, however, must be set in a credible “political” scenario so that cooperation between the various players will not remain just a vague declaration of intent.

In this sense, then, it seems appropriate to also recall the “political” character of the May 19th initiative, testified since the very opening of the proceedings - and reiterated in several speeches - to mark the need for an effective cooperation between the disciplinary domains involved, starting with those of book, library and document sciences.

In that regard, we may observe that for some time now, there has been a particular insistence on disciplinary cooperation and integration between Archival Science and Library Science, but never defining the concrete territories of possible convergence. In this sense, we are convinced that the interesting indications that emerged during the Seminar certainly deserve to be relaunched.

Moreover, each of these disciplines derives from an established tradition of method and practice, and this biodiversity must be protected to avoid disintegration. Paradoxically, but not excessively so, the best way to calibrate the meeting point between the various domains requires intellectual honesty to state the respective positions explicitly. Not a simple addition of competencies is needed, but a multiplication of common and shared/shareable factors.

Then, one can say that the first result of the Seminar was to clarify two complementary approaches to knowledge. Indeed, on the one hand, the library sector’s deep and evolved focus on identifying and managing information objects emerged. Valuable reflections on titles, i.e., on the names and surnames of resources, on the management and restitution of entities, on the complex role of conceptual models and standards, and possible application responses indicate the depth of scholars’ research on the many forms of the book and its description in terms of incremental knowledge mastery.

On the archival side, historically and epistemologically oriented to an inevitable approach made of successive approximations to data, there is instead growing attention to relational mechanisms, the reciprocal relationship between different information entities, and, above all, the role of the many contexts within which we are called to browse. The archival practice seems to contribute with other systemic traits even when the most classic of archival functions are put at the service of operational needs that are far more cogent than the rights and duties of the method.

It is, therefore, a matter of managing two different complexities, content and context, which have historically been a source of mutual distress on both sides. The recipe for achieving the disciplinary union to which the entire Seminar was aiming lies in the actors’ pliability to reflect on the canons and founding concepts of their respective domains, even re-modeling the canons themselves, without forfeiting their previous skills. The concrete objective of integration - going beyond general theoretical proclamations - is certainly to be sought and constructed in the application dimension, particularly in the spaces that may open within new-generation intercultural systems. Artificial intelligence, and more generally the technologies available today, guarantees developments of deep interest in this direction and opens up a constructive confrontation with the generality of Digital Humanities. This potential has already emerged from the presentations of some Seminar’s speakers. Most importantly, it is necessary to ensure a systemic cultural and scientific approach to avoid technicalities that would betray the founding values of the various disciplines at play by indulging in a quantitative and uncritical use of ICT.

Ultimately, the Seminar’s work helped to outline the possible short-term scenarios within which to place a conceptual and applicative reflection on the ways of producing, managing, using and preserving knowledge, respecting inalienable ethical, cultural, scientific, and epistemological values. The hope of this editorial is precisely synthesising the many ideas that emerged during the debate to relaunch an in-depth reflection, to which JLIS.it intends to contribute both in terms of content and spaces open to different sensibilities.

## **Modellare la conoscenza, standard archivistici e bibliografici a confronto.**

### **1° Seminario JLIS.it - 19 maggio 2022**

Il primo seminario di JLIS.it, “Modellare la conoscenza” ha rappresentato un momento di crescita culturale e scientifica per la rivista e, ci auguriamo, anche per la comunità di riferimento.

Porsi l’obiettivo di contribuire a “Modellare la conoscenza” è senz’altro un’operazione ambiziosa che richiede innanzitutto umiltà e piena disponibilità intellettuale. Crediamo di aver constatato questo atteggiamento costruttivo in quanti hanno contribuito al seminario del 19 maggio scorso, sia con le loro relazioni, sia con gli interventi nel dibattito. Anche per questa ragione – e com’era auspicabile – dal confronto sono emersi più problemi che soluzioni e si è manifestato un bisogno fortemente condiviso di approfondire l’analisi concettuale in direzione di un’applicazione sostenibile delle diverse metodologie archivistiche e biblioteconomiche.

Modellare la conoscenza significa infatti prendere atto dell’inevitabile dimensione integrata in cui dobbiamo ormai collocare le discipline LIS: Archivistica, Bibliografia, Biblioteconomia, Digital Humanities, Documentazione. Le logiche di questa integrazione sono quelle di un confronto collaborativo e concreto tra discipline strutturate ma intrinsecamente destinate ad agire insieme, soprattutto in prospettiva futura.

I paradigmi descrittivi dell’Archivistica e della Biblioteconomia, in particolare, sono da tempo sollecitati ad aprirsi a una multidimensionalità capace di andare oltre il solido bagaglio metodologico comunque acquisito nel corso di decenni ricchi di riflessioni su questi temi.

La questione è ampia e complessa e sono molti i possibili approcci al tema, così come sembrano diversificati i contesti applicativi nei quali ogni riflessione deve necessariamente essere calata.

Nella consapevolezza di tali complessità JLIS.it cerca di farsi interprete di queste diffuse istanze e di proporle alla riflessione metodologica, nella convinzione che l’integrazione debba essere declinata prima di tutto sul piano teorico. I modelli a cui si fa riferimento sono sostanzialmente un tentativo d’immaginare assetti non nuovi ma rinnovati che tengano conto del bisogno vagamente palinogenetico che attraversa buona parte del campo visivo delle nostre discipline. La speculazione teorica deve però inevitabilmente misurarsi con la dimensione applicativa, cioè con quella descrizione sul campo che rappresenta il braccio operativo e la cartina di tornasole di qualsiasi considerazione teorica.

Per tutte queste ragioni JLIS.it intende e intenderà sempre più prestare attenzione agli strumenti di modellazione concreta della conoscenza, dai metadati condivisi alle soluzioni interculturali e multicontestuali, sempre meno futuribili, anzi già presenti in diversi contesti professionali.

Gli aspetti tecnici devono però potersi collocare in uno scenario “politico” credibile all’interno del quale la collaborazione tra i diversi soggetti non rimanga una vaga dichiarazione di intenti.

In questo senso ci sembra allora opportuno richiamare anche il carattere “politico” dell’iniziativa del 19 maggio, testimoniato fin dall’apertura dei lavori - e ribadito in diversi interventi - per marcare il bisogno di una efficace cooperazione tra i domini disciplinari coinvolti, a partire da quelli delle scienze del libro, della biblioteca e del documento.

Al riguardo va notato che da tempo si parla con una certa insistenza di cooperazione disciplinare e di integrazione tra Archivistica, Bibliografia e Biblioteconomia, senza però mai definire davvero i territori concreti delle possibili convergenze. Anche in questo senso ci sembra che nel corso del seminario siano emerse indicazioni interessanti che meritano sicuramente di essere rilanciate.

Ognuna di queste discipline arriva peraltro da una consolidata tradizione di metodo e di prassi e tale biodiversità va senza alcun dubbio protetta, per evitare dis-integrazioni. Paradossalmente, ma nemmeno troppo, il modo migliore per calibrare il punto di incontro tra i vari domini richiede allora l'onestà intellettuale di dichiarare esplicitamente le rispettive posizioni. Quella che serve non è infatti una semplice addizione di competenze, ma una moltiplicazione di fattori comuni e condivisi/condivisibili.

In linea generale si può dire allora che il primo risultato del seminario sia stato quello di chiarire due approcci complementari alla conoscenza. Infatti, da un lato emerge la forte ed evoluta attenzione del comparto biblioteconomico all'individuazione e alla gestione degli oggetti informativi. Preziose riflessioni sui titoli, cioè su nomi e cognomi delle risorse, sulla gestione e la restituzione delle entità, sul ruolo complesso dei modelli concettuali e degli standard e sulle possibili risposte applicative indicano con chiarezza la profondità del pensiero degli studiosi delle molte forme del libro e della sua descrizione in termini di padronanza incrementale della conoscenza.

Sul fronte archivistico, storicamente ed epistemologicamente orientato a un inevitabile approccio per approssimazioni successive al dato, si registra invece una maggiore attenzione ai meccanismi relazionali, al rapporto reciproco tra le diverse entità informative e, soprattutto, al ruolo dei molti contesti dentro ai quali si è chiamati a muoversi. La pratica archivistica per questi motivi sembra portare un contributo dai tratti più sistemici anche quando le più classiche delle funzioni archivistiche vengono messe al servizio di esigenze operative ben più cogenti dei diritti e dei doveri del metodo.

Si tratta quindi di gestire due diverse complessità, il contenuto e il contesto, che proprio per quanto appena detto sono state storicamente sorgente di reciproco affanno sull'uno e sull'altro fronte. La ricetta per ottenere il connubio disciplinare cui l'intero seminario tendeva, sta nella duttilità degli attori e nella disponibilità a riflettere sui canoni e sui concetti fondanti dei rispettivi domini in termini di nuova modellazione dei canoni stessi, senza per questo rinunciare alle competenze pregresse. L'obiettivo concreto dell'integrazione – andando oltre ai proclami teorici di ordine generale – va sicuramente cercato e costruito nella dimensione applicativa, in particolare negli spazi che possono aprirsi dentro a sistemi interculturali di nuova generazione. Le intelligenze artificiali, e più in generale le tecnologie oggi disponibili, garantiscono in prospettiva sviluppi di grande interesse in questa direzione e aprono la riflessione a un confronto costruttivo con la generalità delle DH. Tali potenzialità sono del resto emerse già dagli interventi di alcuni relatori del seminario. Occorre però garantire innanzitutto un approccio culturale e scientifico di natura sistemica per evitare derive tecnicistiche che tradirebbero i valori fondanti delle diverse discipline in gioco, assecondando un uso quantitativo e acritico delle ICT.

In definitiva, quindi, i lavori del seminario hanno contribuito a tratteggiare i possibili scenari entro i quali collocare nel breve periodo una riflessione concettuale e applicativa sulle modalità di produzione, gestione, uso e conservazione della conoscenza, nel rispetto di irrinunciabili valori di ordine etico, culturale, scientifico ed epistemologico.

L'auspicio di questo editoriale è perciò proprio quello di sintetizzare i molti spunti emersi dal dibattito per rilanciare una riflessione approfondita, cui JLIS.it intende contribuire sia in termini di contenuti che di spazi aperti alle diverse sensibilità.

## The parts and the whole. Integrate knowledge

Federico Valacchi<sup>(a)</sup>

a) University of Macerata, <https://orcid.org/0000-0003-2710-9316>

**Contact:** Federico Valacchi, [federico.valacchi@unimc.it](mailto:federico.valacchi@unimc.it)

**Received:** 2 May 2022; **Accepted:** 9 May 2022; **First Published:** 15 September 2022

### ABSTRACT

The article deals with the themes of integration between the different cataloguing and disciplinary descriptive traditions. It initially evaluates the transition phase from methods to established practices up to the new solicitations caused by the transformations of information production models. Lastly, the paper focus on the theme of the reconceptualization of domain languages to move on to outline the concept of metacataloguing as a possible descriptive synthesis.

### KEYWORDS

Cataloguing; Metacataloguing; Archival description; Heritage.

## Le parti e il tutto: integrare la conoscenza

### ABSTRACT

L'articolo si confronta con i temi dell'integrazione tra le diverse tradizioni catalografiche e descrittive disciplinari. Valuta inizialmente la fase di passaggio da metodi e prassi consolidate a nuove sollecitazioni determinate dalle trasformazioni dei modelli di produzione dell'informazione. Ci si sofferma poi sul tema della riconcettualizzazione dei linguaggi di dominio per passare infine a trattare il concetto di meta catalogazione come possibile sintesi descrittiva.

### PAROLE CHIAVE

Catalogazione; Metacatalogazione; Descrizione archivistica; Patrimonio.

*“L’artista deve nell’esecuzione superare  
non solo i limiti che porta con sé  
il carattere specifico della sua arte,  
ma anche quelli che dipendono  
dal soggetto particolare del suo lavoro”*  
Schiller 1927, 106

## 1. Sul crinale

Ipotizzare di modellare la conoscenza, in una società che sperimenta quotidianamente il paradosso nascosto nell’ingovernabilità delle valanghe di informazioni che genera senza pietà per sentirsi viva, significa muoversi su un frastagliato crinale. Si oscilla tra un’ipermemoria compulsiva e un oblio incombente nel quale si specchiano i fantasmi di una società postmnemonica (Connerton 2010).

Jacques Le Goff scriveva in tempi non digitalmente sospetti che la pubblica opinione è ossessionata “dal timore di una perdita di memoria, di un’amnesia collettiva, che trova una goffa espressione nella cosiddetta *mode retro*, o moda del passato, sfruttata spudoratamente dai mercanti di memoria, dal momento che la memoria è diventata uno degli oggetti della società dei consumi che si vendono bene” (Le Goff 1984, 394). A ben vedere, quindi, il terrore di un oblio venturo genera uno smarrimento profondo, che rischia di spingerci nelle braccia di un passato mitizzato dove il profilo stesso della memoria viene alterato, facendone ansiolitico prodotto di consumo piuttosto che riserva di progettualità.

Sappiamo da tempo che questa informazione ipertrofica dilaga oltre i suoi stessi veicoli documentari e suggerisce azioni di gestione della conoscenza che vanno al di là le competenze delle singole discipline in gioco (Guarasci 2008). La questione non è semplicemente quantitativa, non basta processare volumi crescenti di dati per poter parlare di conoscenza. La gestione della conoscenza deve puntare anche alla qualità, in termini di affidabilità e reperibilità dei risultati. C’è la necessità di documentare la documentazione e i processi secondo quale viene generata, gestita e fruita, ricorrendo non solo ai metadati ma anche a paradata funzionali a un potenziamento ad ampio raggio di qualificate funzionalità informative (Huvila, 2022). L’elemento culturale, inteso soprattutto come adeguata capacità di valutazione, non si deve liquefare di fronte all’incalzare massificante dei calcolatori, per quanto essi siano nostri preziosi alleati.

La lunga ondata digitale in tutto ciò ha certamente un ruolo, sia pure non esclusivo, soprattutto nel momento in cui mette in crisi l’identità fisica dei luoghi che del ricordo erano rifugi sicuri, senza sostituirli con soluzioni altrettanto riconoscibili e confortevoli. La memoria, nella sua versione homeless, tende a moltiplicare i suoi messaggi e ad accatastarli in maniera frammentaria nei molti contenitori possibili, siano essi gli usuali ‘archivi’ che la sfumata sedimentazione dei social o altro ancora. Lo spazio diviene quindi una categoria interpretativa ancora più forte del tempo che, peraltro, non è mai stato l’unico arbitro del ricordo. Sembra quindi necessario continuare a riconoscere il giusto ruolo ai luoghi della memoria, ancora prima di affannarci a comprendere, contenere e rilanciare l’azione del tempo.

Come possiamo difenderci da questo sgretolamento fisico o, meglio, come fare per ricostruire una fisicità che resta essenziale anche nei non luoghi digitali? Innanzitutto riconoscendo tale fisicità,

per quanto fortemente mediata. Gli spazi della memoria si sono delocalizzati, sono scivolati lungo il piano inclinato di geografie impalpabilmente binarie ma non si sfugge alla materialità ferrigna dell'hardware e dei supporti che ospitano i bit. L'idea di spazio non può essere abbandonata, non si può rinunciare alla percezione e alla tutela fisica di ciò che leggiamo, vediamo ed elaboriamo a partire da un'immaterialità solo presunta. Negare la virtualità di ciò che chiamiamo virtuale ci può servire a ricomporre la latente schizofrenia da formato.

L'esigenza di inseguire l'informazione dentro ai molti serbatoi digitali dove si deposita non ci esime dalle esigenze classificatorie con cui da sempre governiamo quella stessa informazione. Se possibile, il ruolo delle nostre tassonomie in questi ambienti è anzi ancora più qualificante. Si fa stringente il bisogno di organizzare la conoscenza e di collocarla dentro a schemi sempre più elastici e interconnessi, dentro ai quali possiamo tenere sotto controllo la potenza di macchine che si credono pensanti e puntano all'emulazione sinaptica (Tomasì 2022).

Siamo ormai nei paraggi dello spartiacque che separa consolidati bacini metodologici da territori dove si muovono suggestioni ancora non del tutto formalizzate, per quanto già efficaci nei diversi domini del governo delle informazioni. Sono proprio queste suggestioni, che peraltro non hanno solo fattezze digitali, a suggerirci la strada verso la costruzione di una conoscenza consapevole.

Le dicotomie, o quanto meno le distinzioni possibili, sono molte e muovono dalle specificità delle singole risorse documentarie (libri, documenti e diverse tipologie di risorse e oggetti digitali) e dalle diverse finalità che l'organizzazione della conoscenza può porsi.

A questo riguardo è opportuno precisare che nelle pagine che seguono si farà riferimento essenzialmente alla dimensione storica e culturale delle diverse risorse documentarie con cui ci si confronterà. Porsi il problema nei termini più ampi della gestione, conservazione e accesso ai sistemi documentari digitali in formazione, magari valutandoli in una potenziale prospettiva storica, significherebbe infatti allargare a dismisura l'area di analisi. Ciò non significa però che non si debba essere consapevoli della centralità di questi aspetti, che nel medio periodo porranno a loro volta problemi di classificazione e accesso destinati a diventare esclusivi. Resta vero, infatti, che le responsabilità dei professionisti dell'informazione sono ormai nitidamente proiettate in un futuro sempre più decifrabile, almeno per quanto riguarda il formato delle cose documentarie con cui ci dovremo confrontare. Tenerlo presente, assumendosene la responsabilità etica e deontologica che ne deriva, sarà utile anche all'interno degli scenari in cui ci muoveremo qui, dal momento che l'idea di progettazione sottesa alla *long time preservation* ha in sé il bisogno di saldarsi a pratiche e sensibilità pregresse (Pigliapoco 2016; Ciandrini 2020).

Un'altra premessa deve riferirsi alla accresciuta perimetrazione del contesto di riferimento, poiché la conoscenza che vorremmo modellare non può più rimanere chiusa nei singoli domini di provenienza. Ridurre la realtà in autarchici brandelli disciplinari ci allontana dalla comprensione della sua complessità. Le logiche di un'inevitabile alleanza tra le discipline ci spingono verso la raffigurazione integrata di ciò che con un'espressione un po' modaiola si definisce *heritage*. Ciò che dobbiamo descrivere è il patrimonio culturale inteso nella sua colorata pienezza. Le diverse strategie descrittive possono e devono confluire verso questa lettura tridimensionale, nel tentativo di dare corpo anche alle suggestioni che da tempo ormai in Italia hanno trovato espressione nelle logiche del MAB (Bruni et al. 2016).

Andare in questa direzione significa però confrontarsi anche con le potenzialità in parte non codificate delle digital humanities, sia pure nella consapevolezza che "definire il concetto di DH

non è impresa facile. Lo statuto di questo ambito, o anche settore o area di ricerca, varia al variare del punto di vista dell'osservatore”(Tomasi 2022, 7). Di sicuro, in ogni caso, oltre a un innegabile relativismo, c'è “lo stretto rapporto tra DH e scienze dell'informazione”(Tomasi 2022, 8), rapporto da cui possono scaturire opportunità imprevedibili.

Non si può liquidare la questione in termini tecnologici ma l'approccio che le tecnologie ci suggeriscono sembra inevitabilmente collaborativo, lontano da reciproci autoreferenzialismi di scuola o di maniera e dalle tentazioni di un 'beneculturalismo' generalista, dal sapore più turistico che scientifico. Integrare può aiutare a dare un senso tangibile all'infido concetto di 'valorizzazione' e alle sue troppo generiche declinazioni di azione a sostegno dell'*heritage*. È appena il caso di notare al riguardo come l'insistenza sulla valorizzazione sia in ultima analisi la constatazione che un valore ancora non c'è, con buona pace dei proclami che da decenni esaltano la ricchezza del nostro patrimonio culturale e lo rilanciano a vuoto dentro a contenitori *very belli*. Il valore però è una ricaduta della conoscenza e quindi prima di valorizzare bisogna conoscere e per conoscere bisogna catalogare e descrivere, senza indulgere a fumose strategie di 'comunicazione' (o peggio ancora 'narrazione') fondate su percezioni superficiali condite in una insipida salsa digitale.

Sappiamo che non si parte da zero perché negli anni si è venuta accumulando una significativa massa critica di descrizioni raffinate, talvolta perfino troppo raffinate. Se ci concentriamo sul caso archivistico, ad esempio, emerge come il lavoro di progettazione, realizzazione e implementazione dei sistemi informativi abbia prodotto risultati più che apprezzabili. Pur con i limiti fisiologici che tale sperimentazione ha conosciuto ai suoi esordi, nel tempo è venuta sedimentandosi una solida base descrittiva che costituisce un valido e qualificato presupposto per successivi sviluppi, al di là del suo intrinseco valore informativo<sup>1</sup>. Proprio SAN, 'il sistema dei sistemi', segna per molti versi uno spartiacque tra due diverse modalità di concepire i processi descrittivi digitali. Nel Sistema Archivistico Nazionale ineludibili risposte di natura gestionale e catalografica convivono infatti con una metabolizzazione delle descrizioni che si apre alla dimensione nuova e allargata dei portali tematici, nei quali i dati archivistici di sistema sono messi al servizio di una più che apprezzabile creatività redazionale<sup>2</sup>.

Senza enfatizzare la portata di questa distinzione, ancora in fieri, credo che si possa cogliere una cesura “generazionale” tra i diversi sistemi. Da un lato abbiamo risorse concepite secondo un uso tutto sommato ancora meccanico delle ICT, che privilegiano approcci fondati su informazione strutturata nei database relazionali. Le *archival relationship* di 'isadiana' memoria sono le chiavi di ricostruzione di un contesto ancora indissolubilmente legato al principio di provenienza. Il risultato è una costruzione/restituzione di autorità, nella quale la natura dei contenuti è determinata dal paradigma metodologico che scaturisce dalla volontà di chi descrive. Disponiamo in altre parole di descrizioni di assoluta importanza, ma, per così dire, ancora comprese in sé stesse. Dall'archivio partono e all'archivio ritornano, fedeli a un'eterna circolarità. Le tecniche euristiche che ne derivano, di conseguenza, sono in qualche modo indotte. Sono sferraglianti sistemi di reciprocità tra entità la cui natura gerarchica e multilivellare non si sgancia mai dalle logiche proprietarie degli

---

<sup>1</sup> Sulla genesi di SIAS, un sistema tra i più importanti, si veda: Feliciati e Grana 2005. Per uno sguardo di insieme si veda: Valacchi 2015.

<sup>2</sup> Sulle motivazioni alla base della progettazione dei sistemi informativi si veda: Feliciati 2009. Sui portali tematici si veda invece: <http://san.beniculturali.it/web/san/archivi-tematici>.



erogatori di contenuti. Si tratta però di un patrimonio prezioso, senza il quale non sarebbe possibile ragionare di sviluppi ulteriori, perché nei SIA c'è il carburante indispensabile a far decollare il razzo dell'integrazione multicontestuale.

Su un altro versante stanno infatti logiche che si basano su esigenze informative a maglie più larghe e meno strutturate. Il web delle cose mette insieme le cose e i diversi strumenti di organizzazione di una conoscenza allargata ridefiniscono obiettivi e priorità (Fortino et al. 2016), arrivando ad influenzare apertamente le filosofie e le indicazioni degli standard di seconda generazione come RiC. La costruzione della conoscenza si va facendo incrementale e si svincola dall'originario soggetto descrittore. È alimentata da istanze di provenienza diversa, che non disdegnano di valutare il contributo portato da utenti che si fanno parte attiva dell'azione descrittiva<sup>3</sup>. Questi *prosumer* giocano un ruolo importante e ormai ineludibile nella costruzione di una conoscenza condivisa ed è sconsigliabile insistere solo su modelli top down, fatti di trasmissione a circuito chiuso. Ciò non toglie che esistano e debbano essere difese peculiarità e garanzie scientifiche ad alta specializzazione, da cui derivano comportamenti adeguati e il coinvolgimento di comunità che abbiano competenze altrettanto adeguate. La conoscenza deve essere aperta fuori da una certa ideologia del libero accesso e senza che ne soffra la qualità, per evitare il rischio di demagogiche semplificazioni.

Queste istanze di trasversalità descrittiva attraversano il mondo delle biblioteche ormai da decenni, come sottolinea Giovanna Lambroni quando allude alla conferenza parigina sui principi di catalogazione del 1961, da cui emergeva già “una nuova concezione del catalogo, non più mero contenitore di record che descrivono libri ma aggregatore di dati su tutta la produzione del mondo della conoscenza registrata [...]”(Lambroni 2021; Bergamin e Guerrini 2021).

Come spesso accade, quindi, il presente non ha bisogno di inventare sé stesso dal nulla. Basta saper ascoltare il passato per confezionare la rotta, sia pure nella consapevolezza che gli ultimi decenni tecnologici hanno contribuito a dare corpo reale a quelle che un tempo erano solo lungimiranti aspirazioni. Ci vuole magari l'accortezza di non lasciarsi travolgere dal tecnicamente possibile, per evitare che le fascinazioni digitali spacchino in due la realtà, ma detto questo è chiaro che le filosofie alla base del machine learning hanno ormai un loro senso profondo in questo contesto e le reti neurali dei computer non sembrano più così disumane. Machine learning e deep learning non sono fantascienza ma pratica quotidiana e le intelligenze artificiali ci chiedono a gran voce di essere utilizzate (Colavizza et al. 2021).

## 2. Le parole che usiamo

Mauro Guerrini e Roberto Guarasci hanno scritto recentemente a proposito delle tecniche di indicizzazione che il termine “indica un'operazione quanto mai complessa: è l'attività di valutazione del messaggio, cioè del contenuto concettuale di un testo” (Guarasci e Guerrini 2022, 13). Non si può che concordare: alle radici di qualsiasi intento descrittivo sta la comprensione culturale, ci sono cioè le parole con le quali riusciamo a dare spessore alle cose, oltre la nudità dei nomi.

Questa fatica onomastica, che diventa fatica delle cose, è probabilmente uno dei tratti di discontinuità più marcati nell'evoluzione recente delle discipline documentarie. Il trasformismo delle

---

<sup>3</sup> Neppure questa è però una novità. Sugli archivi che crescono a partire dagli archivi si veda: Gardini 2021.

risorse non è una novità. Marcia da sempre al passo delle evoluzioni “delle materia scritte e degli arnesi usati per scrivere nei vari tempi e nei vari luoghi” (Cencetti 1978, 7). Certe violente accelerazioni tecnologiche enfatizzano però questo evolucionismo fisiologico e mettono in discussione i tempi e i luoghi che nella lezione cencettiana erano parametri affidabili e monitorabili.

Confrontarsi con il bisogno di descrivere per conoscere e, soprattutto, far conoscere, significa accettare la sfida di questa realtà in lunga trasformazione. Quella che ci si para davanti è però una realtà tormentata, ancora prima che aumentata. Le discipline e le loro statue epistemologiche sono perplesse di fronte alla natura cangiante di ciò che a diverso titolo chiamiamo, documenti, risorse o, ancora più genericamente, ‘oggetti’<sup>4</sup>.

Il problema della concettualizzazione o, meglio, della riconcettualizzazione dei linguaggi di dominio si pone in tutta la sua urgenza e va oltre le ricadute semantiche più appariscenti. Rimodulare il linguaggio significa essere disponibili a metabolizzare cambiamenti che vanno oltre le parole.

Contemporary archival terminology provides a useful and necessary means of specialized communication within the archival profession. Its terms can be precise enough to preserve important distinctions among types of materials and archival institutions, and yet its usage also can be sufficiently flexible to reflect the changing nature of record materials and developments in the administration of archival institutions. As the archival profession grows and matures and as new technologies and records media affect the practice of archives administration, both the precision and flexibility of archival terminology will prove to be of continuing benefit to archivists<sup>5</sup>.

Ci servono ‘precisione’ e ‘flessibilità’. I linguaggi disciplinari devono guardarsi allo specchio nel tentativo di pensare a parole e comportamenti che diano conto di un adeguamento epistemologico e metodologico.

Le stesse parole base, archivio e biblioteca, non bastano più a sé stesse. Tentare oggi una definizione di archivio, ad esempio, significa mettere in gioco una serie di variabili inimmaginabili venti anni fa (Penzo Doria 2022). L’interoperabilità delle istituzioni alimenta quella dei sistemi documentari e l’intangibile linearità produttore/produzione/conservazione/uso giunge a esiti parossistici<sup>6</sup>. La frammentazione e la delocalizzazione della produzione, insieme alla fluidità della conservazione, consigliano di cercare una definizione di archivio che tenga conto del polimorfismo galoppante dei sistemi documentari, sempre ammesso che una sola definizione possa bastare. L’idea stessa di documento ha talvolta il fiato corto. Le viste documentali e le aggregazioni dinamiche di dati provenienti da sistemi diversi fanno dei documenti registrazioni potenzialmente effimere, comunque distanti dalla materialità analogica che le inchiodava al supporto. Come ha notato Alessandro Alfier “l’attuale sistema di documentazione, che appare così promettente e sicuro di sé sull’onda del ricorso alle tecnologie dell’informazione e della comunicazione, attraversa in realtà una fase di assestamento, come se non avesse raggiunto ancora una posizione consolidata” (Alfier 2020, 91). Le conseguenze, evidentemente, non sono semplicemente meccaniche ma si al-

<sup>4</sup> A questo riguardo si veda la “Premessa terminologica” di Guarasci e Guerrini 2022. Si veda anche Giglio 2021.

<sup>5</sup> <https://www.archives.gov/research/alic/reference/archives-resources/terminology.html>

<sup>6</sup> Conviene ricordare come proprio la progressiva rottura del rapporto tra produzione, uso e conservazione che si inizia a manifestare a partire dagli ultimi anni del XVIII secolo sia ritenuta una svolta nel percorso di avvicinamento ad un’archivistica intesa come disciplina scientifica, per quanto di natura prevalentemente storica.

largano alla fisionomia di insieme dei sistemi documentari. Mettono in discussione categorie fondanti come autenticità, affidabilità e conservazione nel tempo dei requisiti complessivi di qualità del documento e dei processi nei quali esso si manifesta, riproponendo in questo senso l'esigenza di disporre anche di più ampi sistemi di paradata (Duranti e Rogers 2019).

### 3. Modellare la mediazione

La variegata complessità che abbiamo evocato ci induce a una lettura moltiplicata del concetto stesso di descrizione/catalogazione, magari mettendo sullo sfondo sterili purismi corporativi che ci allontanano da soluzioni praticabili.

Modellare la conoscenza significa porsi il problema di dar vita a tassonomie dinamiche, capaci di rispondere a un'incessante produzione di dati che si presentano in ordine sparso (Marrè e Rovella 2008). Continuiamo a muoverci nel campo ampio e plurale della mediazione, cioè della più nobile delle arti documentarie, per quanto ormai anch'essa ci si manifesti moltiplicata fin dai modi e dai tempi del suo esercizio, sospesa tra l'umanità dei servizi di reference e strumenti di decodificazione digitale dei ruoli, dei servizi e dei contenuti.

Le pratiche di *knowledge management* ci insegnano che l'informazione ha un valore specifico, si tratta solo di stabilire se esso si possa tradurre in moneta sonante o se, come nel caso nostro, abbia ricadute di diversa natura. La mediazione, comunque la si eserciti, è ricchezza. La capacità di conoscere, interpretare e restituire in forma accessibile contenuti per loro natura ostici o comunque difficili da trovare, capire e usare è un irrinunciabile valore aggiunto. Tecnicamente mediare significa mettere in gioco la descrizione, tema di lunghissima e accidentata durata che per gli archivi muove dal metodo storico e dalle sue piroette per arrivare agli standard di prima e seconda generazione, con tutto ciò che ne deriva. L'archivista, si sa, è mediatore per definizione ed è destinato a presidiare molti confini. Nel più ampio panorama delle discipline documentarie, però, l'archivista non è l'unico mediatore. Mediazione e intermediazione sono valori altrettanto fondanti di molteplici metodologie e prassi genericamente documentarie. Queste pulsioni sono il vero cuore del nostro lavoro e della nostra ricerca. Ammetterlo non è difficile, più complicato risulta invece risolvere il problema della mediazione in rapporto ai bisogni e alle abitudini della ricerca contemporanea. La crescita incontenibile di risorse digitali buone e cattive, ulteriormente sospinta dai ripetuti eremitaggi pandemici, è una tentazione molto forte. Essa si manifesta ormai compiutamente, con le sue opportunità e le sue criticità. La ricerca, a partire da quella storica, di fronte a questo apparente uovo di Colombo è sempre più tentata dalle scorciatoie digitali, da un tutto e subito che rischia di far saltare il banco di un più meditato rigore interpretativo. Se andiamo oltre la retorica dell'onniscienza a portata di clic sembra di poter dire che dematerializzare la mediazione, il tratto più fisico e inclusivo del mestiere di archivisti e bibliotecari, è un'indiscutibile opportunità, direi un dovere, giunti a questo punto. Bisogna però essere cauti e cercare disperatamente di riprodurre contesti che diano qualità a ciò che si dematerializza. Nella rete ci si può imbattere in archivi senza archivisti e biblioteche senza bibliotecari. Sarà più facile trovarli ma anche più arduo interpretarli correttamente: non è importante come cerco o quanto trovo ma cosa ottengo, e non solo in termini di quantità informativa. Il punto focale non coincide con la maggiore reperibilità ma sta nella qualità dei risultati di ogni ricerca e della possibilità di usarli

consapevolmente. Questo è particolarmente vero per la ricerca archivistica, soprattutto perché i metodi di organizzazione dell'informazione negli archivi, a cominciare dal metodo storico, si basano su un processo di approssimazione al dato che deve tenere conto di componenti che possono sfuggire ad un uso meccanico dei sistemi. La descrizione digitale si carica infatti di un peso ancora maggiore di quello che sostiene in ambiente analogico. Deve tentare di riprodurre un'intermediazione fisica, potremmo dire intuitiva ed empatica, che altrimenti sarà perduta.

L'informazione che sottoponiamo a trattamenti massicci di intermediazione digitale deve fornire anche e soprattutto garanzie qualitative. I grandi sistemi informativi archivistici, fino alle loro manifestazioni federate come SAN, hanno dato risposte esaurienti in questo senso, collocando dentro opportuni modelli di restituzione, santificati da ISAD, le entità auto esplicative e i loro sistemi di relazioni. Come abbiamo visto l'uso dei metadati che si fa proprio in SAN, per quanto perfettibile, dimostra come la ricontestualizzazione digitale sia possibile ed anzi ineludibile.

L'autorevolezza, al di là dei bollini scientifici e istituzionali, mi sembra risieda nella capacità che abbiamo di contestualizzare efficacemente gli oggetti. In una logica di sistemi integrati si può fare affidamento sull'incrocio tra 'vocabolari di cose' che arrivano direttamente dalle universitas rerum che li hanno generati nel rispetto dei propri canoni e che dai canoni ereditano autorevolezza e affidabilità. In questo senso gli archivi o le biblioteche non producono più soltanto descrizioni interne ma generano stimoli informativi molto potenti, che stanno alla base della rappresentazione integrata e multidimensionale.

#### 4. Classificare l'infinito? L'ipotesi della metacatalogazione

Abbiamo molte frecce al nostro arco, si tratta di capire meglio quale sia la distanza e la natura dei bersagli, senza dimenticare che nessuna intelligenza, umana o artificiale, potrà mai davvero prosciugare il fiume della conoscenza. Capita a proposito, nella sua urticante paradossalità, quanto scrisse a suo tempo Fernando Pessoa: "I classificatori di cose, che sono quegli uomini di scienza la cui scienza consiste solo nel classificare, ignorano in genere che il classificabile è infinito e che dunque non si può classificare" (Pessoa 2020).

Non possiamo classificare perché l'infinito ci sfugge, eppure vogliamo farlo ad ogni costo, perché l'infinito ci fa paura e le nostre descrizioni forse sono solo battaglie contro l'*horror vacui* della scheda bianca. Più prosaicamente, però, classificare è la nostra professione e non ammetteremo mai che sia impossibile. Per questa ragione ogni disciplina documentaria nei secoli si è battuta strenuamente per tracciare l'informazione, cullandosi nell'utopia di poterla domare. Le diverse tassonomie si sono nei secoli sovrapposte, incrementate, arenate. Sono resuscitate, si sono trasformate e sono arrivate ancora agguerrite alle soglie della modernità digitale che le ha ulteriormente elaborate, lasciandole in qualche caso stordite ad osservare sé stesse.

Oggi però disponiamo di calcolatori potenti, a cui possiamo chiedere di fare almeno una parte del lavoro sporco, in cerca di possibili sintesi dei nostri saperi pragmatici dentro all'utopia digitale, la più performante di tutte le utopie mai vagheggiate.

La sintesi va cercata oltre alla catalogazione, alla metadattazione e alla descrizione<sup>7</sup>. Può diventa-

---

<sup>7</sup> Al riguardo si vedano le considerazioni di Guerrini 2020.

re metacatalogazione, un gesto che va oltre la catalogazione senza dimenticarla. In questo senso quindi l'idea di metacatalogazione può coniugare evoluzioni tecniche e tecnologiche con evidenti persistenze metodologiche, ponendosi l'obiettivo di concettualizzare la conoscenza ancora prima di classificarla per diffonderla.

La metacatalogazione, neologismo di sapore vagamente postmoderno, è la possibile etichetta da collocare sul vaso di Pandora di una descrizione sempre più ricca e potente e per questo sempre più complicata e diversa da sé stessa. Con questo termine ci si vuole riferire all'insieme dei concetti, degli strumenti e delle prassi che sostanziano oggi le attività di catalogazione, metadattazione e descrizione archivistica.

Metacatalogazione è una parola che accoglie nella sua definizione approcci metodologici di lunga durata, in direzione dell'inevitabile superamento degli steccati di dominio indotto dalle trasformazioni profonde che attraversano le discipline documentarie.

Non si intende mettere in discussione la centralità e la peculiarità di determinati schemi descrittivi ma si prende atto che esiste il rischio che tali schemi, soprattutto nel rimescolamento digitale, finiscano col sovrapporsi agli stessi oggetti della descrizione. Abbiamo ormai sistemi di metadattazione di rara complessità ed efficacia ma c'è il rischio che nell'accanimento terapeutico sotteso ai nostri metadati si perda di vista la natura di ciò che i metadati si affannano a catturare.

Tra 'catalogazione e la metadattazione' devono mantenersi ben saldi gli elementi descrittivi e i loro parametri. Rimane essenziale costruire adeguate rappresentazioni basiche delle singole unità di descrizione perché le cose devono avere nomi e attributi identificabili. Altro è dire, poi, che la combinazione delle descrizioni possa rompere gli argini di una tassonomia monodimensionale per aprirsi alla molteplicità di possibili contesti cui comunque appartiene e che contribuisce a definire dinamicamente.

L'idea di una descrizione costruita attraverso combinazioni di metadati permette infatti di dare spessore alle pagine, di collocare le parole simulacro dei fatti nel tempo e nello spazio, in un processo descrittivo puntuale e dinamico, capace di integrare le singole risorse in contesti più ampi. I modelli concettuali sottesi alle tecnologie che possiamo utilizzare per la descrizione, almeno in ambito archivistico, vanno oltre l'identificazione delle entità, enfatizzando sistemi di relazioni sempre più larghi e suggerendo di affiancare alla consolidata multilivellarietà una multidimensionalità capace di rendere 'visibili' le idee o i fatti di cui le diverse entità sono veicoli. Le logiche del *deep learning* sembrano poter contribuire efficacemente a quel complicato processo neuronale che è da sempre la descrizione archivistica, intesa come combinazione tendenzialmente inesauribile di entità descrittive desunte dalla realtà della rappresentazione. Indipendentemente dai suoi limiti e dalle sue reali ricadute, del resto, lo standard RiC riconosce un ruolo forte alle ICT nel quadro della descrizione archivistica. I modelli descrittivi in questo approccio sono orientati dalle tecnologie disponibili, con un salto logico e applicativo rispetto agli standard di prima generazione che, pur avendo di fatto indirizzato e consentito le applicazioni di ICT agli archivi, erano esplicitamente liberi da suggestioni tecnologiche.

Tutto il resto sarà intelligenza. Artificiale?<sup>8</sup>

---

<sup>8</sup> Sul tema, che merita ben altri approfondimenti basterà qui ricordare InterPARES Trust AI: <https://interparestrustai.org/>.

## Riferimenti bibliografici

(Ultima consultazione dei siti web: 26 apr. 2022)

- Alfier, Alessandro. 2020. *Il sistema di documentazione digitale*. Milano: Editrice Bibliografica.
- Bergamin, Giovanni, e Mauro Guerrini. 2021. «The Paris Principles 60 years later. Conversation with Diego Maltese». *JLIS.it* 12 (3): I–IV. <https://doi.org/10.4403/jlis.it-12769>.
- Bruni, Silvia, Francesca Capetta, Anna Lucarelli, Maria Grazia Pepe, Susanna Peruginelli, e Marco Rulent. 2016. «Towards the integration of archives, libraries and museums». *JLIS.it* 7 (1): 225–44. <https://jlis.it/index.php/jlis/article/view/183>.
- Cencetti, Giorgio. 1978. *Paleografia latina*. Roma: Jouvence.
- Ciandrini, Paola. 2020. *Records management : ISO 15489: progettare sistemi documentali*. Milano: Editrice Bibliografica.
- Colavizza, Giovanni, Tobias Blanke, Charles Jeurgens, e Julia Noordegraaf. 2021. «Archives and AI: An Overview of Current Debates and Future Perspectives». *Journal on Computing and Cultural Heritage* 15 (1): 1–15. <https://doi.org/10.1145/3479010>.
- Connerton, Paul. 2010. *Come la modernità dimentica*. Torino: Einaudi.
- Duranti, Luciana, e Corinne Rogers, a c. di. 2019. *Trusting Records and Data in the Cloud. The Creation, Management, and Preservation of Trustworthy Digital Content*. Cambridge: Facet Publishing.
- Feliciati, Pierluigi. 2009. «I requisiti di fattibilità di un sistema informativo archivistico: modelli organizzativi, informatici e soddisfazione degli utenti». *Archivi* 1: 13–32.
- Feliciati, Pierluigi, Daniela Grana. 2005. «Dal labirinto alla piazza. Il progetto “Sistema Informativo degli Archivi di Stato”». *Scrinia* II (2–3): 9–18.
- Fortino, Giancarlo, Anna Rovella, Wilma Russo, e Claudio Savaglio. 2016. «Towards Cyberphysical Digital Libraries: Integrating IoT Smart Objects into Digital Libraries». In *Management of Cyber Physical Objects in the Future Internet of Things. Internet of Things*, a cura di Antonio Guerrieri, Valeria Loscri, Anna Rovella, e Giancarlo Fortino, 135–56. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-26869-9\\_7](https://doi.org/10.1007/978-3-319-26869-9_7).
- Gardini, Stefano. 2021. «Economie circolari dell’archivio: la carte di utenti e studiosi come archivi derivati». *Nuovi Annali della Scuola speciale per archivisti e bibliotecari XXXV*: 237–77.
- Giglio, Daniela. 2021. «Biblioteca: digitale, elettronica o virtuale? Evoluzione terminologica e definizioni». In *La trasmissione della conoscenza registrata*, a cura di Carlo Bianchini e Lucia Sardo, 235–44. Milano: Editrice Bibliografica.
- Guarasci, Roberto. 2008. *Dal documento all’informazione*. Milano: ITER.
- Guarasci, Roberto, e Mauro Guerrini. 2022. *Cos’è l’indicizzazione*. Milano: Editrice Bibliografica.
- Guerrini, Mauro. 2020. *Dalla catalogazione alla metadatoazione. Tracce di un percorso*. Roma: Associazione Biblioteche.

Humila, Isto. 2022. «Improving the usefulness of research data with better paradata». *Open Information Science* 6: 28-48. <https://doi.org/10.1515/opis-2022-0129>.

Lambroni, Giovanna. 2021. «Descrizione e accesso ai beni culturali dai principi di Parigi a FR-BRoo». In *La trasmissione della conoscenza registrata*, a cura di Carlo Bianchini e Lucia Sardo, 357–67. Milano: Editrice Bibliografica.

Le Goff, Jacques. 1984. *Storia e memoria*. Torino: Einaudi.

Marrè, Giovanni, e Anna Rovella. 2008. «Strutturazione dell'informazione e integrazione della conoscenza». *AIDA Informazioni* 1 (2): 265–74.

Penzo Doria, Gianni. «A new archive definition». *JLIS.it* 13 (2): 156–173. <https://doi.org/10.36253/jlis.it-465>.

Pessoa, Fernando. 2020. *Il libro dell'inquietudine di Bernardo Soares*. Milano: Feltrinelli.

Pigliapoco, Stefano. 2016. *Progetto archivio digitale. Metodologia sistemi professionalità*. Torre del lago: Civita editoriale.

Schiller, Federico. 1927. *Lettere sull'educazione estetica e altri scritti*. Firenze: Sansoni.

Tomasi, Francesca. 2022. *Organizzare la conoscenza: digital humanities e web semantico*. Milano: Editrice Bibliografica.

Valacchi, Federico. 2015. «I sistemi informativi archivistici tra locale, nazionale e internazionale». In *Archivistica. Teorie, metodi, pratiche*, a cura di Linda Giuva e Maria Guercio, 357–80. Roma: Carocci.

## Entity modeling: traces of an evolving path

Tiziana Possemato<sup>(a)</sup>

a) @Cult, <https://orcid.org/0000-0002-7184-4070>

**Contact:** Tiziana Possemato, [tiziana.possemato@atcult.it](mailto:tiziana.possemato@atcult.it)

**Received:** 26 May 2022; **Accepted:** 3 June 2022; **First Published:** 15 September 2022

### ABSTRACT

In this work we will deal with the subject of entities understood as real-world objects and how this concept is used in the context of entity modeling, that process of identification and modeling of entities that plays so much part in projects of conversion of catalogs into linked open data. The help in understanding what this concept expresses within the bibliographic universe comes from object-oriented programming, which introduces the concept of modeling and management of an “object” by defining its state and behavior. But to model an object it is necessary to identify it and this process must often take place dealing with massive amounts of data, not necessarily homogeneously structured: the Entity Resolution is this set of machine processes that tries to resolve the ambiguities given by the inhomogeneity of the descriptions referable to the same entity. The adoption of these practices, in the bibliographic field, still moves the horizon of the cataloging action, which had already extended towards the more general metadating, towards that web of data that imposes a new way of understanding objects and treating them: entity modeling promises to be the third generational step in the management of bibliographic data.

### KEYWORDS

Real world object; Entity; Entity resolution; Entity modeling.

## Entity modeling: tracce di un percorso in evoluzione

### ABSTRACT

In questo lavoro tratteremo il tema delle entità intese come oggetti reali del mondo (real-world object) e di come questo concetto sia utilizzato nell’ambito dell’entity modeling, quel processo di identificazione e modellamento delle entità che tanta parte occupa nei progetti di conversione dei cataloghi in linked open data. L’aiuto alla comprensione di cosa questo concetto esprima nell’ambito dell’universo bibliografico ci viene dalla programmazione orientata agli oggetti, che introduce il concetto di modellamento e gestione di un “oggetto” definendone uno *stato* e un *comportamento*. Ma per modellare un oggetto è necessario identificarlo e questo processo deve avvenire, spesso, trattando quantità imponenti di dati, non necessariamente omogeneamente strutturati: l’Entity Resolution è questo insieme di processi macchina che cerca di risolvere le ambiguità date dalla disomogeneità delle descrizioni riferibili alla medesima entità. L’adozione di queste pratiche, in ambito bibliografico, muove ancora l’orizzonte dell’azione catalografica, che già si era esteso verso la più generale metadating, verso quel web di dati che impone un nuovo modo di intendere gli oggetti e di trattarli: l’entity modeling si annuncia come il terzo passaggio generazionale nella gestione del dato bibliografico.

### PAROLE CHIAVE

Real world object; Entità; Entity resolution; Entity modeling.



## Cos'è un Real World Object

Nel suo blog Coyle's InFormation, Karen Coyle pubblica il 16 febbraio 2015 un post dal titolo Real World Objects (RWO), in risposta ad una domanda ricevuta sul significato e l'importanza del concetto di real world object<sup>1</sup>. Nella lista di discussione di BIBFRAME Coyle richiama il post sul RWO definendolo "this mysterious RWO thing"<sup>2</sup> e prova a dare una spiegazione, o meglio ad aprire una possibile riflessione sul tema. L'origine del termine e del concetto è fissato nell'ambito dell'Artificial Intelligence (AI) ed è spiegato in modo chiaro: immaginiamo di dover produrre dei robot che vivano nel nostro stesso mondo, con i quali comunicare parlando di qualsiasi cosa che faccia parte del nostro mondo, sia esso un oggetto fisico o un concetto astratto. Immaginiamo, dunque, di dover stilare un elenco di cose del mondo e immaginiamo di doverlo fare in una modalità condivisibile con le macchine, e comprensibile da esse. Questa è l'ambizione fissata dall'AI e dal web semantico: poter creare e sfruttare, per dialogare con le macchine, delle raccolte (ontologie), in cui sia descritta qualsiasi cosa esistente nel mondo. Per cogliere il significato del termine real world object nell'ambito dell'AI, Coyle usa un esempio relativo al mondo bibliografico: se analizziamo un qualsiasi record bibliografico, scopriamo che i campi del record possono descrivere l'oggetto "pubblicazione", ma anche l'oggetto "record" in sé. La pubblicazione è un RWO (l'oggetto che vorrei avere tra le mani e leggere), ma anche il record è un RWO (quella descrizione testuale che io potrei richiedere ad un'altra biblioteca per migrarla nel mio catalogo). Uno dei problemi principali del record Marc in termini di capacità di veicolare un insieme di messaggi comprensibili ad una macchina, è proprio quello di non riuscire a distinguere "our metadata and the thing it describes" (Coyle 2015).

Ma come mai Karen Coyle si interroga sul real world object? Non era un termine già utilizzato in ambito bibliografico?

## Il Real World Object nella disciplina catalografica

L'uso del termine real world object in ambito catalografico è particolare perché esiste e coesiste in una doppia accezione: quella ereditata dal mondo dell'AI, del web semantico e della programmazione a oggetti, e quella più tradizionalmente radicata in ambito catalografico, che utilizza il termine real world object come sinonimo di *materiale non bibliografico*, *realia*, *materiale non librario* (non-book material), per indicare oggetti tridimensionali come monete, strumenti e tessuti, posseduti dalle biblioteche ma che non rientrano nelle categorie tradizionali del materiale librario. Contestualmente a questo significato, però, il termine real world object viene ufficialmente e definitivamente assorbito nel gergo biblioteconomico: nel 2015 viene costituito il PCC Task Group on URIs in MARC (URI TG) con l'obiettivo di identificare e indirizzare le scelte per l'arricchimento del record MARC con identificatori (URI). Il lavoro del gruppo pone le basi per facilitare la transizione dei dati MARC verso i linked data: tra le diverse proposte e soluzioni quella di includere

<sup>1</sup> La sintassi utilizzata per questo termine non è omogenea: Real-World Object, real-world object e real world object sono le diverse formulazioni sintattiche più frequentemente utilizzate nelle diverse fonti consultate.

<sup>2</sup> <LISTSERV 16.0 - BIBFRAME Archives (loc.gov)>.

un nuovo sottocampo \$1 per la registrazione dell'URI del RWO<sup>3</sup> nei campi relativi a nomi, titoli, soggetti, classificazioni etc. Nel 2017 è pubblicato il documento MARC Proposal no. 2017-08 dal titolo: Use of Subfield \$0 and \$1 to Capture Uniform Resource Identifiers (URIs) in the MARC 21 Format. Scopo della proposta è quella di delineare un criterio per arricchire i record bibliografici e di authority con gli identificatori URI in un modo che siano chiaramente distinguibili:

- URI che identificano un record che descrive un oggetto (URI come puntatore ad una descrizione);
- URI che identificano l'oggetto stesso.

A tal fine, il documento propone di limitare l'uso del sottocampo \$0 agli URI e ai numeri di controllo che si riferiscono *al record che descrive una cosa* e di definire un nuovo sottocampo, \$1, per gli URI che si riferiscono direttamente al *Real World Object*. La proposta diventa operativa e l'Appendix A - Control Subfields del MARC 21 Bibliographic Format ufficialmente introduce questa importante distinzione tra identificatori per descrizioni e identificatori per RWO.

Questa distinzione, in qualche modo, chiude il cerchio rispondendo a quella perplessità chiaramente espressa da Karen Coyle, ma comune a molti esperti di metadattazione e di modellazione dei dati, rispetto alla mancata distinzione, nel record bibliografico e di authority, tra “our metadata and the thing it describes”.

## Real World Object e Real World Entity

A complicare l'intero quadro terminologico si aggiunge l'uso del termine Real World Entity spesso in alternativa o in concomitanza con il termine Real World Object. Il Real World Entity è definito su alcuni siti come un'entità con una posizione fisica all'interno dell'universo, quindi con una definizione che, in sé, non lascia intendere nulla di diverso rispetto al Real World Object: *object* e *entity* sono usati come alternativi l'uno all'altro. E lo stesso accade in molta della letteratura che richiami queste tematiche, in cui i termini *real world object* e *real world entity* sembrano rimandare allo stesso concetto. Tuttavia, nonostante il nuovo rimescolamento di carte che l'uso di questa terminologia, nei termini descritti, sembra produrre, la traccia da seguire per distinguere la sottilissima linea di demarcazione semantica tra questi termini sembra essere stata individuata: sappiamo che viviamo in un mondo costituito da entità di vario tipo (“cose” fisiche e concettuali), sappiamo che spesso queste entità sono estremamente complesse e sappiamo che nessun sistema che voglia rappresentarle potrà avere l'ambizione di esprimerle nella loro interezza: ci sarà sempre un aspetto di un'entità che non sia facile cogliere o che il contesto per il quale quella “cosa” sia descritta non richieda di essere rappresentata. L'ipotesi che stiamo provando a verificare è che nelle espressioni *real world entity* e *real world object* ci possa essere la stessa relazione semantica che possiamo rilevare nel binomio *entità e identità*.

---

<sup>3</sup> Per una chiara e completa disamina sui lavori del gruppo PCC Task Group on URIs in MARC si rimanda all'articolo di Jackie Shieh, dal titolo PCC's Work on URIs in MARC, pubblicato nel 2019 in *Cataloging & Classification Quarterly* (Shieh 2020).

## Il Real World Object nell'Object Oriented Programming

La programmazione orientata agli oggetti, o più semplicemente programmazione ad oggetti, si basa sulla definizione di *classi* che contengono la dichiarazione delle strutture dati e le procedure che operano su di esse. La classe costituisce un modello o un progetto per quelli che poi saranno gli *oggetti* che deriverò da essa. Una sorta di template che definisce la forma dell'oggetto, utilizzato per poi modellare i singoli oggetti. Una classe definisce qualcosa in termini di:

- *stato*: le variabili che costituiscono quel tipo di cosa
- *comportamento*: i comportamenti (metodi, procedure) che ha quel tipo di cosa.

Questo modello definito dalla classe viene poi istanziato nei diversi oggetti, istanze, create a partire dalla medesima classe. Nel contesto dell'OOP un oggetto software viene creato a partire (e come rappresentazione di) un oggetto reale del mondo, un *real world object*: il programmatore è chiamato a riprodurre in una dimensione diversa (quella del software, appunto) il mondo reale, con i suoi tanti oggetti (siano essi fisici o concettuali). E così come gli oggetti del mondo reale, quindi i *real world object*, hanno tutti uno *stato* (degli attributi che li identifichino) e un *comportamento* (la capacità di fare qualcosa) allo stesso modo, gli oggetti software devono poter essere modellati con uno stato e un comportamento. Ma perché rappresenta un *real world object* e non un *real world entity*? La programmazione orientata agli oggetti dichiara, in modo piuttosto esplicito, la difficoltà o addirittura l'impossibilità a rappresentare in un oggetto software la complessità di un'entità: l'oggetto software è la rappresentazione di un *real world object* e non di un *real world entity* perché l'entità, nella sua ricchezza, non potrà mai essere rappresentata in modo esaustivo. L'OOP costruisce oggetti che rappresentino una particolare "faccia" dell'entità, e solo quella, adatta e funzionale allo specifico contesto in cui venga utilizzata. Questo è uno dei pilastri dell'OOP, chiamato *astrazione*: l'informatica non ha l'ambizione di rappresentare tutte insieme le mille caratteristiche di una entità, ma *seleziona* quelle adatte al contesto in cui quell'oggetto debba essere calato (nascondendo o ignorando del tutto le tante altre caratteristiche - di stato e comportamento - ascrivibili a quella medesima entità). Questo principio è collegato con un altro dei pilastri dell'OOP: il *polimorfismo*, che è definito come la capacità di usare lo stesso nome per fare differenti cose, oppure, più chiaramente, la capacità di rappresentare *molte forme della singola entità*.

## L'Entity Resolution e l'Entity Modeling

Il meccanismo del "creare" un oggetto, mettendolo in relazione con altro, così tipico di questo paradigma di programmazione, non è affatto diverso da quel nuovo modo di intendere l'attività del catalogatore nell'universo bibliografico, sempre più orientato alla identificazione delle entità che partecipano a questo universo, e alla loro "modellazione". Modellare un oggetto significa, dunque, individuare quei tratti salienti che lo rendono ciò che è, e che gli consentono di essere o fare qualcosa nel mondo, o per lo meno in un determinato contesto. Ma se concordiamo con l'immagine di un'entità complessa che sia presente in forme diverse sulla medesima fonte informativa o in fonti differenti, possiamo immaginare quanto complesso sia il meccanismo dell'identificazione dell'entità. Rinunciare all'ambizione di descrivere nella sua interezza la complessità dell'entità non significa rinunciare anche all'ambizione di *identificarla* nelle sue tante espressioni: l'informatica

accetta il limite di non poter rappresentare in un unico oggetto la complessa personalità di Lewis Carroll, e probabilmente costruirà due o più “oggetti” (uno come Charles Lutwidge Dodgson, autore dell’opera *The game of logic* (1887) e uno come Lewis Carroll, autore del celebre *Alice’s Adventures in Wonderland*). Ma in una realtà più complessa, come quella del web semantico e dell’intelligenza artificiale, possiamo rinunciare all’idea di riconoscere dietro questi due “profili” la medesima entità? Questo è esattamente il compito dell’Entity Resolution (abbreviato ER), quello di identificare tutte le menzioni che rappresentano la stessa entità all’interno della stessa base di conoscenza oppure in basi di conoscenza multiple (Zhu et al. 2016).

L’Entity Resolution (*risoluzione delle entità*) è il processo che risolve le entità e rileva le relazioni utili a identificarle. Il processo nel suo insieme genera un database di entità, in cui, dunque, i dati che identifichino una entità sono registrati per consumo (ricerche, statistiche etc.) o anche per iterare i processi di “entificazione” e renderli sempre più efficaci.

Il punto di partenza, dunque, per i processi di Entity Resolution, è costituito da un insieme di informazioni che esprimono l’identità di un particolare real world entity: questa unità informativa, nell’ambito dell’ER, viene spesso definita *profilo* e a noi ricorda molto il risultato di quella *astrazione* che abbiamo trovato come pilastro dell’OOP.

In un contesto informativo sempre più ampio, come quello che il web semantico propone, in cui le basi di dati sono eterogenee e non necessariamente autorevoli, i profili sono spesso sporchi, incompleti, incorretti o ridondanti: il successo dell’ER è quello di riuscire a identificare e integrare profili anche molto diversi ma che rimandino, in realtà, alla medesima entità.

Se definiamo due profili relativi alla stessa entità, presenti nella medesima fonte dati, come “duplicati”, allora possiamo dire che una fonte dati pulita è quella in cui non esistano duplicati, mentre una fonte dati è sporca quando esistano profili duplicati. In realtà la classificazione di una fonte è ben più complessa di questa esemplificazione: nella stessa fonte di dati si possono trovare casi di entità molto ben modellate e altre con molteplici profili. La “veracità” di una fonte è un parametro complesso che si basa su numerosi algoritmi, e che affida una percentuale piuttosto alta di successo alla qualità del dato di origine.

L’obiettivo dei processi che puntano alla veracità del dato è quello di *collegare* profili o descrizioni di entità diverse (*record linkage*) e deduplicare, dunque individuare i duplicati e risolverli.

I complessi processi di ER raramente si affidano a singoli dati, o a profili poveri: più il profilo utilizzato è ricco di caratteristiche, più attributi sono espressi a definire una particolare identità dell’oggetto descritto, più possibilità si avranno di identificarlo. Ma perché una macchina possa identificare un oggetto, deve prima di tutto conoscerne i contorni: deve sapere cosa cercare e come si aspetta di trovare quell’oggetto. Deve sapere, dunque, come quella cosa sia *modellata*. La costruzione dell’oggetto, la definizione del profilo in una modalità che sia rispondente alle esigenze prima di tutto di identificazione, è un’attività della massima importanza, soprattutto in un contesto aperto e tremendamente sconfinato come quello del web. E per intendere cosa sia l’entity modeling, ripartiamo da quell’oggetto che l’OOP crea attribuendo uno *stato* e un *comportamento*, e che il web semantico riformula, arricchisce in modo da renderlo non solo usabile, ma anche comprensibile alle macchine. Il meccanismo di creazione di un oggetto nell’OOP parte dalla definizione di un modello, quanto abbiamo definito come una sorta di *matrice* o *template* e che serve per creare tante istanze che da quel modello prendono, appunto, la forma. Queste matrici (le classi, con le loro caratteristiche di stato e comportamento) possono essere costruite specifica-

tamente per il singolo oggetto, nell'ambito di uno specifico progetto (il che rende quegli oggetti poco "usabili" al di fuori del contesto originario) oppure possono essere definite a livello globale, come risultato di un accordo che una comunità decida di sottoscrivere e condividere. Le ontologie e i data model, in tutti i contesti in cui siano utilizzati, sono esattamente questo: una matrice che definisce la forma che avranno le rappresentazioni di un real world object nella maniera che sia il più possibile fedele all'entità rappresentata, in accordo con quanto definito da una comunità.

## I modelli e le ontologie dell'universo bibliografico

Non potendo rappresentare l'entità nella sua interezza, abbiamo concordato di scegliere un criterio di *utilità* per filtrarne o selezionarne i caratteri utili ad un determinato fine: per quali scopi rappresento questa entità? Nel contesto della disciplina catalografica la definizione dei principi e delle regole catalografiche, la scelta delle intestazioni (poi diventati access point), la definizione delle entità da rappresentare, sono tutti fattori strettamente collegati agli obiettivi e funzioni del catalogo che, già a partire da Cutter, quindi già alla fine del XIX secolo, sono stati fortemente influenzati dalle *esigenze dell'utente* (*user's tasks*) Nella sua opera *Rules for a Printed Dictionary Catalogue* Cutter definisce gli obiettivi del catalogo proprio in funzione dell'utilità per gli utenti (Cutter 1876). Questo criterio delle esigenze dell'utente ha guidato tutte le successive revisioni dei modelli e delle regole catalografiche nonché la definizione delle entità necessarie a supportare tali bisogni.

L'input a definire principi, regole, modelli e infine ontologie viene, dunque, prima di tutto da quel che gli utenti si aspettano di vivere nella loro esperienza di utilizzo del catalogo. Per lo meno in ambito bibliografico chi si pone di fronte al compito di definire dei modelli catalografici, di modellare la conoscenza, lo fa avendo ben chiaro in mente quali siano gli obiettivi di ciascun elemento che entri a far parte di quel modello.

## Un esempio di entity modeling: l'Opus in Share-VDE

L'iniziativa Share-VDE (Share Virtual Discovery Environment),<sup>4</sup> che riunisce, arricchisce e integra i cataloghi bibliografici e di authority di una vasta comunità di biblioteche in un ambiente condiviso basato su linked data, ha al proprio interno diversi gruppi di lavoro,<sup>5</sup> che affrontano alcuni ambiti più complessi o per i quali un'analisi ulteriore e specializzata sia particolarmente richiesta. I gruppi sono coordinati da un *Advisory Council* che traccia il percorso dell'iniziativa e coordina le diverse anime che vi partecipano, nonché i rapporti e le relazioni con iniziative e gruppi affini. Il gruppo *Sapientia Entity Identification Working Group - SEIWG*, che si occupa di tutti i temi relativi al modellamento dei dati, ha lavorato alla revisione del modello dati di Share per renderlo più vicino alle esigenze di ricercabilità e identificazione delle risorse, soprattutto in un contesto così esteso e ricco come quello prodotto dall'iniziativa. Il percorso di analisi e studio ha avuto come risultato l'identificazione di una nuova entità, l'*Opus*, non già presente nel modello BIBFRAME

<sup>4</sup> <[https://wiki.share-vde.org/wiki/Main\\_Page](https://wiki.share-vde.org/wiki/Main_Page)>.

<sup>5</sup> <[https://wiki.share-vde.org/wiki/ShareVDE:Members/Share-VDE\\_working\\_groups#SVDE\\_Advisory\\_Council\\_28AC.29](https://wiki.share-vde.org/wiki/ShareVDE:Members/Share-VDE_working_groups#SVDE_Advisory_Council_28AC.29)>.

che la comunità di Share adotta come ontologia principale per la resa dei propri cataloghi in linked open data. Il percorso di analisi e di proposte e ipotesi di soluzione è stato piuttosto lungo e ha visto il coinvolgimento di molte biblioteche nordamericane e europee nella discussione intorno alla possibile estensione del modello BIBFRAME.

## Il punto di partenza: l'ontologia BIBFRAME

Share-VDE è un'iniziativa che nasce con lo scopo di accompagnare e supportare le biblioteche in questo delicato momento di transizione dal record Marc ai modelli basati sulle entità, con particolare attenzione a BIBFRAME, che è stato adottato, appunto, come ontologia centrale nei processi di conversione dei dati in linked open data. La presenza di tante biblioteche americane, a partire dalla Library of Congress, ha orientato fortemente la scelta di adozione di BIBFRAME, anche se l'iniziativa partiva dalle suggestioni offerte da un progetto tutto italiano, quello delle biblioteche universitarie del sud d'Italia, chiamato Share Catalogue, che aveva già operato questa scelta molto determinata di adozione di BIBFRAME. D'altra parte, l'ingresso nella comunità di Share-VDE della National Library of Norway, della National Library of Finland e della British Library, tutte biblioteche con un profondo legame con il contesto culturale europeo e con l'IFLA, ha suggerito una importante riflessione in merito alla compatibilità del modello BIBFRAME con quello definito dall'IFLA come armonizzazione della famiglia FR, quindi con l'IFLA Library Reference Model (abbreviato IFLA LRM).

Fino al giugno 2021, con l'introduzione ufficiale dell'entità Hub, il modello BIBFRAME nella versione 2.0 organizzava le informazioni utili a modellare l'universo bibliografico in tre livelli principali di astrazione:

- **Work:** contenuto intellettuale di una risorsa.
- **Instance:** una singola manifestazione materiale di un Work.
- **Item:** il singolo esempio di un'Istanza.

Queste entità principali sono corredate di una serie corposa di entità relazionate (tra cui, ovviamente, quella degli Agenti).

Di contro, il modello IFLA LRM eredita i livelli delle varie declinazioni della famiglia FR e sviluppa un modello concettuale ben più complesso, che comunque ripropone come entità principali quelle di FRBR:

- **Work:** il contenuto intellettuale o artistico di una determinata creazione.
- **Expression:** una determinata combinazione di segni che veicola un contenuto intellettuale o artistico.
- **Manifestation:** un insieme di tutti i supporti che si presume condividano le stesse caratteristiche per quanto riguarda contenuto intellettuale o artistico e aspetti di forma fisica.
- **Item:** un oggetto o oggetti che recano segni destinati a veicolare un contenuto intellettuale o artistico.

La mancanza di un elemento apicale in BIBFRAME che indirizzi l'utente su una possibile opera originale, quell'opera creativa che era nella mente del suo autore e che si è poi espressa in molti modi differenti, non può essere risolta con la proprietà "translation of" di BIBFRAME perché questa proprietà può indistintamente relazionare un'opera originale e la sua traduzione così come

una traduzione in una lingua ma a partire da un'altra espressione non originale e dunque da un'altra traduzione. Il che pone tutti i Work di BIBFRAME sulla medesima linea orizzontale, senza alcuna indicazione di un'opera originaria.

Ma il problema di un'opera con una storia editoriale molto ricca in cui si perda traccia del punto di partenza da cui il reticolo espressivo ha poi preso forma si sente molto e non è un caso che IFLA LRM introduca il concetto di *Espressione rappresentativa* che dovrebbe essere, qualora identificabile, la prima espressione di un'opera, dunque quella che meglio esprima l'opera originale, quella più vicina a quanto idealmente pensato dal suo creatore.

Il risultato delle prime riflessioni del gruppo SEI su questo tema e i tanti confronti con esperti di dominio, conduce, nel gennaio 2019, ad una prima ipotesi di modellamento di un'entità, inizialmente definita *SuperWork*, che potesse raccogliere sotto di sé tutte le diverse espressioni relazionate, qualcosa, dunque, molto vicino al concetto di Work di FRBR e di IFLA LRM. Ma l'introduzione di una nuova entità come estensione, locale o ufficiale, di un modello riconosciuto da una comunità - BIBFRAME - potrebbe generare, prima di tutto, un problema di interoperabilità. Problema sentito fortemente anche da tutte le biblioteche interessate a modellare i propri dati secondo l'ontologia BIBFRAME ma non disposte a perdere la possibilità di un dialogo e uno scambio agile di dati e servizi con l'altrettanto riconosciuta comunità riferentesi al modello IFLA LRM.

Intanto, la Library of Congress, anche seguendo le suggestioni date dalle conversazioni e i confronti tenuti nel gruppo di lavoro del SEI, proponeva, prima solo in test (nel giugno 2019) e poi in ambiente di produzione (McCallum 2022), quell'elemento "aggregante" costituito dall'Hub, definito come *una risorsa astratta che funziona come collegamento tra due Work*.

A questo punto gli elementi per una comparazione puntuale tra i tre modelli ontologici, quello di Share-VDE con le proposte modifiche al modello originale, quello di BIBFRAME e quello di IFLA LRM erano maturati, dando così il via ad un'attenta valutazione comparativa ai fini, appunto, di garantire una più efficace risposta al bisogno informativo dell'utente ma anche una piena compatibilità con i due maggiori e più diffusi modelli bibliografici. Nel gennaio 2020, anche per sancire la funzione semantica attribuita al livello apicale definito *SuperWork* e per evitare confusioni con altre definizioni passate di "superwork", il gruppo del SEI rinomina questa nuova classe come *Share-VDE Opus* (svde:Opus).

## La definizione della classe svde:Opus e il suo modellamento

Definita a livello teorico la classe svde:Opus, bisognava renderla operativa, capace, cioè, di essere modellata con proprietà e relazioni, in modo da renderla efficace da un punto di vista di finalità ma anche utilizzabile da altre comunità. Il gruppo di lavoro SEI ha ipotizzato, così, tre possibili scenari di modellamento dell'Opus, cui poi se ne è aggiunto un quarto come variante dello scenario terzo: l'esercizio di modellamento è servito al gruppo per valutare tutti i pro e i contro di ciascuno scenario, partendo da alcuni elementi chiave intesi come "desiderata" e senza mai perdere di vista il principio dell'interoperabilità nella più ampia comunità del web. Negli schemi di rappresentazione dei 3 scenari di partenza, che sono quelli originali prodotti e discussi nell'ambito del gruppo di lavoro SEI, la formulazione della nuova entità riporta ancora la nomenclatura *SuperWork* in vece di *Opus*:

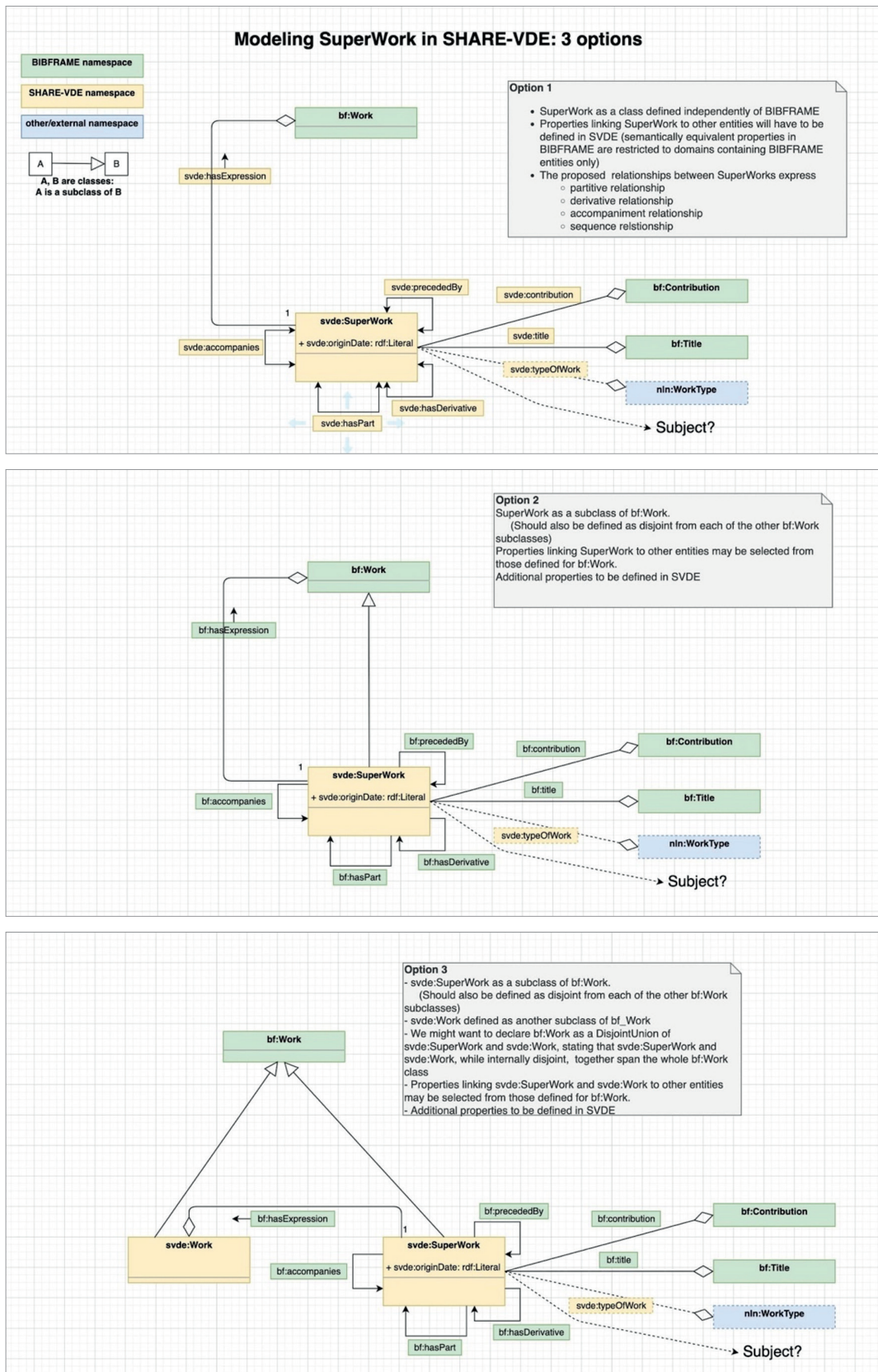


Figure 1-3. I tre scenari di modeling ipotizzati per l'Opus.



Un leggero rimodellamento dell'opzione 3, quella verso la quale il gruppo del SEI sempre più si orientava, è stato proposto e formalizzato in quella poi definita opzione 4, in cui, per altro, appare la formulazione definitiva di Opus:

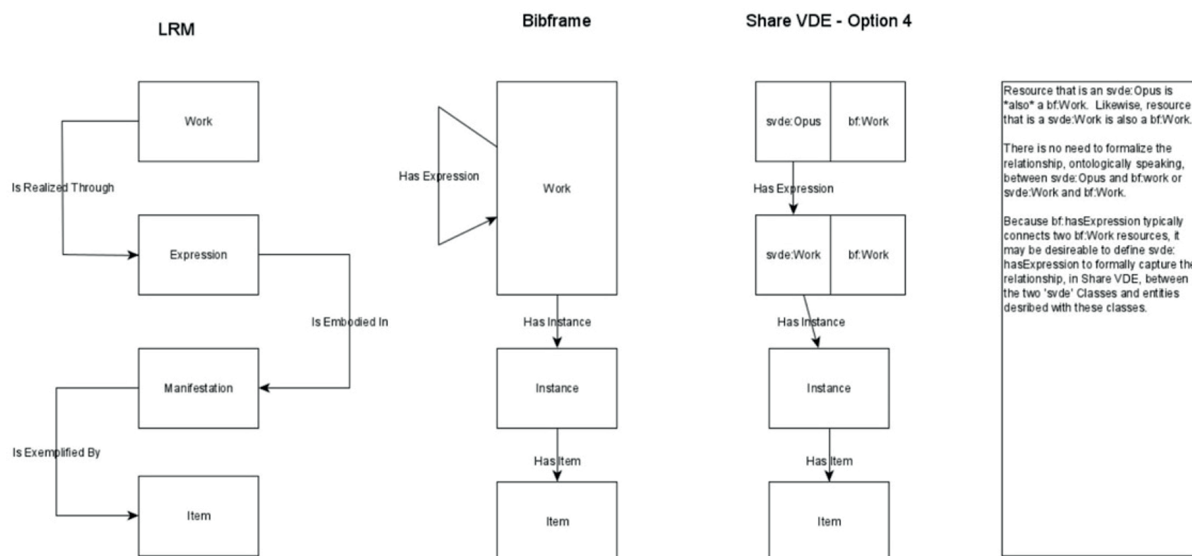


Figura 4. La quarta opzione di modeling dell'entità Opus.

La discussione del gruppo di lavoro si è così concentrata sull'analisi comparativa del modello 3 e della sua versione 4, con il preciso mandato dell'Advisory Council di studiare e formalizzare lo scenario che meglio si adattasse allo scopo di identificazione, descrizione, conversione e mantenimento dei dati delle biblioteche, tenendo conto anche delle sfide dell'interoperabilità con altri modelli. Per arrivare ad una decisione congiunta, dunque, il gruppo SEI ha formulato un elenco di "desiderata" per il modellamento dell'entità, con le rispettive formalizzazioni realizzabili nelle due opzioni selezionate come possibili, elaborando a conclusione dello studio comparativo una tabella di riepilogo per pesare i pro e i contro dei due scenari di modeling, che porterà alla decisione finale, quella di adottare lo scenario nella sua versione 4:

Opzioni del modello	Pro	Contro
Opzione 4	<ol style="list-style-type: none"> <li>Soddisfa la caratteristica del modello desiderata in modo flessibile</li> <li>Concede tempo affinché le migliori pratiche si sviluppino ulteriormente e forniscano miglioramenti agli algoritmi di istanziazione</li> <li>Nonostante le mappature dei cluster di entità e gli algoritmi di istanziazione siano diversi, l'approccio è simile a quello della Library of Congress rispetto al <a href="#">bf:Hub</a></li> </ol>	<ol style="list-style-type: none"> <li>Senza una definizione formale delle sottoclassi, la documentazione delle mappature dei cluster di entità e proprietà diviene imprescindibile</li> <li>(Forse non è un "contro") Con lo sviluppo delle buone pratiche e degli standard, l'applicazione delle caratteristiche del modello qui presentato può essere auspicabile a livello di ontologia BIBFRAME</li> </ol>
Opzione 3	<ol style="list-style-type: none"> <li>Soddisfa la caratteristica del modello desiderata (alcune perplessità sono annotate nella colonna "Contro")</li> <li>Fornisce una definizione formale di <a href="#">svde:Opus</a> e <a href="#">svde:Work</a> come sottoclassi</li> </ol>	<ol style="list-style-type: none"> <li>L'uso di sottoclassi pone l'interrogativo su come preservare l'equivalenza tra <a href="#">svde:Work</a> come sottoclasse di <a href="#">bf:Work</a> e <a href="#">bf:Work</a> in Share-VDE. Ciò rivela questioni di interoperabilità circa l'input e l'output verso/da fonti esterne.</li> <li>Diverse domande potrebbero beneficiare di ulteriori analisi e sviluppo di buone pratiche.</li> </ol>

Figura 5. Schematizzazione dei pro e dei contro degli scenari 3 e 4.

Le valutazioni sulla scelta dell'opzione 4 sono condivise con la comunità di Share e con tutti coloro che hanno contribuito alla riflessione su un così delicato intervento di modeling.

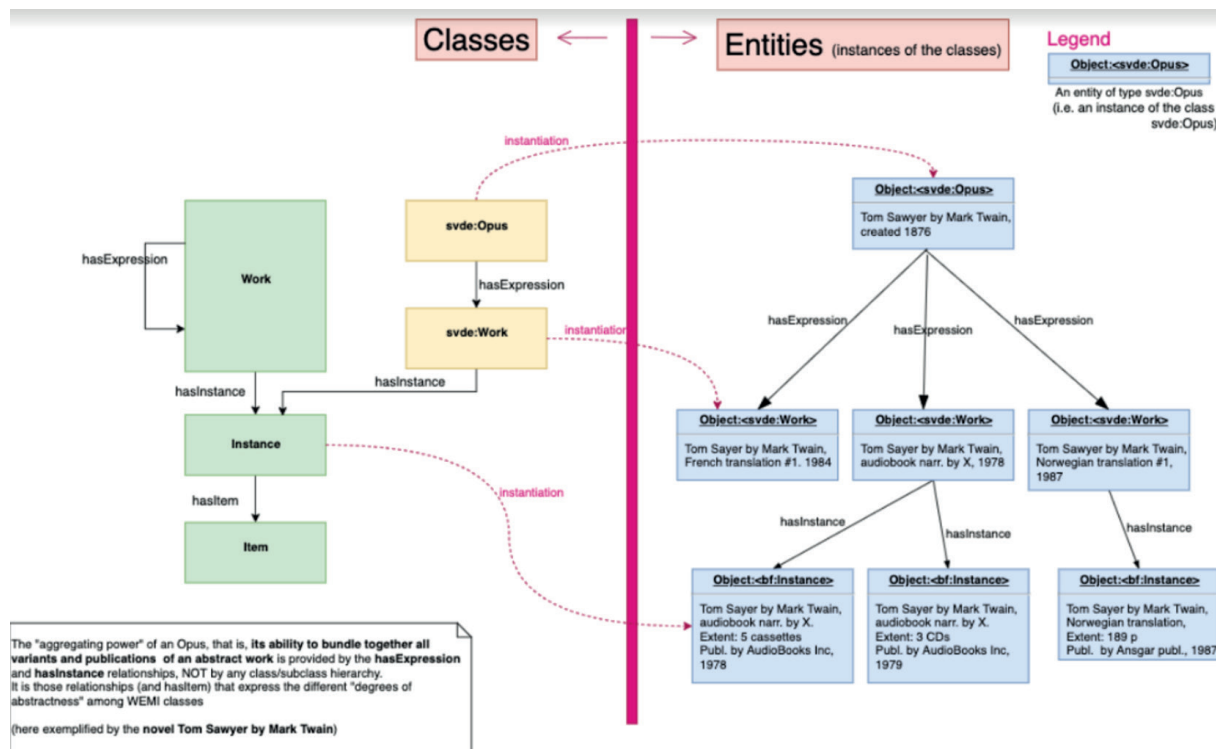


Figura 6. L'Opus e le sue relazioni qui esemplificate nell'istanziamento del romanzo Tom Sawyer di Mark Twain.

In questa soluzione di modellamento la decisione centrale, quella che rende davvero compatibile questa estensione ontologica con BIBFRAME e con IFLA LRM (e RDA) è la seguente: la risorsa di tipo `svde:Opus` è anche un `bf:Work` (quindi è *un tipo di* `bf:Work`) così come la risorsa `svde:Work` è anche un `bf:Work`. Le due entità modellate per Share-VDE diventano *tipi di* `Work` BIBFRAME, essendo così perfettamente compatibili sia con BIBFRAME che con le entità `Work` ed `Expression` di IFLA LRM.

## L'Entity modeling come terza generazione della disciplina catalografica

Quanto fin qui detto dell'Entity Resolution e dell'Entity modeling riguarda soprattutto elaborazioni massive di dati, quindi processi macchina attivati in progetti di conversione da un formato all'altro (per esempio da Marc a RDF) o nell'ambito del machine learning. Fino a pochissimo tempo fa. La logica *entity-oriented* sta maturando e coinvolgendo in modo sempre più evidente le pratiche manuali di catalogazione: come gli ILS e le piattaforme di gestione dei dati si posizioneranno rispetto all'avvicinamento sempre più prossimo della biblioteconomia ai linguaggi e alle tecniche del web, è un quesito aperto, che comincia a dare risposte pratiche sul fronte delle conversioni e sul fronte dei sistemi di discovery. Ma il cambiamento di orizzonte prospettato dal web semantico sta superando il contesto tradizionale delle conversioni e dei discovery e sta avviando ipotesi su nuovi scenari operativi, con definizione di casi d'uso focalizzati sulla creazione nativa di dati in rdf, l'analisi dell'impatto sui processi produttivi, lo sviluppo di moduli di catalogazione in linked data. Solo a titolo esemplificativo si citano:

- il progetto di sviluppo di un modulo di catalogazione in rdf, Libris (XL), voluto già nel 2018 dalla National Library of Sweden per gestire la nuova versione dello Union Catalogue nazionale (Wennerlund e Berggren 2019);
- l'iniziativa LD4P - Linked Data for Production, che nella Fase 2 si è concentrata sullo sviluppo di un ambiente di catalogazione basato su BIBFRAME e chiamato Sinopia (Schreur 2019);
- l'editor BIBFRAME della Library of Congress, un tool pensato per consentire di modellare le descrizioni bibliografiche secondo l'ontologia BIBFRAME.

Guardando questi editor e la loro focalizzazione sull'entità invece che su un record, si capisce a pieno il cambio di prospettiva e di orizzonte che anche nelle biblioteche sta maturando: quando un catalogatore comincia a ragionare su *come modellare un oggetto* (per esempio un libro, oppure una persona, oppure un evento), quali attributi e quali relazioni aggiungere per meglio rappresentare l'oggetto che sta descrivendo, per renderlo il più identificabile possibile anche al di fuori della biblioteca, allora ha già smesso di catalogare: sta *modellando l'entità*. Questo tema si riallaccia ed estende il tema del cammino della catalogazione verso la metadattazione, che è un processo concretamente già da molto tempo avvenuto (Gorman 2018, 121), e che ha ancora recentemente aperto un vivace dibattito teorico in Italia (Guerrini 2020) (Guerrini 2022). Pensando a cosa significhi l'entity modeling e a quanto incarni quell'auspicato passaggio dal record al real world object, mi piacerebbe provare a indicare un altro modo di guardare a questa evoluzione, come ad una sorta di *cambiamento di visione prospettica*, quasi come se il processo fosse quello di guardare la stessa cosa ma da un'altra visuale. Riepiloghiamo gli elementi chiave di ciascuno di questi "scenari",

quello della catalogazione, della metadattazione e dell'entity modeling, per provare a definire i confini di ciascuna di queste attività.

- a) *Catalogazione*: questo termine esprime bene la dimensione più tradizionale dell'attività bibliografica, quella di rappresentare in un catalogo il posseduto di una biblioteca o altre informazioni bibliografiche, attraverso l'applicazione di regole e standard che abbiano come motrice l'interesse dell'utente. La dimensione "spaziale" sufficiente per garantire questa dinamica di relazione tra l'attività della biblioteca e i suoi utenti è quella definita dalla biblioteca stessa e in qualche modo da essa delimitata: il catalogo rispecchia, per lo più, quanto posseduto da una biblioteca o quanto veicolato dalla biblioteca; l'utenza è quella della biblioteca stessa.
- b) *Metadattazione*: la definizione più semplice di questo termine è quella di "dati sui dati" o "informazioni sulle informazioni" (Riley 2004). Nei diversi contesti di utilizzo, il termine è stato usato per indicare le informazioni relative a qualsiasi *cosa*: libri, oggetti museali, finding aids per materiale archivistico, immagini: "*Broadly speaking, metadata encapsulate the information that describes any information-bearing entity*". Come Zeng e Qin ricordano nella loro fondamentale opera *Metadata*, dagli albori dei cataloghi e degli indici scritti a mano e stampati fino ai giorni nostri, quelli dei servizi web e delle app, la natura e l'obiettivo di descrivere le entità portatrici di informazioni sono rimaste più o meno invariate (Zeng e Qin 2016). Tuttavia, i metodi e le tecnologie sono cambiati in modo significativo. L'estensione dell'orizzonte di riferimento della biblioteca e degli istituti della cultura in generale, da una dimensione locale ad una dimensione sempre più ampia, l'arricchirsi dei materiali entrati a far parte delle collezioni o anche solo ad essi referenziati, l'allargamento dei servizi proposti dall'aumento esponenziale del digitale, con anche il moltiplicarsi delle relative mansioni di gestione, sono tutti fattori che hanno portato ad una vera esplosione di metadati, al punto che il termine "metadattazione" sta soppiantando lo stesso termine "catalogazione".
- c) *Entity modeling*: abbiamo parlato dell'entity modeling, di come realizzi quel cambiamento di mentalità prima che tecnologico che pone l'entità, o il real world object, al centro della propria attenzione, come oggetto da costruire e modellare. Il panorama di riferimento non è neanche più quello del web tradizionale ma quello del web semantico, il web dei dati, in cui gli oggetti sono costruiti seguendo il paradigma dei linked data e aggiungendo "semantica" alle informazioni, sì da renderle condivisibili con le macchine. La tecnologia, gli standard, i protocolli apparentemente si complicano, ed in certa misura questo davvero accade. Ma si tratta, prima e soprattutto, di un cambio di ottica, un nuovo passo verso un universo più esteso, quello globale del web, dove anche la terminologia deve superare il proprio limite di riferimento a singoli domini.

Se provassimo ad analizzare il risultato concreto delle tre operazioni elencate sopra, catalogazione, metadattazione, entity modeling, ci accorgeremmo che non c'è alcuna differenza nel risultato pratico: un record bibliografico, un set di metadati, un dataset di dati rdf. Metadati, di differenti tipologie e livelli di profondità, ma comunque metadati.

Ma se invece quei tre scenari provassimo ad analizzarli in modo diverso, cambiando la nostra ottica di visuale, ci accorgeremmo di quanto siano differenti e di quanto sia legittimo un cambio anche radicale di terminologia. Di seguito si rappresenta un oggetto (un appartamento) visto in tre dimensioni "spaziali" diverse:

- la dimensione locale, con una *planimetria* che non ha riferimenti al di fuori di sé e risponde all'esigenza di rappresentare l'oggetto senza alcun riferimento chiave al contesto esterno. I metadati associati alla planimetria sono quelli tipici per identificare uno stabile in una data località: un indirizzo con i suoi campi. Questa dimensione rappresenta la catalogazione, e la sua funzione di rappresentazione di una realtà più locale;
- la dimensione estesa al municipio, con una *mappa catastale* in cui quello stesso appartamento non è più rappresentato nel suo "isolamento", ma rispetto ad un catasto urbano ben più ampio. Lo stesso oggetto, lo stesso appartamento, viene identificato rispetto ad un orizzonte più ampio, e gli elementi descrittivi necessari a identificarlo sono diversi e soprattutto sono in relazione ad altri oggetti, e per altre finalità: sezione urbana, foglio, particella, subalterno. Questa dimensione rappresenta invece quella della metadattazione, e la sua funzione di rappresentare dati più ricchi e che abbracciano funzioni diverse: metadati descrittivi, metadati amministrativi (tecnici, di conservazione, sui diritti), metadati strutturali e linguaggi di marcatura;
- la dimensione geospaziale, con sempre lo stesso oggetto visto da una dimensione ancora più ampia: in questa dimensione i punti di riferimento per l'identificazione diventano quelli di latitudine e longitudine, dunque con riferimento al più vasto "globo". Questa dimensione è quella dell'entity modeling, il cui retroterra di realizzazione è quello ben più ampio del web (semantico).

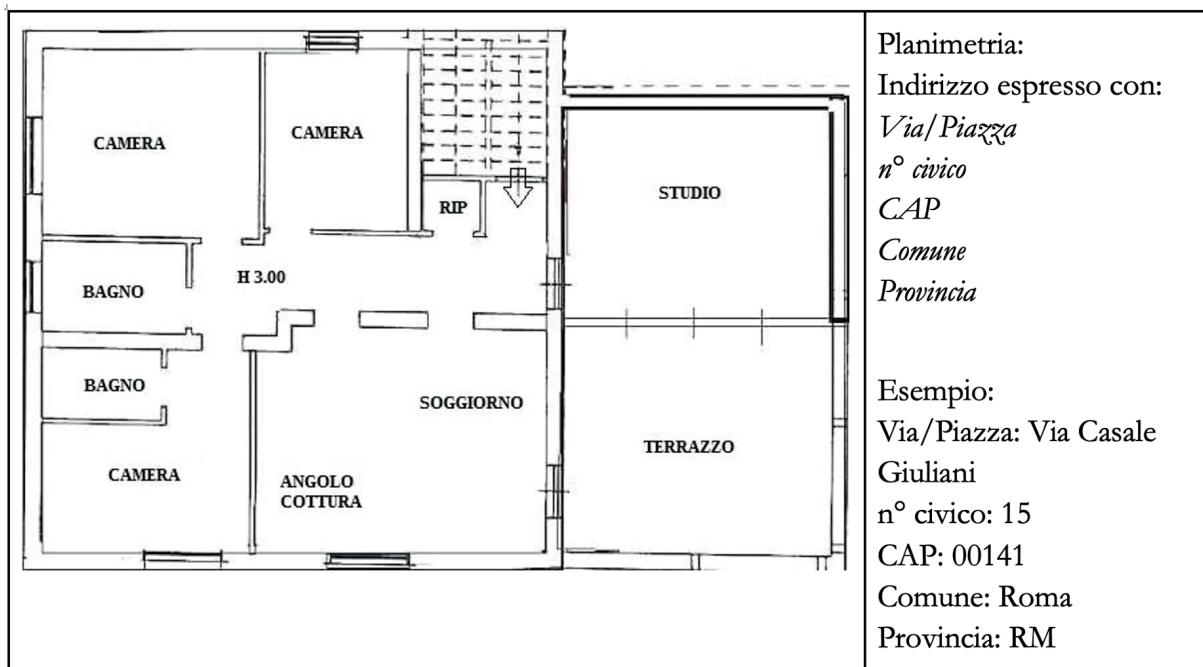




Figura 7-9. Rappresentazione di un medesimo oggetto da tre prospettive visuali e funzionali diverse.

Importante è sottolineare che il cambio di visuale non è fine a sé stesso, non è una scelta filosofica né tanto meno una moda, ma ha una sua necessità funzionale: lo stesso dato viene identificato, descritto, comunicato per finalità differenti. Questo potrebbe sembrare, per una biblioteca, un allontanamento dall'obiettivo principale o tradizionale, quello del proprio catalogo e della propria utenza. In realtà è l'espressione di un positivo e direi inevitabile allargamento dell'orizzonte: è come se le cose si guardassero da una visuale sempre più ampia, in un orizzonte sempre più esteso, partendo dalla dimensione locale, del proprio catalogo, passando per l'inclusione di nuovo materiale, di nuovi servizi, di nuovi tipi di utenza, di contaminazioni sempre più evidenti con mondi "affini" (cross-domain e interoperabilità) e arrivando, per ora, al web, in cui ogni "cosa" locale di-

venta tremendamente piccola e limitata se confrontata con la vastità che esso, il web, rappresenta. Ed è inevitabile che un cambio di orizzonte generi anche un cambio di linguaggio: se prima era sufficiente parlare di *documento*, pur con tutti i dubbi su cosa davvero sia un *documento*, ora bisogna adottare termini più globalmente riconoscibili, come quello di *risorsa*, che nella sua generalità e neutralità meglio esprime la totalità delle cose descrivibili nel web.

## Conclusioni

Le riflessioni proposte in questo studio partono da un concetto ormai quasi abusato quando si parla di linked data applicati all'ambito bibliografico, e cioè il passaggio dal record al real world object, per provare a definire meglio cosa si intenda per real world object e come questo termine sia entrato nel gergo bibliografico e sia utilizzato. L'indagine qui svolta è strettamente collegata ad un'analisi ancora in corso sul binomio tre *entità e identità*, binomio indagato sotto un profilo soprattutto filosofico, per arrivare, ovviamente, a individuare un criterio di applicazione di questa terminologia nell'ambito della disciplina catalografica: la distinzione tra real world entity e real world object, o più semplicemente tra entità e real world object, rimanda alla proposizione di un'entità complessa che si esprime attraverso diverse identità. L'informatica, soprattutto nella sua declinazione di programmazione orientata agli oggetti, ci aiuta a meglio focalizzare questa ipotesi di modellamento dell'universo bibliografico, che già da tempo si era appoggiato ai modelli entità-relazione ma che fatica a focalizzare e creare consenso sul concetto di real world object come nuovo protagonista della scena catalografica.

## Riferimenti bibliografici

- Coyle, Karen. 2015. «Coyle's InFormation: Real World Objects». *Coyle's InFormation* (blog). 16 gennaio 2015. <http://kcoyle.blogspot.com/2015/01/real-world-objects.html>.
- Cutter, Charles A. (Charles Ammi). 1876. *Rules for a Dictionary Catalogue*. U.S. Government Printing Office. <http://archive.org/details/cu31924029519026>.
- Gorman, Michael. 2018. *I nostri valori, rivisti: la biblioteconomia in un mondo in trasformazione*. A cura di Mauro Guerrini. Tradotto da Giuliano Genetasio. Firenze: Firenze University Press.
- Guerrini, Mauro. 2020. *Dalla catalogazione alla metadattazione: tracce di un percorso*. Collana Percorsi AIB 5. Roma: Associazione italiana biblioteche.
- , a c. di. 2022. «La metadattazione: cos'è?» *Biblioteche oggi* 40 (3): 21–50. <https://doi.org/10.3302/0392-8586-202203-021-1>.
- McCallum, Sally. 2022. «Collocation and Hubs. Fundamental and New Version». *JLIS.It* 13 (1): 45–52. <https://doi.org/10.4403/jlis.it-12760>.
- Riley, Jenn. 2004. «Understanding Metadata». NISO website. 2004. <https://www.niso.org/publications/understanding-metadata>.
- Schreur, Philip E. 2019. «Sinopia: A New Linked-Data Editing Environment Designed for Libraries». In *Metadata and Semantic Research*, a cura di Emmanouel Garoufallou, Francesca Fallucchi, e Ernesto William De Luca, 425–30. Communications in Computer and Information Science. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-36599-8\\_39](https://doi.org/10.1007/978-3-030-36599-8_39).
- Wennerlund, Bodil, e Anna Berggren. 2019. «Leaving Comfort Behind: A National Union Catalogue Transition to Linked Data», 10.
- Zeng, Marcia Lei, e Jian Qin. 2016. *Metadata*. 2nd edition. Chicago: Neal-Schuman.
- Zhu, Linhong, Majid Ghasemi-Gol, Pedro Szekely, Aram Galstyan, e Craig A. Knoblock. 2016. «Unsupervised Entity Resolution on Multi-Type Graphs». In *The Semantic Web – ISWC 2016*, a cura di Paul Groth, Elena Simperl, Alasdair Gray, Marta Sabou, Markus Krötzsch, Freddy Lecue, Fabian Flöck, e Yolanda Gil, 9981:649–67. Lecture Notes in Computer Science. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-46523-4\\_39](https://doi.org/10.1007/978-3-319-46523-4_39).



## A look at metadata processing beyond Libraries

Valdo Pasqui<sup>(a)</sup>

a) University of Florence, <https://orcid.org/0000-0002-4105-2208>

**Contact:** Valdo Pasqui, [valdo.pasqui@unifi.it](mailto:valdo.pasqui@unifi.it)

**Received:** 18 June 2022; **Accepted:** 30 June 2022; **First Published:** 15 September 2022

### ABSTRACT

Since many years libraries, archives and museums, the institutions entrusted with the dissemination and conservation of cultural heritage, contributed to metadata standards definition, meta-dating methodologies and metadata representation in different syntaxes, by participating to national projects and international initiatives.

Management systems and catalogs used in these contexts borrow from information and telecommunication technology tools, methodologies and techniques to generate, organize, share and use various types of metadata. But the Information and Communication Technology (ICT) area is not only a supplier of technological tools and solutions as it also constitutes a rich basin in which metadata plays a fundamental role in designing architectures, modeling information systems and implementing services.

Through an overview that includes national and European initiatives, especially in the digital transformation process of the public sector, this paper aims to offer a look to metadata beyond the traditional boundary of libraries and other cultural institutions and to underline some relevant aspects such as standardization, sharing, reuse and metadata quality assessment.

This framework highlights the need to carry on with the path of cooperation between different functional domains and organizational contexts in order to consolidate and extend the (re)use of metadata schemes, ontologies and controlled vocabularies both in the redesign of digitized processes and in the implementation of services supporting them. By undertaking since design early stages a multidisciplinary approach based on metadata standards can ensure greater flexibility and higher interoperability. This vision requires the enhancement of intersectorial skills that meld metadata methodologies and syntaxes representation basic knowledge with the ability to model functional domains using metadata schemes and ontologies.

### KEYWORDS

Metadata; Ontologies; Controlled vocabularies; Open Data; Web Services; API; Interoperability.

## Uno sguardo alla metadattazione oltre le Biblioteche

### ABSTRACT

I metadattati, le metodologie di metadattazione e le sintassi per rappresentare i metadattati sono ambiti nei quali le biblioteche, gli archivi e i musei, istituzioni alle quali è affidata la diffusione e la conservazione del patrimonio culturale, da molti anni hanno contribuito alla definizione di standard e allo sviluppo di metodologie grazie anche a progetti nazionali e iniziative internazionali. I sistemi gestionali e i cataloghi di cui si avvalgono questi contesti mutuano dalle tecnologie informatiche e delle telecomunicazioni tecniche e strumenti che consentono di generare, organizzare, condividere e usare le varie tipologie di metadattati. Ma l'area Information and Communication Technology (ICT) non è solo un fornitore di strumenti e soluzioni tecnologiche poiché costituisce anche un ricco bacino nel quale i metadattati svolgono un ruolo fondamentale per il disegno delle architetture, per la modellazione dei sistemi informativi e per l'implementazione dei servizi.

Attraverso una panoramica che comprende iniziative nazionali e europee, soprattutto nell'ambito della trasformazione digitale del settore pubblico, questo contributo intende offrire uno sguardo oltre al confine dei metadattati del mondo bibliotecario e delle altre istituzioni culturali e mettere in evidenza alcuni aspetti rilevanti come la standardizzazione, la condivisione, il riuso e la qualità dei metadattati.

Dal quadro esaminato emerge la necessità di proseguire nel cammino intrapreso per promuovere la cooperazione tra i diversi domini funzionali e i vari contesti organizzativi al fine di consolidare e ampliare l'uso degli schemi di metadattati, delle ontologie e dei vocabolari controllati nella progettazione dei processi digitalizzati e dei servizi che li supportano. Un approccio multidisciplinare che si fonda sull'uso di schemi di metadattati può assicurare fin dalle prime fasi della progettazione maggiore flessibilità e una più elevata interoperabilità. Questa visione richiede lo sviluppo di competenze intersectoriali che uniscano alla conoscenza delle metodologie di metadattazione e delle sintassi di rappresentazione dei metadattati anche la capacità di modellare i domini funzionali avvalendosi degli schemi di metadattati e delle ontologie.

### PAROLE CHIAVE

Metadata; Ontologie; Vocabolari controllati; Open Data; Web Services; API; Interoperabilità.

## 1. Spigolature in retrospettiva sulla metadattazione

Nell'affascinante libro “Payprus. L’infinito in un giunco” (Vallejo 2021) la studiosa spagnola Irene Vallejo ricorda che le tavolette di argilla usate dai Sumeri cinquemila anni fa sono state la prima forma di libro. A Nippur, nel sud della Mesopotamia, gli scavi archeologici hanno portato alla luce tavolette usate come cataloghi nelle quali ogni opera veniva identificata attraverso la prima riga del testo o un breve riassunto del contenuto, talvolta erano riportati anche il nome dell’autore, altri dati e per evitare la dispersione dei testi lunghi veniva indicato il numero delle tavolette che componevano l’opera. Nella biblioteca di Khattusa, capitale del regno ittita situata nell’odierna Turchia, sono state rinvenute tavolette contenenti i cataloghi di alcune collezioni. Si tratta di alcuni degli esempi più antichi di metadattazione a testimonianza di come, già con le prime tipologie di supporti fisici assimilabili ai libri, gli esseri umani abbiano avvertito la necessità di descrivere i libri e organizzare tali descrizioni in cataloghi. L’autrice afferma che “Un catalogo non è una semplice appendice della biblioteca: è il concetto su cui si basa, ne è il nesso e il culmine”. Questa essenza e centralità del catalogo si afferma e consolida proprio grazie ai metadati e alla metadattazione per mezzo dei quali prende forma, evolve nel tempo e svolge la propria funzione.

L’evoluzione degli strumenti e delle soluzioni offerte dall’ICT ha consentito non solo di strutturare e gestire in modo sempre più articolato e ricco i metadati delle biblioteche, per esempio passando dall’approccio basato sui record dei formati MARC a quello orientato alle relazioni basato su linked data, ma ha anche favorito l’uso dei metadati e lo sviluppo dei modelli di metadattazione in molti altri bacini di applicazione. La diffusione delle reti di telecomunicazione e del Web, la disponibilità dei dispositivi di memorizzazione (storage) che consentono l’acquisizione e l’archiviazione di grandi volumi di dati (database, testi, video, immagini, big data), l’elevata potenza di calcolo e le tecnologie del Cloud hanno consentito la digitalizzazione di processi sempre più complessi, distribuiti in rete e tra loro interoperanti, una trasformazione nella quale i metadati svolgono un ruolo fondamentale e che è avvenuta in tempi molto rapidi e in un arco temporale molto limitato se paragonati alla storia dei libri e delle biblioteche. In questo recente percorso si iscrive anche una curiosa storia che si colloca a cavallo tra tecnologia e linguistica. Nel 1955, quando negli Stati Uniti i primi calcolatori venivano chiamati Electronic Data Processing System (EDPS), denominazione poi sostituita da “computer”, ormai entrato a far parte del lessico comune anche nella lingua italiana, François Girard, responsabile del settore pubblicitario della IBM France, scrisse una lettera al suo ex insegnante Jacques Perret, professore di filologia latina presso la Sorbonne, per chiedergli suggerimenti sul nome francese da attribuire al nuovo modello di calcolatore che era in fase di realizzazione presso la fabbrica Corbeil-Essonnes (IBM 2014). La risposta dell’accademico esamina varie ipotesi e propone la parola “ordinateur”, anzi la declinazione al femminile “ordinatrice” aggettivata con “électronique” sia perché allora IBM aveva adottato alcuni nomi femminili come selezionatrice e tabulatrice, sia perché così sarebbe stato più chiara la differenziazione dal significato ecclesiastico dei termini “ordinare” e “ordinazione”. Perret motiva la proposta citando anche il Littré, Dizionario della lingua francese, in cui il termine “ordinateur” viene riferito a Dio che mette ordine nel mondo. La scelta dell’IBM sarà di adottare il termine “ordinateur” che da allora è il vocabolo correntemente usato in francese per denominare i computer.

Di fronte alla complessità delle informazioni e alla quantità dei dati questa aspirazione umana di “mettere ordine” si attua proprio attraverso gli schemi di metadati, la metadattazione e i cataloghi. I metadati sono stati usati fin dall’inizio anche nell’informatica sebbene assumendo nomi diversi come attributi, descrittori, proprietà. Il Multics, sviluppato nella seconda metà degli anni Sessanta in collaborazione tra i laboratori Bell, la General Electric e il MIT, è stato uno dei primi sistemi operativi che ha ispirato molti sistemi successivi tra cui UNIX. Una delle innovazioni di quel progetto fu l’ideazione del file system come il componente del sistema operativo cui è affidata la gestione dei dati e dei dispositivi di registrazione. Secondo una definizione classica (Wiederhold 1977) un file è un insieme di record simili composti di campi che contengono dati elementari. Nell’ambito bibliotecario ancora oggi i formati MARC si rifanno a questo approccio poiché sono stati concepiti proprio agli inizi dello sviluppo dell’automazione. Nei sistemi operativi dei nostri computer e tablet e nei servizi di condivisione come Google Drive, PCloud o Dropbox siamo abituati a chiamare file anche un documento PDF o Word e un’immagine JPEG, ovvero una qualunque sequenza di bit strutturata in base a un determinato formato. Le caratteristiche di questi file sono descritte da proprietà come il nome, la dimensione, il formato e i diritti di accesso/modifica, a loro volta questi metadati sono organizzati in cataloghi, chiamati “files directories” o “volume table of contents” (VTOC), utilizzati dal file system del sistema operativo per l’identificazione, la localizzazione e l’accesso ai dati. L’utente del sistema ne ha visibilità anche in forma leggibile come in Windows attraverso Esplora file o, a più basso livello, il classico comando “dir” del DOS (vedi Figura 1).

```
Directory di D:\DOCS\Articoli\Metadattazione_maggio_2022
14/04/2022  10:18  <DIR>          S301L\Valdo  .
14/04/2022  10:18  <DIR>          S301L\Valdo  ..
31/03/2022  10:51          301.710      S301L\Valdo  16 avril 1955 _ _Que diriez-vous d'ordinateur __.pdf
10/04/2022  19:00          66.433      S301L\Valdo  Appunti.docx
07/04/2022  18:23       12.745.981  S301L\Valdo  Bagley.pdf
31/03/2022  12:25       1.281.186  S301L\Valdo  Core_Vocabularies-Business_Location_Person_v1.00_Specif
11/04/2022  18:37          18.731      S301L\Valdo  Fwd I Seminario di JLIS.it del 19 maggio 2022 - Valdo
21/03/2022  22:26       113.388     S301L\Valdo  Guerrini_BibliotecheOggi_Metadattazione_2022.docx
12/03/2022  20:09       123.972     S301L\Valdo  Progetto DAFNE (progetto MIUR) Progetti e realizzazioni
11/04/2022  18:35       536.611     S301L\Valdo  Submissions JLIS.it.pdf
10/04/2022  17:33       552.706     S301L\Valdo  T-REC-X.500-200811-S!!PDF-E.pdf
08/04/2022  18:30          2.861       S301L\Valdo  testo_email.txt
14/04/2022  10:18         43.954      S301L\Valdo  Testo.docx
24/03/2022  19:26       1.647.254  S301L\Valdo  Understanding metadata NISO.pdf
          12 File      17.434.787 byte
          2 Directory 184.197.332.992 byte disponibili
```

Figura 1 Metadati che descrivono una directory DOS mostrati dal comando “dir”.

I file testo, immagini, audio e video (es. TIFF, JPEG, MP3, WAV, DOCX, PDF) riservano una porzione ai metadati (descrittivi, tecnici e di struttura) grazie ai quali i sistemi operativi li possono riconoscere e gestire. La specifica del formato PDF/A (PDF/A 2005), il cui uso è contemplato dalle norme nazionali in materia di conservazione degli archivi e dei documenti, fa sì che i documenti generati in conformità a tale formato contengano tutte le informazioni relative a testo, immagini raster e vettoriali, font di caratteri, colori, etc. necessarie per essere interpretati e visualizzati senza dipendere da dati residenti in sorgenti esterne.

Il termine metadata, “dati sui dati” (data about data), è originato in ambito ICT e la sua paternità è attribuita a Philip Bagley che nel 1968 propose alcuni concetti innovativi per la progettazione dei linguaggi di programmazione (Bagley 1968) quando i linguaggi più evoluti (es. Algol, Cobol, Fortran, PL/I) avevano una limitata capacità di gestire strutture di dati complesse. Nel seguente passaggio:

As important as being able to combine data elements to make composite data elements is the ability to associate explicitly with a data element a second data element which represents data “about” the first data element. This second data element we might term a “metadata element”. Examples of such metadata elements are: an identifier, a domain “prescriptor” which specifies from what domain the values of the first element must be taken, an access code which limits the which the first data element can be accessed

sono espressi il concetto di “metadata element”, ormai entrato nel linguaggio corrente dei principali schemi di metadati, la possibilità di restringere il dominio di valori che il dato descritto può assumere, l’identificatore, gli elementi compositi per rappresentare dati strutturati, le relazioni tra elementi, i domini di valori e i modelli utilizzati per raggruppare i metadati comuni ad una classe di elementi compositi. Si tratta di principi poi ripresi e sviluppati nella modellazione a oggetti e nei formalismi di definizione dei metadati e degli schemi di metadati come XML Schema.

Infine in questo breve excursus non si può rinunciare a citare la famiglia degli standard X.500 (ISO 9554), sviluppati<sup>1</sup> negli anni Novanta per assicurare l’interoperabilità dei directory services, servizi concepiti per gestire dati relativi a persone, organizzazioni (es. elenchi telefonici “pagine bianche”, “pagine gialle”, contatti aziendali, indirizzari etc.), risorse informatiche (computer, stampanti, dispositivi) e servizi applicativi. Attualmente l’autenticazione federata, il controllo degli accessi a servizi e dispositivi, l’infrastruttura a chiave pubblica (PKI) per i certificati digitali usati ai fini della trasmissione sicura (es. HTTPS) e la firma digitale si avvalgono di sistemi che implementano il Lightweight Directory Access Protocol (LDAP) basato sugli standard X.500. Il modello definito da X.500 prevede un insieme distribuito di Directory Information Tree (DIT) organizzati secondo una struttura gerarchica ad albero costituita da oggetti, chiamati entry, ciascuno dei quali è un insieme di attributi che hanno un tipo e assumono uno o più valori. Il directory schema definisce le classi di oggetti e per ciascuna gli attributi e se questi sono obbligatori o opzionali. Ogni oggetto (entry) può essere associato ad una o più classi di oggetti e ha un nome univoco detto distinguished name (DN), costituito dalla concatenazione di alcuni attributi dalla radice dell’albero fino alla entry. L’insieme delle definizioni dei tipi degli attributi, delle classi di oggetti e di altre informazioni che permettono di operare nel database che implementa la struttura gerarchica formano lo schema. Il documento (RFC2252 1997) del 1997 e i suoi successivi aggiornamenti formalizzano i tipi degli attributi, le sintassi e le classi di oggetti, a ciascuno di questi elementi è associato un identificatore, object identifier (OID) univoco e un nome descrittivo. Gli standard X.520 e X.521 definiscono insiemi di attributi e

---

<sup>1</sup> Lo standard è stato elaborato dal Consultative Committee for International Telephony and Telegraphy (CCITT), poi diventato International Telecommunication Union (ITU-T), in collaborazione con The International Organization for Standardization (ISO).

classi di oggetti che permettono di rappresentare le persone e le organizzazioni. I documenti (RFC 2256 1997), (RFC 2798 2000) e (RFC 4519 2006) definiscono una serie di attributi, alcuni dei quali sono riportati nella Tabella 1, e un insieme di classi di oggetti come “organization”, “organizationalPerson”, “person”, “device” che aggregano gli attributi base e permettono di descrivere oggetti appartenenti a determinate categorie.

Attributo	Descrizione
c	Country code / Codice della nazione
cn	Person's full name /Nome e Cognome
sn	Surname (family name) /Gognome
givenName	Part of a person's name which is not their surname nor middle name /Nome
l	LocalityName (e.g. city) /Città
st	stateOrProvinceName /Stato o Provincia
street	Street / Indirizzo
o	organizationName /Denominazione dell' organizzazione
title	Title of a person in their organizational context / Ruolo della persona nel contesto organizzativo
businessCategory	the kind of business performed by an organization / Attività dell'organizzazione
postalAddress	Indirizzo postale
postalCode	Codice postale
telephoneNumber	Numero telefonico
mail	Indirizzo di posta elettronica
mobile	Numero telefonico mobile
userPassword	Password
userCertificate	Certificato digitale dell'utente
cACertificate	Certificato digitale della Certification Authority

Tabella 1. Alcuni attributi usati negli schemi LDAP.

Nell'ambito delle istituzioni accademiche e di ricerca sono stati consolidati gli schemi eduPerson (eduPerson 2016), eduOrg e Schema for Academia (SCHAC) (SCHAC 2022) che consentono di descrivere persone ed enti che operano nei contesti della formazione superiore<sup>2</sup>.

I concetti di identificatore univoco, per identificare gli attributi e gli oggetti, di classi astratte di oggetti, per descrivere oggetti più complessi a partire da elementi descrittivi elementari (attributi), di schema, per organizzare gli attributi/metadati attraverso meccanismi di inclusione e riuso di insiemi già definiti, sono analoghi ai principi fondanti delle correnti modalità di metadattazione basate su XML e su XML Schema Definition.

---

<sup>2</sup> Attualmente la manutenzione e lo sviluppo di questi schemi e dei servizi di cooperazione interistituzionale correlati è svolto dal Research and Education FEDerations group (REFEDS) (REFEDS n.d.) che a sua volta opera nell'ambito di GÉANT (GÉANTn.d.) costituito dalla European National Research and Education Networks (NRENs).

## 2. Metadati e Metadazione nell'ecosistema digitale

Nel paragrafo precedente abbiamo visto come la metadazione abbia accompagnato fin dall'inizio lo sviluppo dell'ICT risultando nella definizione di insiemi (schemi) di metadati per descrivere oggetti del mondo reale e oggetti nativi del mondo digitale come i file, i data set, i dispositivi, i servizi e le applicazioni. I metadati sono ormai pervasivi nello sviluppo dei sistemi informativi e nella digitalizzazione dei processi sia nell'ambito detto business to business (B2B) sia nel settore pubblico attraverso i piani per la trasformazione digitale della Pubblica Amministrazione.

Le due organizzazioni W3C e OASIS (OASIS n.d.) hanno definito standard, metodologie e strumenti per la produzione, la gestione e la condivisione dei metadati. Risale ai primi anni Duemila la diffusione di XML, la definizione di XML schema per descrivere i documenti XML e la successiva specifica XML Schema Definition (XSD) (XSD 2012) che hanno consentito la formalizzazione, l'importazione e il riuso di molteplici schemi di metadati. Il termine "namespace", che in XML esprime il principio della modularità mediante la condivisione di elementi e attributi ("markup vocabulary") definiti in moduli software distinti, è ormai diventato di uso corrente per indicare l'importazione e il riutilizzo di insiemi di metadati. Parallelamente, l'attività di sviluppo e di standardizzazione del Web Semantico, ispirata dal modello Linked Data ideato da Tim Berners-Lee (Berners-Lee 2006), ha portato alla definizione nel 2014 del Resource Description Framework (RDF) che rappresenta l'informazione del Web come un grafo connesso (RDF 2014) costituito da triple soggetto, oggetto, predicato. Gli RDF Schema (RDFS) permettono di formalizzare classi, tipi di dati, domini delle proprietà, relazioni gerarchiche tra le classi e le proprietà, detti anche RDF Vocabulary, e sono diventati una delle modalità più diffuse per definire schemi di metadati. Recentemente si è affermata la notazione JSON-LD (JavaScript Object Notation for Linked Data), una "raccomandazione" nell'ambito del W3C che specifica come codificare (serializzare) i linked data secondo la sintassi JSON (JavaScript Object Notation) molto diffusa nell'ambito degli sviluppatori delle interfacce programmatiche (API). Questi modelli e le relative specifiche hanno agevolato l'utilizzo dei metadati e la definizione di schemi metadazione in molti settori dell'ICT che per comodità di esposizione e in base alle finalità di applicazione possono essere schematizzati in tre ambiti:

- a) assicurare l'interoperabilità dei dati tra domini funzionali diversi
- b) facilitare la produzione, diffusione e il riuso dei data aperti (Open Data)
- c) agevolare l'interoperabilità applicativa tra i servizi di rete.

Per ciascuna delle tre aree individuate esamineremo alcuni aspetti relativi alla definizione e all'uso dei metadati con un'attenzione prioritaria per le iniziative rivolte al settore pubblico in ambito europeo e italiano. Giova ricordare l'esistenza di altri ambiti come quello dei siti Web, dei social network, dei servizi multimediali per la diffusione di video e musica e della gestione dei diritti d'autore nei quali sono stati definiti schemi di metadati (es. FOAF, Schema.org, ODRL, PRISM) che per ragioni di spazio non è possibile trattare in questo contesto.

## 3. Metadati e interoperabilità dei dati

La necessità di assicurare l'interoperabilità dei dati appartenenti ad ambiti organizzativi e a domini funzionali eterogenei è un'esigenza ormai prioritaria per le pubbliche amministrazioni (e-government), nel settore commerciale e in quello della ricerca, oltre che per la conservazione e valorizzazione

zione del patrimonio culturale. Le ontologie, i vocabolari controllati, il riuso di schemi di metadati sono punti di riferimento imprescindibili per ogni progetto che intenda soddisfare tali requisiti.

### 3.1 Il contesto europeo

La “Strategia europea per i dati” sottolinea la centralità dei dati per il futuro dell’economia europea e esprime la volontà di creare uno “spazio unico dei dati”, cioè un mercato unico che favorisca lo sviluppo di nuovi prodotti e servizi basati sui dati (personali, non personali e dati rilevanti in ambito commerciale e industriale) accessibili in modo sicuro (EUDati 2020). Una delle prime iniziative europee con queste finalità è stata la Direttiva europea “Infrastructure for spatial information in Europe (INSPIRE)” volta a creare un’infrastruttura europea di supporto all’attuazione delle politiche ambientali e di salvaguardia del territorio (INSPIRE, 2007), attraverso la condivisione e l’accesso pubblico ai dati territoriali e ai dati inerenti le attività che impattano sull’ambiente. Tra le regole di implementazione comuni la Direttiva individua i metadati poiché essi assicurano l’interoperabilità tra i dati spaziali/territoriali e i servizi che li rendono disponibili (INSPIRE\_TG 2022).

Un altro obiettivo delle strategie messe in atto a livello europeo e rafforzate dalla Dichiarazione di Tallin del 2017 (TALLIN 2017) è realizzare il principio once-only mediante lo sviluppo di servizi pubblici digitali transnazionali (cross-border) grazie a i quali i cittadini non siano più costretti a fornire alle amministrazioni documenti e dati già in possesso di altre amministrazioni o autorità pubbliche. A tale finalità sono riconducibili due iniziative:

- la piattaforma collaborativa Joinup, realizzata per favorire la condivisione e lo sviluppo di servizi interoperabili (Joinup\_CV 2022);
- il Single Digital Gateway (SDG), attivo dal Dicembre 2020, un portale concepito per aiutare i cittadini e le imprese a trovare informazioni sulle norme, i diritti e i procedimenti sia europei sia nazionali con l’obiettivo di attivare entro il 2023 un insieme di 21 procedure amministrative e di rendere pienamente accessibili a livello transnazionale i servizi online nazionali (SDG n.d.).

Questi progetti sono volti a promuovere la cultura del riuso dei dati nelle amministrazioni, prevedono la creazione di registri autoritativi (basic registries) che assicurino l’autenticità dei dati, la loro reperibilità, qualità e accessibilità, e la condivisione di schemi di metadati, vocabolari controllati e ontologie che permettono il funzionamento dell’intera infrastruttura di servizi. Nell’ambito Joinup si segnala il lavoro che ha condotto alla definizione degli e-Government Core Vocabularies (e-GOV n.d.) che comprendono:

- Core Business Vocabulary - Rappresenta le caratteristiche delle entità legali (“trading bodies” formalmente registrate da qualche autorità e presenti in qualche registro) come il nome legale, l’attività, l’indirizzo, la tipologia etc.
- The Core Location Vocabulary - Coglie le caratteristiche dei luoghi come l’indirizzo, l’identificatore geografico (un URI definito da GeoNames.org oppure un DBpedia URI), il nome geografico (secondo le specifiche INSPIRE) etc.
- The Core Person Vocabulary - Esprime le caratteristiche fondamentali di una persona intesa come individuo reale (nome, sesso, data di nascita, etc.)

- The Core Public Organization Vocabulary (CPOV) - Descrive le organizzazioni pubbliche che operano nell'Unione Europea (organizzazione, le unità organizzative, le unità amministrative territoriali, i contatti, le ore di apertura etc.)
- Core Public Service Vocabulary Application Profile (CPSV-AP) - Descrive le caratteristiche di servizi offerti dalle pubbliche amministrazioni (titolo, descrizione, input, output, fornitori, luoghi, etc.)
- The Core Evidence and Criterion Vocabulary - Descrive i principi e le modalità che un'entità privata deve soddisfare per essere qualificata a realizzare servizi pubblici o a partecipare alle forniture pubbliche. Per "criterio" si intende una regola usata per valutare, giudicare o testare e per "evidenza" il modo attraverso il quale un criterio può essere comprovato.

Questi vocabolari, espressi in sintassi RDF, sono modelli riusabili e estensibili che catturano e descrivono le caratteristiche di entità tipiche del settore pubblico e che possono essere usati per:

- avviare la progettazione dei modelli logici e concettuali di nuovi sistemi informativi
- modellare dati specifici usati per l'interscambio tra sistemi informativi esistenti
- integrare dati provenienti da differenti sorgenti
- pubblicare dati aperti attraverso cataloghi, registri e portali pubblici (vedi il paragrafo seguente)

al fine di assicurare il riuso e la condivisione dei concetti base, l'interoperabilità semantica dei modelli concettuali e l'utilizzo di sintassi standard per esprimerle.

### 3.2 Il contesto italiano

La centralità dei dati è uno degli elementi cardine del Piano Triennale per l'informatica nella Pubblica Amministrazione fin dalla sua prima versione (2017 – 2019) che ha impostato una strategia basata sull'individuazione delle basi dati di interesse nazionale e sulla condivisione di modelli dati (ontologie) e vocabolari controllati. La versione 2022-2023 del Piano, nell'ambito dell'Obiettivo 2.2 volto a migliorare la qualità dei dati e dei metadati, ha introdotto una nuova linea d'azione che fissa per Giugno 2022 il rilascio del "National Data Catalog per l'interoperabilità semantica" a cura del Dipartimento per la Trasformazione Digitale (PTIPA 2021).

Il Portale dei dati aperti della pubblica amministrazione (cfr. il prossimo paragrafo), il Sistema dei Registri INSPIRE Italia (INSPIRE\_REG\_IT, n.d.), l'insieme delle ontologie OntoPiA e i vocabolari relativi ad alcuni settori pubblici costituiscono un insieme di strumenti concepiti allo scopo di normalizzare i dati in possesso della PA, offrire alle imprese e ai privati punti di riferimento ufficiali per il popolamento delle loro basi di dati e agevolare lo scambio di dati tra le PA.

La Figura 2 è una rappresentazione funzionale della rete di ontologie OntoPiA nella quale si rintracciano quelle relative alle Persone, Organizzazioni e Luoghi, la Cultural Heritage Ontology che sistematizza le ontologie sviluppate nel biennio 2017-2018 dal progetto ArCo del MiC (ArCo 2017), oltre alle classificazioni per il settore cultura, le tipologie di luoghi di interesse culturale e alcune classificazioni per l'ambito universitario (ruoli accademici, settori scientifico-disciplinari).



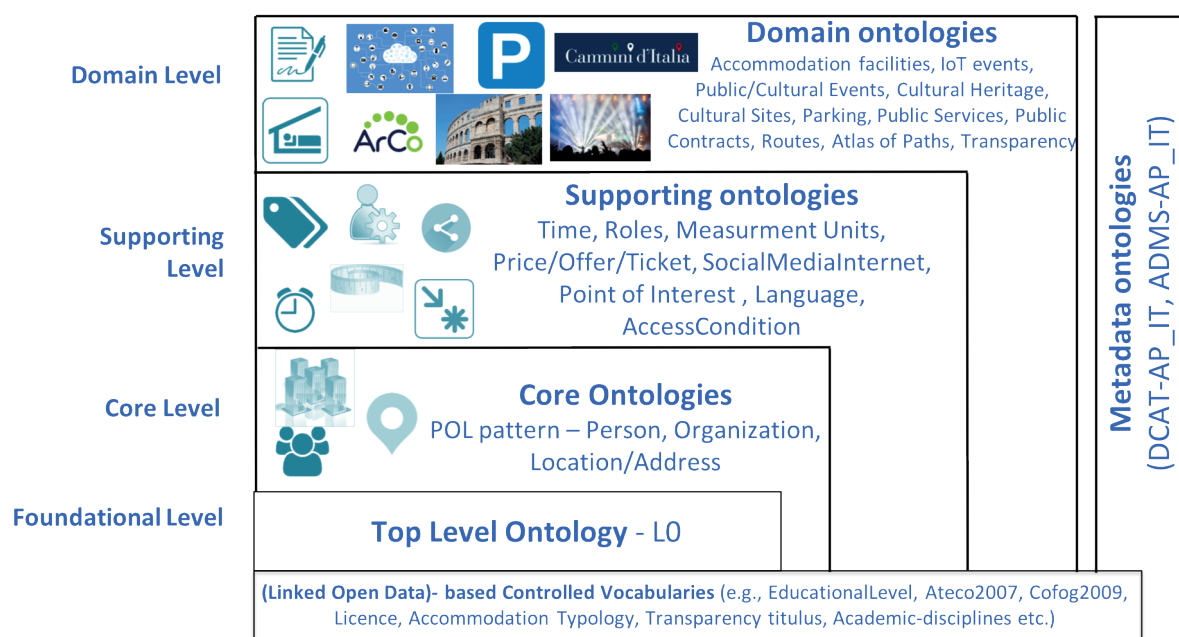


Figura 2 Rete OntoPiA tratta da <https://github.com/italia/daf-ontologie-vocabolari-controllati>

il PNRR comprende anche una specifica linea d'azione ("misura") denominata "1.3.2 Sportello Digitale Unico" volta alla implementazione della componente italiana del Single Digital Gateway, alla quale sono interessati 230 enti (PA Centrali e Locali, Università e Istituti di Ricerca) di cui 55 coinvolti nel progetto che prevede di realizzare 10 procedure online entro dicembre 2022 (target nazionale) e giungere a 21 entro dicembre 2023 (target finale europeo). L'AgID è il soggetto attuatore (AgID, 2022) di questo progetto, declinato come processo di reingegnerizzazione e introduzione di nuove tecnologie digitali. Le Università hanno attivato un tavolo tecnico al fine di impostare le soluzioni architetture e funzionali più idonee per l'attivazione dei servizi relativi all'orientamento, all'iscrizione ai test ai corsi a numero programmato e alle immatricolazioni che tengano conto delle complesse e differenti articolazioni degli Atenei. I metadati e i vocabolari controllati svolgono un ruolo abilitante per la realizzazione dei servizi SDG ai fini della normalizzazione e riconciliazione di dati e procedure eterogenei.

### 3.3 Esempi di interoperabilità tramite metadati

Avvalersi di schemi di metadati consolidati agevola l'integrazione dei dati tra domini diversi anche attraverso la mappatura di dati eterogenei verso schemi comuni. Un esempio assai diffuso di questa pratica sono i metadati Dublin Core spesso usati per trasformare dati bibliografici provenienti da sorgenti eterogenee per esempio nei processi di raccolta tramite protocolli di harvesting come il ben noto OAI-PMH. In ambito informatico è molto diffusa la pratica della mappatura per riconciliare tra loro i dati provenienti da basi di dati differenti per poter effettuare elaborazioni complesse, trasversali e che richiedono l'integrazione e la correlazione di dati appartenenti a diversi domini applicativi.

Nei sistemi di Business Intelligence (BI) e Data Warehouse (DW) e nell'analisi di grandi volumi di dati (big data) esistono processi, metodologie e linguaggi, denominati Extract-Transform-Load (ETL) che permettono di interpretare, elaborare, trasformare, riconciliare e normalizzare dati provenienti da sorgenti diverse. I metadati che descrivono i dati e le procedure (raccolta, estrazione, pulizia, trasformazione, mappatura dei tipi e caricamento dei dati, struttura delle aree intermedie usate per la trasformazione, modellazione logica fisica del database del DW, uso degli strumenti di analisi e reportistica) guidano questi articolati e complessi processi al punto che si usa anche l'espressione metadata driven ETL.

Un altro esempio di metadati concepiti ai fini dell'interoperabilità semantica è la recente diffusione nell'ambito dell'istruzione superiore di pacchetti informativi digitali contenenti metadati che certificano le abilità, le competenze e i titoli conseguiti da una persona (es. certificazioni linguistiche e informatiche/digitali, crediti e titoli accademici), il soggetto che li ha conseguiti e quello che li ha emessi. Lo IMS Global Learning Consortium (IMS Global/IMS) dal 2018 ha definito gli Open Badge la cui specifica più recente (OpenBadges 2021) include lo Open Badges Vocabulary che formalizza classi, proprietà e valori per esprimere gli Open Badge come linked data rappresentati in sintassi JSON-LD. Grazie ai linked data la specifica garantisce l'estendibilità, il multilinguismo e l'uso di identificatori unici per identificare gli oggetti descritti e i valori delle proprietà. La conformità a questo schema di metadattazione consente agli istituti di formazione pubblici e privati di emettere certificazioni e titoli che possono essere interpretati e elaborati automaticamente ai fini della valutazione e del riconoscimento in contesti lavorativi e formativi diversi, a vantaggio dei soggetti che detengono i badge per o quali così si semplifica e si velocizza il conferimento dei dati.

Il Sistema Pubblico di Identità Digitale (SPID) permette di accedere a tutti i servizi online della Pubblica Amministrazione italiana con un'unica Identità Digitale. Oltre agli utenti (cittadini e imprese) il sistema comprende due tipologie di attori: i) i gestori dell'identità digitale (identity provider o IdP), accreditati da AgID per la gestione, la creazione e il rilascio agli utenti delle identità digitali; ii) i fornitori di servizi (service provider o SP), organizzazioni pubbliche o private che consentono l'accesso protetto ai propri servizi tramite l'identità digitale. Ciascuna di queste entità è descritta da un insieme di metadati (SPID\_RT n.d.) conformi allo standard SAML v2.0 (SAML-Metadata). I metadati sui gestori e i fornitori dei servizi sono consultabili e scaricabili in formato aperto da un registro pubblico (SPID\_REG n.d.). L'interazione tra gli Identity Provider e i Service Provider avviene tramite asserzioni espresse in SAML (SAML n.d) che contengono metadati che descrivono i soggetti, persone fisiche, enti o aziende (SPID\_AT n.d.). Quando un utente accede ai servizi di una pubblica amministrazione (es. Agenzia delle Entrate, INPS, etc.) viene rinvio al proprio provider SPID e dopo aver inserito le proprie credenziali è reindirizzato al servizio iniziale insieme ad un pacchetto di metadati<sup>3</sup> che lo identificano. Grazie a questi metadati i fornitori di servizio possono riconoscere l'identità dei soggetti, attuare politiche di controllo degli accessi e realizzare controlli di tipo autorizzativo.

---

<sup>3</sup> L'accesso al portale INPS con SPID comporta, previa autorizzazione esplicita dell'utente, il trasferimento del seguente pacchetto di metadati: Luogo di Nascita, Sesso, Numero di telefono mobile, Cognome, Domicilio digitale, Nome, Codice identificativo SPID, Data di nascita, Provincia di nascita, Codice Fiscale, Indirizzo di posta elettronica.

## 4. Metadati come elemento abilitante per il riuso dei dati aperti

I dati ad accesso aperto comprendono una molteplicità di tipologie e di produttori, da quelli generati nell'ambito amministrativo ai dati prodotti nell'ambito della ricerca finanziata con fondi pubblici secondo i principi FAIR (FAIR n.d.). Per condividere dati aperti non è sufficiente pubblicare file in formato aperto (es. csv, xml, json etc.), ma occorre corredarli di metadati che ne consentano la scoperta e il riuso attraverso cataloghi e motori di ricerca. Il Codice per l'Amministrazione Digitale (CAD 2021) all'Articolo 1 lettera l-ter precisa che i dati aperti "sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi metadati".

La Direttiva Europea sui Dati Aperti (Directive (EU) 2019/1024 2019), entrata in vigore il 19 Luglio 2019, si rivolge alle pubbliche amministrazioni, agli enti e alle imprese che operano in questo ambito con l'intento di promuovere l'utilizzo dei dati aperti e il riuso dei dati fornendo un quadro di riferimento omogeneo e condiviso che include la disponibilità di interfacce programmatiche (API) aperte e l'utilizzo di formati di dati aperti riconosciuti a livello internazionale. Il "portale ufficiale dei dati europei" data.europa.eu mette a disposizione un rilevante patrimonio di dati aperti, 174 Cataloghi, 36 Paesi e 1.444.438 Set di dati (al 21 aprile 2022), e include due cataloghi italiani:

- il Catalogo Italiano dei Dati Territoriali (Spatial Data Portal Italy) creato per dare seguito alla direttiva europea INSPIRE (<https://geodati.gov.it/geoportale>)
- il Portale dei dati aperti della Pubblica Amministrazione (Open Data Portal) dati.gov.it (<https://www.dati.gov.it/>)

Questa iniziativa europea consente di estrapolare tre punti che costituiscono una linea guida per qualunque progetto relativo alla produzione di metadati:

- a) l'uso di uno schema di metadattazione di riferimento comune;
- b) il consistente riuso di termini e metadati mutuati da altre specifiche per la definizione di tale schema;
- c) l'attenzione per la qualità dei metadati attraverso una metodologia di analisi e valutazione basata su specifiche metriche.

Al fine di garantire il corretto harvesting dei metadati e dei dati e per accrescere la capacità di scoperta (discoverability) dei datasets e dei servizi i vari cataloghi si devono conformare al "DCAT Application Profile for data portals in Europe" (DCAT-AP), definito da un'iniziativa congiunta di tre organismi europei (DG CONNECT, EU Publications Office, Interoperable Europe Programme). Questo profilo è basato sul Data Catalog Vocabulary (DCAT), un modello e un vocabolario espresso in RDF per descrivere i dataset e i servizi offerti da cataloghi pubblicati sul web (DCAT 2020), e consente di descrivere i dataset delle pubbliche amministrazioni europee e i relativi cataloghi in modo da supportare lo scambio dei metadati e le ricerche tra i portali dei dati aperti del settore pubblico (DCAT-AP 2022). Il profilo prevede una serie di classi (Agent, Catalogue, Dataset, Data Service, Document, Media type, Relationship, Provenance statement, Rights statement etc.) ciascuna delle quali ha delle proprietà (obbligatorie, raccomandate, opzionali), identificate tramite un URI, in gran parte mutate da namespace standard: Asset Description Metadata Schema (ADMS), Location Core Vocabulary (LOCN), DCMI Metadata Terms (DCTT), FOAF, Open Digital Rights Language (ODRL), W3C PROV Namespace, Schema.org, Simple Knowledge Organization System (SKOS), Software Package Data Exchange (SPDX), Vocabulary of a Friend (VOAF), vCARD.

Il portale europeo attua anche una metodologia di misurazione della qualità dei metadati basata sui criteri FAIR e mediante alcuni strumenti di analisi e valutazione viene periodicamente generato un rapporto che classifica la qualità dei metadati prodotti dai servizi e cataloghi attivati dai vari Paesi. La Figura 3 mostra gli indicatori di qualità, raggruppati nei cinque ambiti FAIR, che vengono misurati dallo strumento di valutazione della qualità dei metadati denominato Metadata Quality Assessment (MQA) e il punteggio riportato dal portale italiano dei dati aperti.

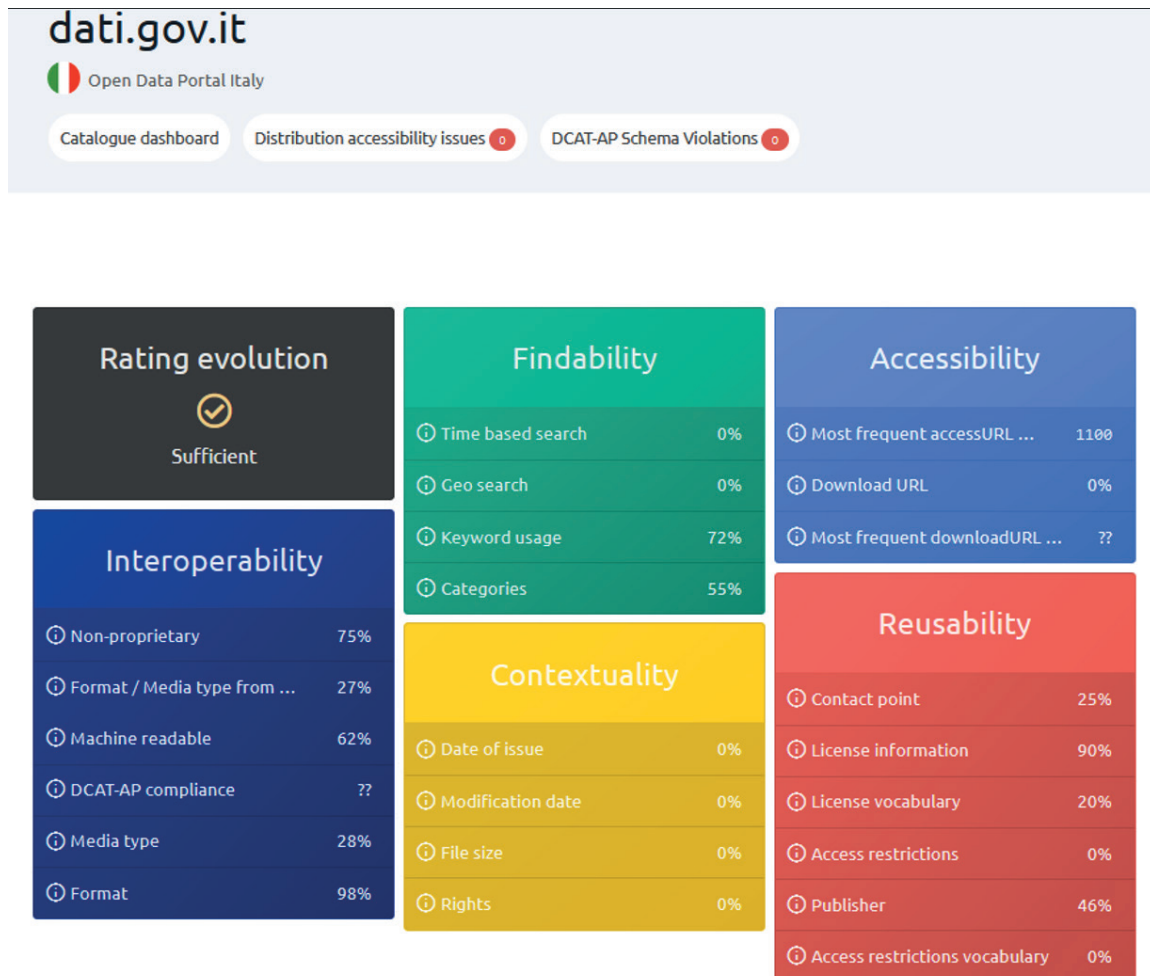


Figura 3 Dashboard del portale europeo relativa alla valutazione della qualità del portale italiano dei dati aperti dati.gov.it (<https://data.europa.eu/mqa/catalogues/dati-gov-it/?locale=en>)

Tra gli indicatori dell'interoperabilità Format misura se nella descrizione dei dataset è usata il meta-dato "dct:format", il valore assunto da Machine readable deriva dalla verifica se il formato del dataset è uno di quelli previsti (JSON, CSV, XLS, XLSX, vari formati RDF etc.), Non proprietary misura se il formato del file appartiene a quelli classificati come non proprietari (es. TXT, RTF, PNG, ZIP, GZIP, JPEG2002, TAR, TIFF etc.), l'indicatore Format/Media type from vocabulary valuta se i valori dei due metadati dct:format e dcat:mediaType appartengono rispettivamente ai vocabolari controllati previsti da data.europa e ai Media Types definiti da IANA (Media Types 2002).

La valutazione dei due portali italiani nel rapporto del 3 Aprile 2022 è “appena sufficiente” per dati.gov.it e “scarso” per il catalogo dei dati territoriali a dimostrazione del fatto che occorrono ulteriori investimenti (organizzativi, infrastrutture, risorse umane) per migliorare la metadatozione dei dati aperti e per renderli effettivamente utilizzabili. Il CAD, con l’Art.1 e il Titolo 5, le Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo Pubblico (LGNVPIP 2020) e le varie versioni del Piano Triennale per l’informatica nella Pubblica Amministrazione hanno promosso la produzione e la diffusione dei dati aperti, attraverso l’attivazione del portale italiano e la definizione del profilo italiano dei metadati DCAT-AP\_IT a cui dovrebbero conformarsi tutte le amministrazioni per alimentare il catalogo nazionale (DCAT-AP\_IT 2022). Lo scorso 15 dicembre 2021 è entrato in vigore il decreto legislativo n. 200/2021 che recepisce la direttiva (UE) 2019/1024 sui Dati Aperti e l’AgID ha attivato il coordinamento di un tavolo tecnico con i Responsabili della Transizione per definire delle Linee guida contenenti regole tecniche relative al tema dei dati aperti, la cui adozione secondo l’attuale versione del Piano Triennale (CAP2.LA08) è prevista per dicembre 2022. Infine, un’ulteriore interessante iniziativa da citare è il portale Ocds (OCDS\_Port n.d.) realizzato dall’Agenzia Nazionale Anticorruzione (ANAC) che consente di consultare e confrontare le informazioni degli appalti pubblici contenute nella banca dati ANAC<sup>4</sup>. Il portale segue le specifiche dello standard OCDS, elaborate dall’organizzazione non-profit Open Contracting Partnership, che comprende alcuni schemi di metadati concepiti per modellare i soggetti e rappresentare tutte le entità dei contratti e che sono espressi in JSON (OCDS, n.d.).

## 5. Metadati e interoperabilità applicativa

La disponibilità di infrastrutture di rete sempre più performanti per velocità, volumi di traffico sostenibile e ridondanza ha consentito di digitalizzare molti processi sia delle pubbliche amministrazioni sia nelle aziende private, utilizzando modalità di cooperazione basate sull’integrazione dei servizi distribuiti in rete e operanti nel Cloud. Le applicazioni interagiscono e cooperano tra di loro grazie alle interfacce programmatiche (API) e a web services che consentono di esporre funzionalità per la creazione, la modifica e la cancellazione dei dati.

Le architetture orientate ai servizi si sono particolarmente affermate a partire dagli anni Duemila e la W3C “Web services Architecture” definisce un Web service come un “software system designed to support interoperable machine-to-machine interaction over a network” (W3CWS 2004). Inizialmente i WS erano basati sul protocollo SOAP, concepito per scambiare messaggi in formato XML, tra un mittente e un ricevente (sender e receiver), in modo indipendente dal sottostante protocollo di rete (di solito HTTP/HTTPS). Più di recente si sono diffusi i web service realizzati in conformità allo stile architetturale Representational State Transfer (REST) (Fielding 2000) che ormai è il più comunemente utilizzato per sviluppare interfacce programmatiche dette Web API o REST API. REST si basa su alcuni principi tra i quali il più pertinente per la tematica che stiamo trattando è il concetto astratto di risorsa, “Any information that can be named can be a resource”. Ogni risorsa è identificata da un identificatore (URI), in un determinato istante di tempo assume

---

<sup>4</sup> ANAC ha lanciato una competizione (datathon) nella quale i concorrenti sono stati invitati a proporre, entro il 22 maggio 2022, esempi di utilizzo dei dati Ocds (Open Contracting Data Standard) dell’Autorità.

uno stato chiamato “resource representation” costituito dai dati (es. un’immagine, un documento), i metadati che descrivono la risorsa e i link ipermediali che permettono di modificarne lo stato. In entrambe gli ambiti, SOAP e REST, sono state definite modalità di descrizione dei servizi. Il Web Services Description Language (WSDL) è solitamente utilizzato per pubblicare le specifiche dei web service SOAP in base a XML Schema avvalendosi di elementi come types (descrive i tipi di messaggi che il servizio invia e riceve), interface (attraverso gli elementi operation, input e output descrive le funzioni astratte che il Web service fornisce), binding (indica con quale protocollo e quali operazioni si accede al servizio), service (mediante l’elemento endpoint il cui valore è un URI descrive dove accedere al servizio). La specifica WSDL 2.0 (WSDL 2.0 2007) consente anche la descrizione dei web services di tipo RESTful. La OpenAPI Specification (OAS) è uno schema di metadate appositamente concepito per descrivere, attraverso un OpenAPI document in sintassi JSON, questa tipologia di web services avvalendosi di una serie di oggetti (OpenAPI Object, Info Object, Contact Object, License Object, Server Object, Paths Object, Operation Object, etc.) e delle relative proprietà/metadati (OAS n.d.). I metadati strutturati di WSDL e OAS permettono di testare e integrare rapidamente le API esposte da altri servizi e di generare, pubblicare e condividere nuove interfacce programmatiche consentendo agli sviluppatori di generare automaticamente il codice applicativo (es. classi e metodi Java, librerie PHP) senza doversi preoccupare dei dettagli e di svolgere attività soggette ad errore.

Nel contesto nazionale nel corso degli ultimi anni si è assistito ad un notevole sviluppo di sistemi orientati ai servizi, basti pensare alle modalità di pagamento elettronico (dall’internet banking all’uso delle carte di credito tramite POS virtuali fino al sistema di pagamento per le Pubbliche amministrazioni PagoPA), alle transazioni commerciali business-to-business (B2B), al recupero delle attestazioni ISEE, al Sistema informativo sulle operazioni degli enti pubblici (SIOPE+) e l’Anagrafe Nazionale della Popolazione Residente (ANPR). Il Sistema d’Interscambio (SdI) dell’Agenzia delle Entrate per l’invio e la ricezione delle fatture elettronica con le PA (SDI n.d.) e il Nodo di smistamento degli ordini di acquisto delle amministrazioni pubbliche (NSO) gestito dalla Ragioneria dello Stato (NSO n.d.) espongono API che permettono ai gestionali locali (es. i sistemi di gestione ordini e contabilità) di integrarsi per gestire lo scambio delle fatture e degli ordini in formato elettronico. È utile notare come da un lato gli oggetti digitali così scambiati (certificati, dati relativi al reddito, incassi e pagamenti delle pubbliche amministrazioni ai propri tesorieri/cassieri, fatture, ordini di acquisto etc.) sono accompagnati da metadati strutturati che ne permettono la corretta costruzione e interpretazione e dall’altro i servizi invocati per cooperare sono a loro volta descritti tramite appropriati metadati che permettono lo sviluppo delle applicazioni (programmi) che li utilizzano. Il Dipartimento per la trasformazione digitale del Ministero per l’innovazione tecnologica e la transizione digitale (MITD) in collaborazione con l’AgID ha definito le linee guida e il Modello di Interoperabilità allo scopo di rendere possibile la collaborazione tra le Pubbliche Amministrazioni e tra queste e soggetti terzi attraverso interfacce programmatiche (API) conformi agli standard consolidati<sup>5</sup>. Inoltre Il Dipartimento supervisiona la realizzazione della Piattaforma Digitale Nazionale

---

<sup>5</sup> AgID con la Determinazione n. 547 del 1 ottobre 2021, ha adottato le Linee guida sull’interoperabilità tecnica delle Pubbliche Amministrazioni e le Linee guida Tecnologie e standard per la sicurezza dell’interoperabilità tramite API dei sistemi informatici [https://www.agid.gov.it/sites/default/files/repository\\_files/547\\_dt\\_dg\\_n\\_547\\_1\\_ott\\_2021\\_adozione\\_lg\\_interoperabilit\\_tecnica\\_e\\_sicurezza.pdf](https://www.agid.gov.it/sites/default/files/repository_files/547_dt_dg_n_547_1_ott_2021_adozione_lg_interoperabilit_tecnica_e_sicurezza.pdf).

Dati (PDND), affidata alla società PagoPA Spa, che renderà possibile l'interoperabilità dei sistemi informativi e delle basi di dati delle pubbliche amministrazioni e dei gestori di servizi pubblici<sup>6</sup>.

## 6. Considerazioni e prospettive

Le iniziative citate e gli esempi esaminati nei paragrafi precedenti mostrano come i metadati abbiano ormai raggiunto un elevato livello di pervasività che si estende dai ben noti strumenti di ricerca e cataloghi in rete fino all'interoperabilità semantica tra diversi contesti applicativi e domini funzionali eterogenei. Questo scenario comprende sistemi e servizi che interagiscono e cooperano secondo modalità guidate dai metadati per integrazione diretta, attraverso web service e interfacce programmatiche, oppure mediante la mappatura di diverse sorgenti dati in uno schema di metadati condiviso o, infine, attraverso processi di trasformazione, riconciliazione e integrazione di dati in origine diversi avvalendosi di specifici insiemi di metadati.

Gli schemi di metadati e le ontologie sono gli strumenti attraverso i quali è possibile realizzare soluzioni come il Single Digital Gateway, gli Open Badge e i sistemi di identità federata che permettono di ridisegnare molti processi adottando i principi dell'only-once e della cooperazione interistituzionale. I metadati sono un fattore abilitante per l'evoluzione dell'ecosistema digitale formato dalla rete di servizi e applicazioni e per implementare processi complessi nei quali operano in ambito transazionale enti e soggetti del settore pubblico e di quello privato. Svariate iniziative e progetti, a livello europeo e italiano, hanno dato luogo alla definizione di schemi di metadati e ontologie che coprono porzioni significative dell'universo informativo di questo bacino di applicazione e hanno promosso la diffusione dei metadati e la conoscenza delle tecniche di metadatezione. Tuttavia i sistemi informativi e le applicazioni delle Pubbliche Amministrazioni usano ancora in modo insufficiente e parziale schemi di metadati standardizzati. Tra le cause di questa carenza si possono annoverare l'esistenza di molte applicazioni datate, la tendenza delle aziende a privilegiare soluzioni verticali e poco standardizzate per mantenere vincolati i clienti ai propri prodotti (fenomeno noto come vendor lock-in) e la maggiore priorità degli adeguamenti dei software alle linee guida sulla usabilità, l'accessibilità e la sicurezza. Ma nel settore pubblico vanno posti in evidenza anche altri due fattori inibitori: la nota inerzia al cambiamento e i limiti delle risorse economiche disponibili per lo sviluppo di soluzioni innovative, per la formazione delle competenze e per l'acquisizione di nuove risorse umane. Alcune linee d'azione del PNRR sono destinate a dare un impulso al processo di cambiamento, ma resta da chiarire quali risorse saranno effettivamente disponibili dopo il 2026 per garantire la gestione, la manutenzione e l'evoluzione di quanto realizzato, dato che fino ad oggi tutti i piani di trasformazione digitale e innovazione nelle PA sono stati declinati escludendo costi aggiuntivi, anzi prevedendo riduzione di spesa per gli investimenti nel settore ICT.

Occorre ancora molto lavoro e impegno affinché la cultura dei metadati si consolidi, ovvero af-

---

<sup>6</sup> In base al cronoprogramma del Piano Triennale è previsto che entro Maggio 2022 AgID emetta le Linee Guida sulla Piattaforma Digitale Nazionale Dati e che entro Dicembre 2022 siano completati gli sviluppi tecnologici delle componenti della Piattaforma Digitale Nazionale Dati (PDND) e del Catalogo Nazionale dei Dati necessarie per integrare il Single Digital Gateway. Il 16 giugno 2022 AgID ha pubblicato la consultazione pubblica fino al 17 luglio le "Linee Guida relative alle regole tecniche sull'apertura dei dati e il riutilizzo dell'informazione del settore pubblico" <https://www.agid.gov.it/agenzia/stampa-e-comunicazione/notizie/2022/06/16/open-data-consultazione-linee-guida>.

finché si diffonda e permei un atteggiamento mentale che non consideri la metadattazione soltanto come un insieme di strumenti e di tecniche ma come un approccio sistematico che deve accompagnare l'intero ciclo della trasformazione digitale iniziando dalla fase di riorganizzazione e di progettazione dei processi, proseguendo con la reingegnerizzazione e lo sviluppo dei sistemi informativi fino a giungere alla realizzazione e all'attivazione di nuove applicazioni.

Il primo passo consiste nella modellazione dei processi e dei sistemi informativi attraverso il riuso di schemi di metadati, ontologie e vocabolari controllati già consolidati. Esiste già un consistente insieme di schemi ma occorre promuovere la conoscenza dei contesti e le finalità delle iniziative che li hanno realizzati, costruendo e alimentando repertori, come quello europeo di Joinup e il portale delle ontologie italiano, che consentano ai soggetti che si occupano dell'analisi e della modellazione di avvalersi dei modelli già disponibili eventualmente estendendoli con nuove definizioni per coprire ambiti ancora non trattati.

Il secondo passo comprende lo sviluppo di strategie e l'attuazione di metodologie per la misurazione e il controllo costanti della qualità dei metadati prodotti e utilizzati. Se infatti i metadati non rispettano gli schemi e i vocabolari controllati adottati, sono creati in modo parziale o vengono aggiornati in modo irregolare l'intero ecosistema che su di essi si basa non può funzionare in modo soddisfacente.

Il terzo passo include la collaborazione intersettoriale e lo sviluppo di specifiche competenze professionali. Molto spesso nelle aziende che producono software non si dispone delle adeguate conoscenze dei domini funzionali per cogliere le specifiche e delle competenze per razionalizzare i processi riprogettandoli e semplificandoli avvalendosi delle potenzialità offerte dai metadati e dalle ontologie. Il settore informatico è caratterizzato da un elevato livello di tecnicismo e da una forte specializzazione, una logica che rischia di essere poco fruttuosa ai fini della revisione dei processi e della modellazione dei sistemi informativi. Al contrario è necessario un approccio intersettoriale e orientato al riuso dei metadati che coinvolga figure adeguatamente preparate in grado di pensare in termini di schemi e di ontologie. Per poter attuare questa trasformazione occorre investire maggiormente, anche a livello accademico, per creare figure professionali dotate di competenze sull'uso e la definizione di schemi di metadati, di vocabolari controllati e la costruzione di ontologie. Alcuni argomenti (es. linked data, RDF, OWL) vengono già trattati in alcuni corsi di informatica o ingegneria informatica (per lo più di secondo livello), tuttavia le figure professionali da formare non dovrebbero necessariamente appartenere all'ambito scientifico e tecnologico, ma potrebbero provenire anche dall'ambito umanistico attraverso un'adeguata formazione sui linked data, sulle sintassi di rappresentazione e sugli standard esistenti. Un'impostazione multidisciplinare e un nuovo umanesimo declinato sulla base delle potenzialità che il contesto ICT mette a disposizione possono assicurare lo sviluppo di competenze trasversali e intersettoriali che uniscano all'apprendimento delle sintassi e delle tecniche di rappresentazione dei metadati le conoscenze per progettare secondo astrazioni e modelli che si avvalgono degli schemi e delle ontologie. Nuove figure professionali con questo tipo di preparazione possono sorgere da una forte collaborazione tra l'accademia, le aziende che operano nel mercato ITC e il settore pubblico e privato che chiedono soluzioni volte a semplificare e razionalizzare i propri processi, assicurando così le necessarie sinergie col mondo del lavoro e l'aggiornamento costante del personale.



## Riferimenti bibliografici

AgID (Agenzia per l'Italia Digitale).2022. *Attuazione Misure PNRR*. <https://www.agid.gov.it/it/agenzia/attuazione-misure-pnrr>

ArCo. 2017. MiC (Ministero della Cultura). *Progetto Arco – Architettura della conoscenza*. Ultimo accesso 5 Maggio 2022. <https://dati.beniculturali.it/progetto-arco-architettura-della-conoscenza/>

Bagley, Philip.1968. *Extension of programming language concepts*. Philadelphia: University City Science Center. Archived (PDF) from the original on November 30, 2012. <https://web.archive.org/web/20121130091053/http://www.dtic.mil/dtic/tr/fulltext/u2/680815.pdf>

CAD. 2021 *Decreto Legislativo 7 marzo 2005, n. 82 - Codice dell'amministrazione digitale*. Successive modifiche (*Ultimo aggiornamento all'atto pubblicato il 31/12/2021*). <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2005-03-07;82!vig=>

DCAT. 2020. *Data Catalog Vocabulary (DCAT) - Version 2 - W3C Recommendation 04 February 2020*. <https://www.w3.org/TR/vocab-dcat-2/>

DCAT-AP. 2022. *Data Catalog Vocabulary (DCAT) Application Profile for data portals*. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/210>

DCAT-AP\_IT. 2022. *DCAT-AP\_IT v1.1 – Profilo italiano di DCAT-AP*. AgID (Agenzia per l'Italia Digitale). <https://dati.gov.it/content/dcat-ap-it-v10-profilo-italiano-dcat-ap-0>

Directive (EU) 2019/1024. 2019 *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on opendata and the re-use of public sector information*. Ultimo accesso 5 Maggio 2022. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L1024>

eduPerson. 2016. *eduPerson Object Class Specification (201602)*. Internet2. February 2016. <http://software.internet2.edu/eduperson/internet2-mace-dir-eduperson-201602.html>

e-GOV. n.d. *Releases for e-Government Core Vocabularies solution*. Ultimo accesso 5 Maggio 2022. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/e-government-core-vocabularies>

EUDati. 2020. *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni. Una strategia europea per i dati*. Document 52020DC0066.2020.<https://eurlex.europa.eu/legal-content/IT/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066>

FAIR. n.d. *FAIR (Findability, Accessibility, Interoperability, Interoperabilità and Reusability) Principles*. Ultimo accesso 5 Maggio 2022. <https://www.go-fair.org/fair-principles/>

Fielding, Roy T. 2000. *Architectural Styles and the Design of Network-based Software Architectures* University of California, Irvine, 2000. <https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>

GÉANT.n.d. Ultimo accesso 5 Maggio 2022. <http://www.geant.org/>

IBM.2014. *100 ans d'IBM en France*. <http://centenaireibmfrance.blogspot.com/2014/04/1955-terme-ordinateur-invente-par-jacques-perret.html>

INSPIRE. 2007. *Directive 2007/2/EC of March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*. <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32007L0002>

INSPIRE\_REG\_IT. n.d. *INSPIRE Italia Registry*. Ultimo accesso 5 Maggio 2022. <https://registry.geodati.gov.it/registry>

INSPIRE\_TG. 2022 *Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007*. <https://github.com/INSPIRE-MIF/technical-guidelines/blob/2022.1/metadata/metadata-iso19139/metadata-iso19139.adoc#conformance-classes-for-spatial-data-services>

Joinup-CV. 2022. *e-Government Core-Vocabularies*. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/e-government-core-vocabularies/release/200-0>

LGNVPIP. 2020. *Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo Pubblico* 13 Febbraio 2020. <https://docs.italia.it/media/pdf/lg-patrimonio-pubblico/bozza/lg-patrimonio-pubblico.pdf>

Media Types. 2022. *Media Types*. Internet Assigned Numbers Authority (IANA) 2022. <https://www.iana.org/assignments/media-types/media-types.xhtml>

NSO n.d. *Nodo di smistamento degli ordini di acquisto delle amministrazioni pubbliche (NSO)*. MEF (Ministero dell'Economia e Finanze) Ragioneria Generale dello Stato. [https://www.rgs.mef.gov.it/VERSIONE-I/e\\_government/amministrazioni\\_pubbliche/acquisti\\_publici\\_in\\_rete\\_apir/nodo\\_di\\_smistamento\\_degli\\_ordini\\_di\\_acquisto\\_delle\\_amministrazioni\\_pubbliche\\_nso/](https://www.rgs.mef.gov.it/VERSIONE-I/e_government/amministrazioni_pubbliche/acquisti_publici_in_rete_apir/nodo_di_smistamento_degli_ordini_di_acquisto_delle_amministrazioni_pubbliche_nso/)

OAIS. 2012. *Reference Model For An Open Archival Information System (OAIS)*. Consultative Committee for Space Data Systems. June 2012. Standard ISO 14721:2012. <https://public.ccsds.org/pubs/650x0m2.pdf>

OASIS. n.d. *Organization for the Advancement of Structured Information Standards*. Ultimo accesso 5 Maggio 2022. <https://www.oasis-open.org/>

OAS. n.d. *OpenAPI Specification Version 3.0.3* Ultimo accesso 5 Maggio 2022. <https://swagger.io/specification/>

OpenBadges 2.0. 2021. *Open Badges v2.0 IMS Final Release*. <https://www.imsglobal.org/sites/default/files/Badges/OBv2p0Final/index.html>

OCDS. n.d. *The Open Contracting Data Standard*. Ultimo accesso 5 Maggio 2022. <https://www.open-contracting.org/data-standard/>

OCDS\_Port. n.d. *Portale dei dati aperti dell'Autorità Nazionale Anticorruzione*. Ultimo accesso 5 Maggio 2022. <https://dati.anticorruzione.it/opendata/ocds>

PDF/A. 2005. *ISO 19005-1:2005 Document Management - Electronic document file format for long term preservation - Part 1: Use of PDF 1.4 (PDF/A-1)*. <https://www.iso.org/standard/38920.html>

PTIPA. 2021. *Piano triennale per l'informatica nella PA – Aggiornamento 2021 – 2023 (versione notificata alla Commissione Europea)* - Agenzia per l'Italia Digitale e Dipartimento per la Trasfor-

- mazione Digitale. Roma, ottobre 2021. [https://www.agid.gov.it/sites/default/files/repository\\_files/piano\\_triennale\\_per\\_linformatica\\_nella\\_pubblica\\_amministrazione\\_2021-2023.pdf](https://www.agid.gov.it/sites/default/files/repository_files/piano_triennale_per_linformatica_nella_pubblica_amministrazione_2021-2023.pdf)
- RDF. 2014. *RDF 1.1 Concepts and Abstract Syntax*. W3C Recommendation 25 February 2014. <https://www.w3.org/TR/rdf11-concepts/>
- REFEDS. n.d. REFEDS (the Research and Education FEDerations group). Ultimo accesso 5 Maggio 2022. <https://refeds.org/>
- RFC 2252. 1997. *Lightweight Directory Access Protocol (v3): Attribute Syntax Definitions*. RFC 2252 December 1997. <https://datatracker.ietf.org/doc/html/rfc2252>
- RFC 2256. 1997. *A Summary of the X.500(96) User Schema for use with LDAPv3*. RFC 2256 December 1997. <https://www.ietf.org/rfc/rfc2256.txt>
- RFC 2798. 2000. *Definition of the inetOrgPerson LDAP Object Class*. RFC 2798 April 2000. <https://www.ietf.org/rfc/rfc2798.txt>
- RFC 4519. 2006. *Lightweight Directory Access Protocol (LDAP): Schema for User Applications*. RFC 4519 June 2006. <https://www.ietf.org/rfc/rfc4519.txt>
- SAML. n.d. *Security Assertion Markup Language (SAML) v2.0*. Ultimo accesso 5 Maggio 2022. <https://www.oasis-open.org/standards/>
- SHACH. 2022. *SCHAC - SCHEMA for Academia*. <https://wiki.refeds.org/display/STAN/SCHAC>
- SDG. n.d. *The single digital gateway and Your Europe*. Ultimo accesso 5 Maggio 2022. [https://ec.europa.eu/growth/single-market/single-digital-gateway\\_it](https://ec.europa.eu/growth/single-market/single-digital-gateway_it)
- SDI n.d. *Fatturazione elettronica verso la Pubblica amministrazione. Sistema di Interscambio*. Ultimo accesso 5 Maggio 2022. <https://www.fatturapa.gov.it/it/sistemainterscambio/cose-il-sdi/>
- SPID\_AT. n.d. *SPID Tabella attributi*. Ultimo accesso 5 Maggio 2022. [https://www.agid.gov.it/sites/default/files/repository\\_files/regole\\_tecniche/tabella\\_attributi\\_idp\\_v1\\_0.pdf](https://www.agid.gov.it/sites/default/files/repository_files/regole_tecniche/tabella_attributi_idp_v1_0.pdf)
- SPID\_REG. n.d. *SPID Registry*. Ultimo accesso 5 Maggio 2022. <https://registry.spid.gov.it/>
- SPID\_RT.n.d. *SPID Metadata*. <https://github.com/italia/spid-regole-tecniche/blob/96287a228c8db-f817a4b0b1c41e234d4e231a830/metadata.rst>
- TALLIN. 2017. *Tallinn Declaration on eGovernment*. [https://ec.europa.eu/newsroom/document.cfm?doc\\_id=47559](https://ec.europa.eu/newsroom/document.cfm?doc_id=47559)
- Berners-Lee, Tim. 2006. *Linked Data—Design Issues*. <http://www.w3.org/DesignIssues/Linked-Data.html>
- Vallejo, Irene. 2021. *Payprus. L'infinito in un giunco. La grande avventura del libro nel mondo antico*. Firenze: Giunti Editore/Bompiani.
- Wiederhold, Gio. 1997. *Database Design*, Tokio: McGraw-Hill computer science series.
- W3CWS. 2004. *Web Services Glossary W3C Working Group Note*. 11 February 2004. <https://www.w3.org/TR/2004/NOTE-ws-gloss-20040211/#webservice>

WSDL 2.0. 2007 *Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language*. <http://www.w3.org/TR/wsdl20>

XSD. 2012. *W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures W3C Recommendation*. 5 April 2012 <https://www.w3.org/TR/xmlschema11-1/> e *W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes W3C Recommendation* 5 April 2012 <https://www.w3.org/TR/xmlschema11-2/>

## Wikibase, or The search for the unicorn

Giovanni Bergamin<sup>(a)</sup>

a) <https://orcid.org/0000-0002-2912-5662>

**Contact:** Giovanni Bergamin, [giovanni.bergamin@gmail.com](mailto:giovanni.bergamin@gmail.com)

**Received:** 30 May 2022; **Accepted:** 7 July 2022; **First Published:** 15 September 2022

### ABSTRACT

The aim of this paper is to study the technological implementation of the emergent bibliographic models (IFLA LRM in particular) taking into account one of the most widespread platforms for the semantic web, namely Wikibase. Different initiatives of implementation of LRM have been taken into account, included: a) a prototype cataloging interface; b) the implementation of the new cataloging system for the Bibliothèque Nationale de France (Bnf); c) a test of use of one of the features of the *Wikibase data model* – namely the “qualifier” – to find a sustainable solution for the LRM *nomen* entity. Wikibase and his data model cannot be considered the magical unicorn that solves all problems. More in-depth analysis and tests are needed, but – as an intermediate result – we can consider Wikibase a promising platform also in the bibliographic field with a low entry barrier.

### KEYWORDS

Nomen (IFLA LRM); Wikibase data model; MARC; Schema.org; Linked open data.

## Wikibase, o La ricerca dell'unicorno

### ABSTRACT

Lo scopo di questo lavoro è quello di studiare l'implementazione tecnologica dei modelli bibliografici emergenti (IFLA LRM in particolare) tenendo conto di una delle piattaforme più diffuse per il web semantico, ovvero Wikibase. Sono state prese in considerazione diverse iniziative di implementazione di LRM, tra cui: a) un prototipo per una interfaccia di catalogazione; b) l'attuazione del nuovo sistema di catalogazione della Bibliothèque Nationale de France (Bnf); c) un test di utilizzo di una delle caratteristiche del modello di dati Wikibase – ovvero il “qualificatore” – per trovare una soluzione sostenibile per l'entità LRM *nomen*. Wikibase e il suo modello di dati non possono essere considerati l'unicorno magico che risolve tutti i problemi. Occorrono analisi e test più approfonditi, ma – come risultato intermedio – possiamo considerare Wikibase una piattaforma promettente anche in ambito bibliografico con una bassa barriera all'ingresso.

### PAROLE CHIAVE

Nomen (IFLA LRM); Wikibase data model; MARC; Schema.org; Linked open data.

Nella discussione in corso sui modelli dei dati bibliografici, Wikibase occupa certamente un posto singolare. Con Wikibase ci si riferisce alla piattaforma tecnologica con la quale viene offerto il servizio Wikidata<sup>1</sup>. Il servizio Wikidata è operativo da ormai dieci anni. Nasce come base dati di supporto alle varie versioni linguistiche di Wikipedia, ma si propone anche di essere una base di conoscenza generale “for the world at large”. Può trovare posto in Wikidata tutto quello che si riferisce a un’entità concettuale o materiale chiaramente identificabile e descritta utilizzando fonti affidabili e disponibili al pubblico: in altre parole non è necessario che una “informazione strutturata” rispetti il canone della *notability* che caratterizza Wikipedia.<sup>2</sup> Chiunque, incluse le biblioteche, può utilizzare Wikidata per creare, pubblicare e utilizzare metadati come *linked open data*: quello che caratterizza Wikidata è la modalità di lavoro bottom-up e altamente inclusiva (Bianchini e Sardo 2022).

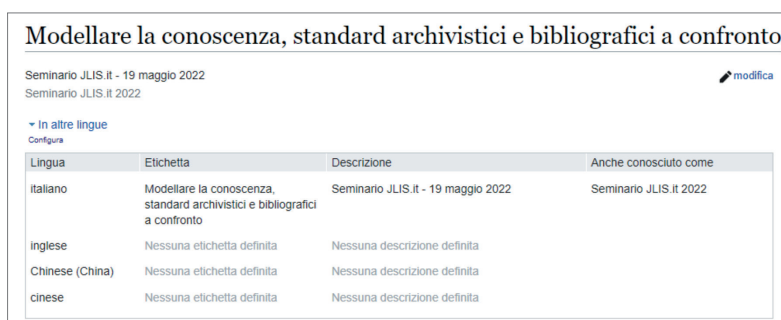
In questo intervento l’interesse per Wikibase sarà centrato su un particolare aspetto della piattaforma tecnologica, ovvero sul modello dei dati: il *Wikibase data model*.<sup>3</sup> Si tratta in realtà di un metamodello che offre la struttura di base per costruire ontologie relative a particolari domini (ad esempio il dominio dei dati bibliografici). Scopo di questo intervento è valutare le potenzialità del *Wikibase data model* sia nel contesto dei modelli di dati bibliografici, sia nel contesto dei *linked open data* dove l’apertura ad altre tipologie di dati si presenta come un requisito fondamentale.

## 1. Le particolarità del Wikibase data model

In questo paragrafo vengono presentate due particolarità del *Wikibase data model*. La prima riguarda la struttura dei dati per esprimere il livello di informazione di base. La seconda riguarda la possibilità di esprimere anche informazioni che completano una determinata tripla, ovvero la possibilità di “arricchire” una tripla.

Per creare a livello base un elemento (un *wikibase:item*) Wikibase prevede la possibilità di indicare tre proprietà:

- *etichetta (label)*;
- *descrizione (description)*;
- *Anche conosciuto come (alias)*.



Lingua	Etichetta	Descrizione	Anche conosciuto come
italiano	Modellare la conoscenza, standard archivistici e bibliografici a confronto	Seminario JLIS.it - 19 maggio 2022	Seminario JLIS.it 2022
inglese	Nessuna etichetta definita	Nessuna descrizione definita	
Chinese (China)	Nessuna etichetta definita	Nessuna descrizione definita	
cinese	Nessuna etichetta definita	Nessuna descrizione definita	

Figura 1: Esempio di inserimento in Wikidata (base di test)

<sup>1</sup> <https://wikidata.org>.

<sup>2</sup> Per la *notability*: in Wikipedia: <https://en.wikipedia.org/wiki/Wikipedia:Notability>; in Wikidata: <https://www.wikidata.org/wiki/Wikidata:Notability/en>.

<sup>3</sup> Descritto in (Wikibase/DataModel 2020).

In Figura 1 troviamo (nella base dati Wikidata di test) un nuovo elemento creato ad hoc <<https://test.wikidata.org/wiki/Q224835>>. La creazione di un nuovo elemento comporta l'immediata pubblicazione come *linked open data* dei dati inseriti. In Figura 2 vengono riportate – usando le convenzioni della serializzazione Turtle<sup>4</sup> – alcune delle triple pubblicate a partire dalla prima che ci informa che stiamo descrivendo un *wikibase:item* (un *item* definito dalla ontologia *wikibase*). Le tre triple successive riportano lo stesso *oggetto* con il contenuto dell'*etichetta* (o *label*). Quello che cambia è il predicato che fa riferimento a tre ontologie differenti evidenziate dai differenti prefissi: *rdfs*, *skos* e *schema*. In pratica potremmo dire che queste tre triple “dicono la stessa cosa”. La quinta tripla è dedicata alla *descrizione* (*description*), mentre la sesta all'indicazione della proprietà *Anche conosciuto come* (*Alias*).

Soggetto	Predicato	Oggetto
1) wd:Q224835	a	wikibase:Item
2) wd:Q224835	rdfs:label	“Modellare la conoscenza, standard archivistici e bibliografici a confronto”@it
3) wd:Q224835	skos:prefLabel	“Modellare la conoscenza, standard archivistici e bibliografici a confronto”@it
4) wd:Q224835	schema:name	“Modellare la conoscenza, standard archivistici e bibliografici a confronto”@it
5) wd:Q224835	schema:description	“Seminario JLIS.it – 19 maggio 2022”@it
6) wd:Q224835	skos:altLabel	“Seminario JLIS.it 2022”@it
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> . @prefix wikibase: <http://wikiba.se/ontology#> . @prefix skos: <http://www.w3.org/2004/02/skos/core#> . @prefix schema: <http://schema.org/> . @prefix wd: <http://test.wikidata.org/entity/> .		

Figura 2: Pubblicazione come LOD<sup>5</sup> dei dati di base inseriti (vedi Figura 1) -- serializzazione Turtle <<https://test.wikidata.org/wiki/Special:EntityData/Q224835.ttl>>

Per le informazioni di base il *Wikibase data model* sceglie di chiamare in causa tre riferimenti ontologici consolidati:

- il prefisso *rdfs* fa riferimento al *Resource Description Framework Schema* che si definisce come una estensione del vocabolario RDF (Resource Description Framework) di base: “RDF Schema provides a data-modelling vocabulary for RDF data. RDF Schema is an extension of the basic RDF vocabulary”;<sup>6</sup>
- il prefisso *skos* fa riferimento a *Simple Knowledge Organisation System*: uno strumento costruito con l'intento di facilitare la pubblicazione sul web – utilizzando la grammatica del web semantico – soprattutto di tesauri e classificazioni;<sup>7</sup>

<sup>4</sup> RDF 1.1 Turtle: Terse RDF Triple Language <https://www.w3.org/TR/turtle/>

<sup>5</sup> LOD sta per *linked open data*. Il disegno iniziale di Tim Berners-Lee del 2006: <https://www.w3.org/DesignIssues/LinkedData>

<sup>6</sup> <https://www.w3.org/TR/rdf-schema/>

<sup>7</sup> <https://www.w3.org/TR/skos-reference/>

- il prefisso *schema* fa infine riferimento all'ontologia *schema.org* ovvero da un'ontologia creata a seguito di un accordo (2011) tra i principali motori di ricerca con l'obiettivo di rispondere alla domanda "quali sono le entità che vengono cercate nei motori di ricerca?" Viene utilizzato anche per la proprietà *descrizione* (*description*) oltre che per la proprietà *etichetta* (*label*).<sup>8</sup>

È l'applicazione del principio di riuso dei vocabolari («Vocabulary re-use is generally encouraged on the semantic web»), ma anche del principio di fornire informazioni nelle modalità che in un determinato contesto gli utenti si aspettano: « For example, *RDFS label* is used for labels, but *SKOS prefLabel* and *schema.org name* would be just as suitable. The Wikidata linked data service actually provides label data in each of these, since LOD consumers may expect labels to be given in a specific form» (Erxleben et al. 2014).<sup>9</sup>

La seconda particolarità del *Wikibase data model* che è opportuno mettere in evidenza è la possibilità di arricchire tutte le triple con qualificatori (*qualifiers*) o con riferimenti (*references*). Dal punto di vista tecnico – come prevede RDF – ogni tripla può essere "reificata" e diventare soggetto di un'altra tripla. È un aspetto dei LOD molto discusso e che introduce inevitabili complessità nella gestione delle triple.<sup>10</sup> Si tratta di un requisito fondamentale per Wikidata. Per poter essere utilizzata come fonte secondaria Wikidata deve documentare le sue dichiarazioni (triple) con il contesto (*qualifier*) e la fonte (*references*). Ad esempio<sup>11</sup> la tripla <Italia> <popolazione> <"50.199.700"> 12 ha come contesto <nel> <"1960"> e come fonte <fonte dell'affermazione> <World Bank Open Data>. Così la tripla <Italia> <popolazione> <"60.317.000"> 13 ha come contesto <nel> <"2020"> e come fonte <fonte dell'affermazione> <ISTAT>.

## 2. Il Wikiprogetto Libri di Wikidata

I dati bibliografici sono presenti in Wikidata e presi in conto in uno specifico progetto: il *Wikiprogetto Libri*. L'obiettivo del progetto è quello di coordinare la pubblicazione delle informazioni relative ai libri all'interno del mondo dei servizi "Wiki" (Wikimedia e Wikisource in particolare) attraverso la definizione di un insieme di proprietà.

La pagina di presentazione del progetto fa esplicito riferimento al modello FRBR e ai quattro livelli<sup>14</sup> *work*, *expression*, *manifestation* e *item*. Si precisa tuttavia che per le esigenze di Wikidata si ritiene sufficiente organizzare le informazioni sui libri attorno a due "insiemi" (*sets*): l'*opera* e l'*edizione*. L'insieme *edizione* raggruppa il livello *espressione* e il livello *manifestazione*.<sup>15</sup> Viene citato anche il

<sup>8</sup> <https://schema.org>.

<sup>9</sup> Il corsivo è stato aggiunto per migliorare la leggibilità.

<sup>10</sup> Le ragioni della scelta delle modalità tecniche di reificazione si possono trovare qui (Erxleben et al. 2014). Per un confronto sulle tecniche LOD esistenti per la reificazione: (Hernández, Hogan e Krötzsch 2015).

<sup>11</sup> <https://www.wikidata.org/wiki/Q38>

<sup>12</sup> La tripla "reificata" è dereferenziabile con un suo indirizzo URI <http://www.wikidata.org/entity/statement/q38-152baf03-ec3f-4261-ba39-f74259153118>.

<sup>13</sup> La tripla "reificata" è dereferenziabile a questo indirizzo <http://www.wikidata.org/entity/statement/Q38-90253fce-46a2-056e-acef-edbae14a0128>.

<sup>14</sup> "the different aspects of user interests in the products of intellectual or artistic endeavour" (IFLA 1998).

<sup>15</sup> «Not to complicate too much, we didn't use the FRBR terms "expression" or "manifestation", as the boundary between



modello BIBFRAME che allo stesso modo prevederebbe un raggruppamento di manifestazione e espressione. In realtà in BIBFRAME la classe *bf:Work* raggruppa sia *frbr:work* che *frbr:expression*. BIBFRAME prevede la proprietà *hasExpression*<sup>16</sup> per “compatibilità” con il modello FRBR/RDA/LRM. Sono quindi possibili in BIBFRAME triple del tipo *<bf:Work-1> <bf:hasExpression> <bf:Work-2>* dove in *<bf:Work-2>* può trovare posto una determinata *frbr:expression*.<sup>17</sup>

FRBR/RDA/LRM	BIBFRAME	Wikidata
<i>frbr:work</i>	<i>bf:work</i> <sup>18</sup>	<i>wd:work</i>
<i>frbr:expression</i>		<i>wd:edition</i>
<i>frbr:manifestation</i>	<i>bf:instance</i>	

Figura 3: FRBR, BIBFRAME e Wikidata: “livelli” opera, manifestazione e espressione

In Figura 3 si può trovare un confronto tra i tre “livelli” proposti da FRBR (e definiti da RDA e LRM) e l’implementazione in BIBFRAME e in Wikidata. Come si può vedere il “livello intermedio” *frbr:expression* è trattato come *bf:work* in BIBFRAME e come *wd:edition* in Wikidata.

### 3. L’entità IFLA LRM *Nomen* e la sua applicazione

Una delle novità più rilevanti del modello IFLA LRM<sup>19</sup> è l’entità *nomen* che almeno a prima vista non sembrerebbe trovare posto nel *Wikibase data model*. In realtà – come vedremo – il *Wikibase data model* potrebbe offrire sia elementi utili alla discussione in corso sul ruolo di questa entità, sia anche per ipotizzare applicazioni che possano essere usabili e diffuse.

Per IFLA LRM «un’entità è una classe astratta di oggetti concettuali; ci sono molte istanze di ogni entità descritte nei dati bibliografici» (IFLA 2020, 18). L’uso del termine “classe” per definire “entità” fa riferimento implicito alla definizione RDF di classe: «Resources may be divided into groups called classes. The members of a class are known as instances of the class» (RDF Schema 1.1 2014). Come viene precisato più avanti il «modello è sviluppato avendo ben presenti le tecnologie del web semantico, nell’auspicio che in futuro una versione aggiornata di questo documento fornisca anche esempi RDF» (IFLA 2020, 40).

In generale tutte le entità di una ontologia (e tutte le istanze di quelle entità) devono avere (almeno) un *nome* o *etichetta* per poter essere oggetto di riferimento attraverso un determinato linguaggio. Il vocabolario di base di RDF (ovvero RDFS) definisce il *nome* come una proprietà: *rdfs:label*

the definitions is not easy to grasp” [...] “(other conceptual frameworks similar to FRBR, like Bibframe, also collapse those two layers)»: [https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Books](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Books). La versione consultata è del 29 marzo 2022, ore 21:49 disponibile su <https://tinyurl.com/2p87mbwb>.

<sup>16</sup> «Work that is an expression of a described Work. Use to relate Works under LRM/RDA guidelines or similar implementations»: questo è l’indirizzo URI della proprietà: [https://id.loc.gov/ontologies/bibframe.html#p\\_hasExpression](https://id.loc.gov/ontologies/bibframe.html#p_hasExpression).

<sup>17</sup> Gli esempi che fanno uso di triple racchiudono di solito tra parentesi uncinata i tre elementi della tripla. Ogni elemento della tripla può avere un “namespace” seguito da due punti.

<sup>18</sup> «The BIBFRAME Work class, *bf:Work*, which is the conceptual view of a resource, relates to the FRBR/RDA Work and Expression entities»: (McCallum 2017).

<sup>19</sup> Per una introduzione al modello concettuale IFLA LRM: (Guerrini e Sardo 2018).

è la proprietà utilizzata per fornire una versione del *nome* della risorsa che sia leggibile soprattutto da un essere umano. Non sempre questo meccanismo di base è sufficiente a dare conto della complessità dei dati trattati da una determinata ontologia. Il modello IFLA LRM non fa affidamento a questo meccanismo di base. Le ragioni di questa scelta sono spiegate nel capitolo 4 (Definizione del modello).

Anziché *etichettare direttamente* tutte le entità definite da IFLA LRM, la scelta è quella di passare attraverso un'altra *entità* che viene chiamata “nomen”: «[...] l'entità *nomen* può essere intesa come la reificazione di una relazione tra un'istanza di *res*<sup>20</sup> e una stringa. La stringa stessa non costituisce un'istanza dell'entità *nomen*, ma è modellata come il valore dell'attributo *stringa del nomen* di un'istanza dell'entità *nome*» (IFLA 2020, 31).

In pratica per *etichettare* le entità anziché usare la modalità di base (esempi 2, 3, 4 di Figura 2) il modello IFLA LRM propone per ogni entità la coppia di triple esemplificata in Figura 4.

Soggetto	Predicato	Oggetto
1) <lrn:work-1>	<lrn:ha come appellativo-LRM-R13>	<lrn:nomen-1>
2) <lrn:nomen-1>	<lrn:stringa del nomen-LRM-E9-A2>	<"Modellare la conoscenza, standard archivistici e bibliografici a confronto">

Figura 4: Esempio di collegamento tra nomen e le altre entità del modello.<sup>21</sup>

A differenza della modalità di base *rdfs*, la proposta LRM ha il vantaggio che lo “snodo” *nomen* – «una reificazione della relazione tra un'istanza di *res* e una stringa» – consente di precisare tutta una serie di proprietà. Una entità *oggetto* di una tripla può diventare *soggetto* di altre triple, mentre una stringa letterale può solo fare da oggetto. LRM prevede infatti che l'entità *nomen* possa fare da soggetto per altri attributi (oltre all'attributo *stringa del nomen* necessario affinché l'entità *nomen* possa formarsi) quali ad esempio *categoria*, *schema* e *lingua*.<sup>22</sup> Inoltre l'entità *nomen* può avere relazioni con le altre entità a partire da quella essenziale alla sua ragione d'essere: ovvero essere *oggetto* di una tripla che ha come *predicato* <*ha come appellativo-LRM-R13*> e come *soggetto* è una delle entità del modello LRM.

La proposta del *nomen* LRM presenta anche qualche svantaggio. Il primo è l'astrattezza dell'entità (non è di immediata comprensione) e il secondo la necessità di “moltiplicare” le istanze delle entità LRM visto che ogni entità dovrebbe avere almeno un *nomen* e che ad esempio tutte le forme varianti dei punti di accesso controllati dovrebbero avere ciascuna un proprio *nomen*.<sup>23</sup>

<sup>20</sup> *Res* è la superclasse di tutte le entità IFLA LRM. Tutto quello che viene definito per *Res* viene ereditato dalle sottoclassi (Bianchini 2017).

<sup>21</sup> <work-1> sta per una determinata istanza dell'entità *work*, così come <nomen-1> sta per una determinata istanza dell'entità *nomen*.

<sup>22</sup> L'elenco completo in (IFLA 2020: 53-58).

<sup>23</sup> L'entità *nomen* può dar luogo ad equivoci: «Un'innovazione particolarmente infelice e pasticciata del modello LRM è l'introduzione dell'entità *Nomen*, definita come “Un'associazione tra un'entità e una designazione che le si riferisce”[...]. Mentre quello che LRM chiama “stringa del nomen”, ossia un'espressione o sequenza di segni, è o può essere considerata un'entità (mentre viene invece considerata un attributo), un'associazione (nel senso della definizione citata) è evidentemente una relazione, e non può essere perciò un'entità» (Petrucciani 2022, 171 n.17).

La difficoltà di applicare la proposta LRM del *nomen* si può vedere anche in questo interessante contributo che si propone di presentare un prototipo per una interfaccia di catalogazione facilmente e immediatamente accessibile che si basa sui vantaggi offerti dal modello IFLA LRM (Pau-man Budanović e Žumer 2021). La decisione iniziale di limitare il prototipo solo ad alcune entità del modello (escludendo anche il *nomen*) lascia perplessi: «Of the eleven LRM entities: *work, expression, manifestation, item, agent, person, collective agent, Res, nomen, place, and time span*, the cornerstone are entities *work, expression, manifestation, and item* [...], their attributes and relationships, to which we focused in the design of our prototype cataloging interface». Nell'applicazione del modello IFLA LRM tuttavia il *nomen* non è decisione che viene lasciata all'utilizzatore. La Figura 5 presenta l'alternativa tra *nomen* come entità (la scelta di IFLA LRM) e *nomen* come attributo: «The first model (the one adopted in LRM) shows that a single *res* may be related to two distinct instances of a *nomen* entity by appellation relationships, and all the entities have attribute values. The lower model shows the alternative of treating *nomens* as attributes of the *res* entity. In this case, values of the “name” attribute cannot have attributes in turn, and no relationships can be declared between these terms and any other entities in the model». <sup>24</sup>

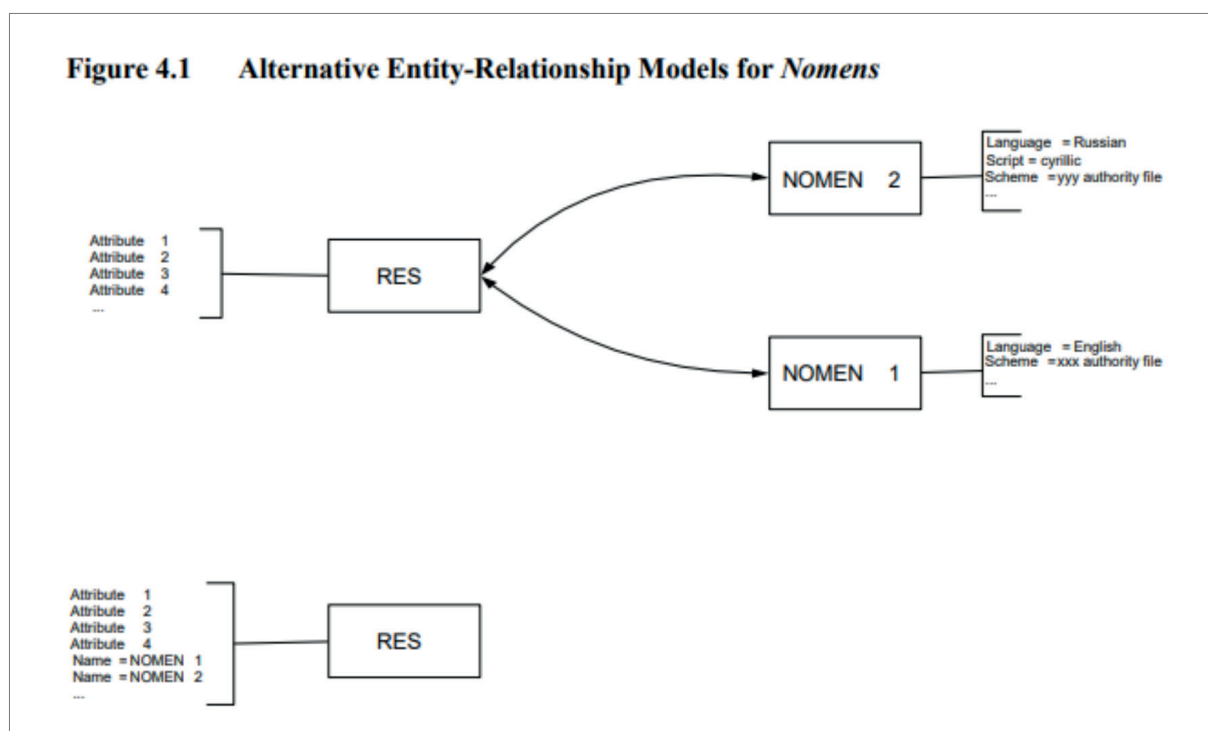


Figura 5: Modelli alternativi per “dare un nome” alle entità

<sup>24</sup> La Figura 5 riporta – per comodità di lettura – la Figura 4.1 collegata al testo citato da (IFLA 2017, 15).

In altre parole la scelta LRM per “dare un nome” alle istanze delle entità è quella di passare sempre attraverso l’entità *nomen*: senza l’entità *nomen* non si possono trattare le altre entità come “*work, expression, manifestation, and item*”.

## 4. MARC e la scelta RDF

In questo paragrafo viene presentata un’esperienza (ancora in corso) di implementazione del modello LRM. L’implementazione si pone tra gli obiettivi di tenere conto sia della tradizione del formato MARC, sia della necessità di tener conto del *web semantico*.

Nel 2002 un famoso articolo di Roy Tennant con un titolo provocatorio – *MARC must die* – portava l’attenzione sul futuro del MARC.<sup>25</sup> Come lo stesso Tennant ebbe modo di precisare successivamente, l’obiettivo dell’articolo era quello di mettere in guardia la comunità professionale sui limiti tecnologici del MARC, non quello di decretarne la sua fine.<sup>26</sup> In ogni caso nel 2018 al Congresso IFLA di Kuala Lumpur un intervento di due bibliotecari della BnF (Bibliothèque nationale de France) dal titolo *Still waiting for that funeral* prendeva in conto il *MARC must die* e si proponeva di fare il punto sulla questione del futuro del MARC all’interno del programma di “transizione bibliografica” in corso alla nazionale francese (Peyrard e Roche 2018). In estrema sintesi la scelta della BnF è quella di rinnovare profondamente il MARC prendendo in conto RDA e il modello IFLA LRM. Il punto di arrivo è INTERMARC-NG ovvero un MARC basato su UNIMARC (non su MARC21).<sup>27</sup> Viene messo in evidenza come INTERMARC-NG sarà un formato interno e che comunque i *linked data* che verranno pubblicati saranno del tutto conformi agli standard del web semantico. Si tratta di un intervento in corso (BnF 2021a) che sarà completato entro il 2023, ma è interessante notare che l’applicativo che la BnF sta sviluppando per la gestione interna dei dati bibliografici è basato sulla piattaforma Wikibase.<sup>28</sup>

A conferma del fatto che l’entità *nomen* non ha percorsi lineari di implementazione nelle applicazioni l’intervento (Peyrard e Roche 2018) precisa: «We nevertheless decided that the *LRM E9-Nomen* and *LRM-E11Time Span* entities would not be treated as distinct records, but as attributes of the record.<sup>29</sup> The relationship between an entity and a Nomen or Time-Span will be expressed directly in the entities that they describe, with the tag or subtag that contains

<sup>25</sup> (Tennant 2002).

<sup>26</sup> «I wanted librarianship to wake up to the fact that our foundational standard was no longer serving us like it should» (Tennant 2017).

<sup>27</sup> INTERMARC-NG sarà un formato interno (NG sta per *nouvelle génération*): «We believe that keeping MARC as a production format allows for professional continuity and enhances the expertise of bibliographers, while exporting it as Linked Data acknowledges and meets the needs of a community of mainly non bibliographers users» (Peyrard e Roche 2018).

<sup>28</sup> «Les enjeux sont multiples: techniques, en développant une nouvelle application full-web fondée sur le socle logiciel libre Wikibase; normatifs, en inscrivant dans le format l’ensemble des modifications nécessaires au passage au modèle entité-relations porté par IFLA LRM» (BnF 2021b). L’acronimo NOEMI sta per «Nouer les Œuvres, Expressions, Manifestations et Items». Non sono ancora pubbliche in questo momento informazioni di dettaglio sulle modalità di implementazione di NOEMI in Wikibase.

<sup>29</sup> «[...] les attributs de Nomen sont gérés comme des méta-métadonnées désignées par des codes de sous-zones utilisant des majuscules» come ad esempio «LRM-E9-A3 Schéma d’encodage: \$N Norme utilisée pour construire le point d’accès; \$X Règles d’indexation»: corrispondenza del 3 marzo 2022 con Françoise Leresche (Bibliothèque nationale de France, Département des Métadonnées, Chargée de Normalisation documentaire).

them expressing how they are used. This has the advantage of staying as close as possible to the existing structure and avoiding entity cluttering». In altre parole il formato interno INTERMARC-NG ospiterà tutte le informazioni previste dal modello IFLA LRM relativamente al *nomen* (inclusi attributi e relazioni del *nomen*) in campi e sottocampi MARC senza creare istanze separate per l'entità *nomen*. Come già ricordato questo non pregiudicherà la pubblicazione dei dati come *linked data*. In altre parole la scelta della BnF è quella di rendere – come si dice in gergo – trasparente per l'utilizzatore la complessità della gestione del *nomen* senza rinunciare alle possibilità offerte dal modello LRM.<sup>30</sup>

## 5. Il *nomen* nel Wikibase data model

All'interno di una sperimentazione dell'utilizzo di Wikibase nel mondo MAB, il tentativo di ospitare le entità IFLA LRM all'interno del *Wikibase data model* ha dovuto fare i conti anche con l'entità *nomen*.<sup>31</sup> Il primo percorso sperimentato (un *wikibase:item* per ogni *nomen*) si è rivelato problematico proprio per quella “proliferazione delle entità” evidenziata nell'esperienza di NOEMI. Il secondo percorso ha cercato invece di sfruttare uno degli aspetti peculiari del *Wikibase data model*: il qualificatore (qualifier). Possiamo infatti trovare una forte vicinanza concettuale tra la definizione dell'entità *nomen* LRM come «reificazione di una relazione tra un'istanza di *res* e una stringa» e il concetto di reificazione di una tripla RDF. In ambedue i casi la reificazione è necessaria per poter esprimere gli attributi che hanno come soggetto la relazione reificata. In Figura 6 è possibile vedere uno dei risultati della sperimentazione. Si è scelto di non usare l'entità LRM *nomen* ma di collegare direttamente l'attributo *Stringa del nomen*(LRM-E9-A2) all'entità Q56. In questo modo si possono applicare – come qualificatori (qualifier) Wikibase – gli attributi previsti per l'entità *nomen* per ogni “reificazione della relazione” tra l'entità Q56 – in questo caso una specifica istanza di un *lrm:agente*) – e le stringhe “Bodoni, Giambattista” e “Alcippo Persejo”: a titolo di esempio sono stati usati gli attributi *Categoria* (LRM-E9-A1) e *Schema* (LRM-E9-A3).

---

<sup>30</sup> «Dans l'ontologie, l'entité Nomen existe pour porter les éléments de méta-métadonnées. Elle existe pour les machines, mais est transparente pour le catalogueur qui ne connaît que les implémentations de l'entité». Corrispondenza del 3 marzo 2022 con Françoise Leresche (Bibliothèque nationale de France, Département des Métadonnées, Chargée de Normalisation documentaire).

<sup>31</sup> Si tratta del Laboratorio 4 nato all'interno di una iniziativa AIB, Rete delle reti, EBLIDA (<https://www.aib.it/struttura/osservatorio-biblioteche-e-sviluppo-sostenibile/2021/94441-10-laboratori-virtuali-per-biblioteche-e-sviluppo-sostenibile/>). Il Laboratorio 4 è stato coordinato da Stefano Bargioni, Giovanni Bergamin e Valentina Lepore. Occorre ringraziare tutti i componenti del Laboratorio (Angelica Bertellini, Anna Molino, Cinzia Martone, Denise Biagiotti, Federica Olivotto, Marina Gatti, Stefania Puccini). Per l'implementazione del *nomen* con il *Wikibase data model* sono stati preziosi i suggerimenti di Carlo Bianchini e Lucia Sardo. Per un resoconto del Laboratorio 4: <https://doi.org/10.5281/zenodo.6816705>.

## Giambattista Bodoni (Q56)

Stringa del nomen	 Bodoni, Giambattista  modifica
	Categoria punto-di-accesso
	Schema REICAT
	▼ 1 riferimento
	ID SBN autore RAVV042068
	+ aggiungi riferimento
	 Alcippo Persejo  modifica
	Categoria variante-punto-di-accesso
	Schema REICAT
	▼ 1 riferimento
ID SBN autore RAVV042068	
+ aggiungi riferimento	
+ aggiungi valore	

Figura 6: Sperimentazione del trattamento del *nomen* in Wikibase usando come qualificatore l'attributo *stringa del nomen* <https://smartmeta.wikibase.cloud/wiki/Item:Q56>

Possiamo vedere la versione pubblicata dei *linked data* relativi all'entità Q56 (con la serializzazione Turtle) a questo indirizzo: <https://smartmeta.wikibase.cloud/wiki/Special:EntityData/Q56.ttl>. Nella Figura 7 sono state messe in evidenza alcune triple che interessano la sperimentazione. In pratica oltre a alla tripla  $\langle wd:Q56 \rangle \langle wdt:P19(LRM-E9-A2) \rangle \langle "Bodoni, Giambattista" \rangle$ , troviamo una nuova entità *wikibase:statement* che viene creata dalla piattaforma (Q56-[identificativo]).<sup>32</sup> Questa nuova entità veicola la relazione reificata tra una istanza dell'entità (wd:Q56) e la stringa ("Bodoni, Giambattista") o se vogliamo è semanticamente coincidente con l'entità LRM *nomen*. Quest'ultima entità (statement) può diventare il soggetto di *triple* in grado di ospitare tutti gli attributi che LRM prevede per il nomen. È interessante notare i collegamenti che collegano il *wikibase:statement* alla tripla di partenza: il *wikibase:statement* è oggetto della tripla che ha come soggetto l'entità Q56 e come predicato la proprietà P19 (differenziata con un namespace specifico: *ps* ovvero *property statement*); come soggetto di una tripla che ha come oggetto la stringa "Bodoni, Giambattista" e come predicato di nuovo la proprietà P19 (anche questa differenziata con il namespace *pv* ovvero *property value*). Vale la pena ricordare che in Wikibase l'entità *wikibase:statement* viene creata automaticamente per ogni dichiarazione (anche quando non vengono aggiunti i qualificatori) che ha come soggetto un *wikibase:item* (es Q56).

<sup>32</sup> Con la serializzazione Turtle: *s:Q56-a64a9b0b-404d-d598-ebb2-e805928e5b44 a wikibase:Statement* .

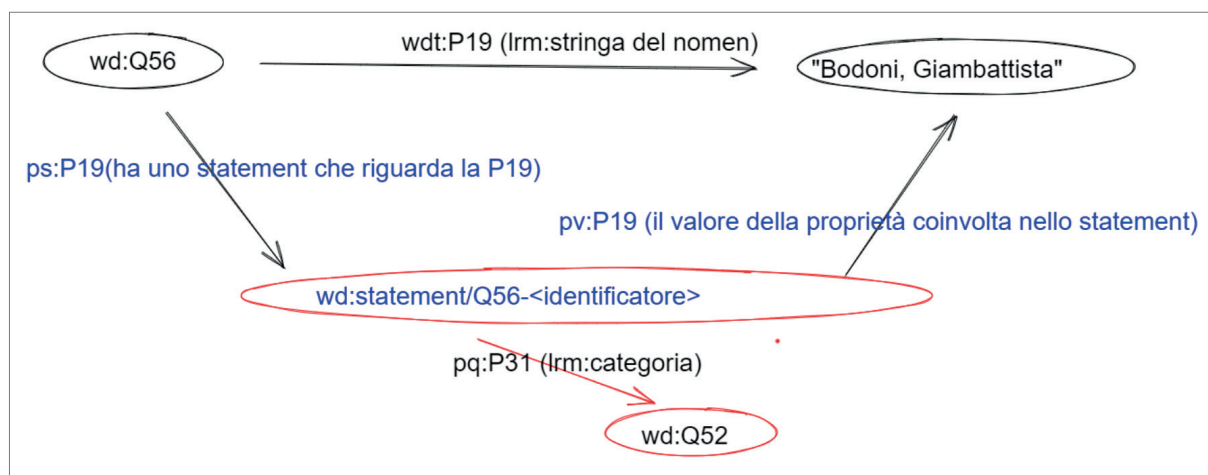


Figura 7: Sperimentazione IFLA LRM *nomen*: in blu l'entità *statement* e il collegamento alla tripla di partenza; in rosso esempio di tripla che supporta gli attributi che LRM prevede per l'entità *nomen* (nel *Wikibase data model* equivalente a quel determinato *statement*)

Indubbiamente da una sperimentazione minimale di questo tipo non possono essere tratte conclusioni di tipo generale. La strada tuttavia sembra essere promettente e sicuramente vale la pena investire in sperimentazioni più estese. Tra i casi da sperimentare accuratamente ci sono le relazioni che LRM definisce per l'entità *nomen*. L'entità *wikibase:statement* in quanto semanticamente coincidente con l'entità LRM *nomen* può fare da soggetto a triple che implementano anche – oltre a attributi – anche relazioni definite da LRM come applicabili all'entità *nomen*. Dato che l'entità LRM *nomen* si può trovare sia come soggetto che come oggetto di una tripla, una entità *wikibase:statement* – in linea di principio, in quanto entità dereferenziabile attraverso un indirizzo di tipo URI – potrebbe svolgere il ruolo anche di oggetto in una tripla.<sup>33</sup> In ogni caso una sperimentazione estesa dovrebbe prendere in conto anche le situazioni più complesse a partire da quelle che LRM contempla. Attraverso l'uso di strumenti tecnologici di larga diffusione (come Wikibase) dovrebbero essere cercate soluzioni che mettono insieme: le tipologie di servizi che gli utenti oggi si aspettano, l'ergonomia nella gestione dei metadati da parte del bibliotecario e il contributo che attraverso i linked open data i metadati possono dare a tutto l'ecosistema bibliografico.

## 6. Conclusioni

In generale il lavoro delle biblioteche nella “organizzazione dell'informazione” non sembra ancora essere adeguatamente valorizzato nell'ecosistema del web. Il riferimento è naturalmente a un web conforme alla visione del web semantico e alla proposta tecnologica dei *linked data*. La strada di prendere in conto le iniziative dei *big player* può offrire risultati importanti. È interessante prendere atto che i *big player* del web semantico in questo momento non siano identificabili con il

<sup>33</sup> Alle note 12 e 13 si possono trovare due indirizzi di tipo URI relativi a due *wikibase:statement*.

ben noto acronimo GAFAM<sup>34</sup>. Le due iniziative più rilevanti (anche in termini di diffusione) che stanno portando il web semantico nella vita quotidiana di chi cerca informazioni strutturate sono *schema.org* e *Wikidata*.<sup>35</sup>

Questo intervento ha preso in conto i lavori in corso sui modelli bibliografici. Senza nessuna pretesa di completezza sono stati esaminati alcuni problemi di applicabilità tecnologica dei modelli bibliografici (in particolare di IFLA LRM). Il punto di vista scelto (il *Wikibase data model*) si è rivelato molto utile per ipotizzare soluzioni applicabili con strumenti tecnologici di larga diffusione. Grazie al lavoro (ancora in corso) della BnF con NOEMI abbiamo visto inoltre che il *Wikibase data model* ci può offrire elementi anche per l'evoluzione del formato MARC.<sup>36</sup>

Il *Wikibase data model* non è certamente l'unicorno magico che risolve tutti i problemi, ma sicuramente offre strumenti molto sofisticati con una bassa barriera d'ingresso e può essere un'ottima fonte di ispirazione per chi progetta modelli e per chi costruisce applicazioni.

---

<sup>34</sup> <https://it.wikipedia.org/wiki/GAFAM> (Google, Apple, Facebook, Amazon, Microsoft).

<sup>35</sup> Le statistiche (2021) di utilizzo di triple *schema.org* codificate nelle pagine web disponibili in <http://webdatacommons.org/>: 82 miliardi di triple per 1,5 miliardi di pagine web su un totale di 3,2 miliardi di pagine esaminate. Per le statistiche su Wikidata: <https://www.wikidata.org/wiki/Special:Statistics>. Per *schema.org* e ricerca dell'informazione bibliografica si può vedere (Wallis 2022).

<sup>36</sup> Per UNIMARC e il Wikibase data model si può vedere (Bergamin e Bacchi 2018).



## Riferimenti bibliografici

- Bergamin, Giovanni e Bacchi, Cristian. 2018. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data." *JLIS.It* 9,3: 35–74. <https://doi.org/10.4403/jlis.it-12458>.
- Bianchini, Carlo. 2017. "Remarks about IFLA Library Reference Model". *JLIS.it* 8, 3: 86–99. <https://doi.org/10.4403/jlis.it-12416>.
- Bianchini, Carlo e Lucia Sardo. 2022. "Wikidata: a new perspective towards universal bibliographic control." *JLIS.it* 13, 1: 291–311. <https://doi.org/10.4403/jlis.it-12725>.
- BnF. 2021a. "Découvrir l'entité expression: son intérêt pour les utilisateurs." <https://youtu.be/mcZXwngW9xw>. Registrazione video dell'evento «6e journée professionnelle organisée par le groupe Systèmes & Données du programme Transition bibliographique».
- BnF. 2021b. "Focus sur NOEMI, le futur outil de production des métadonnées de la BnF." *Actualités du catalogue* 51, mars 2021. <https://www.bnf.fr/fr/actualites-du-catalogue-no-51-mars-2021#bnf-focus-sur-noemi-le-futur-outil-de-production-des-m-tadonn-es-de-la-bnf>.
- Erxleben, Fredo, Michael Günther, Markus Krötzsch, Julian Mendez, and Denny Vrandečić. 2014. "Introducing Wikidata to the Linked Data Web." In *Proceedings of the 13th International Semantic Web Conference*, 50–65. Springer. [https://doi.org/10.1007/978-3-319-11964-9\\_4](https://doi.org/10.1007/978-3-319-11964-9_4).
- Guerrini, Mauro e Lucia Sardo. 2018. *IFLA Library Reference Model (LRM): un modello concettuale per le biblioteche del XXI secolo*. Milano: Editrice Bibliografica.
- Hernández, Daniel, Aidan Hogan, e Markus Krötzsch. 2015. "Reifying RDF: What works well with wikidata?" In *Proceedings of the Conference on Scalable Semantic Web Knowledge Base Systems (SWSS'15)*. 32–47. [http://ceur-ws.org/Vol-1457/SSWS2015\\_paper3.pdf](http://ceur-ws.org/Vol-1457/SSWS2015_paper3.pdf).
- IFLA Study Group on the Functional Requirements for Bibliographic Records. 1998. *Functional Requirements for Bibliographic Records (FRBR): Final Report*. Munchen: K.G. Saur. <https://archive.ifla.org/VII/s13/frbr/frbr.pdf>.
- IFLA (International Federation of Library Associations and Institutions). 2017. *IFLA Library Reference Model. A Conceptual Model for Bibliographic Information*. Edited by Pat Riva, Patrick Le Boeuf, and Maja Zumer. Den Haag: IFLA. [https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla\\_lrm\\_2017-03.pdf](https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla_lrm_2017-03.pdf).
- IFLA (International Federation of Library Associations and Institutions). 2020. *Definizione di un modello concettuale di riferimento per fornire una struttura per l'analisi dei metadati non amministrativi relativi alle risorse di biblioteca*. A cura di Pat Riva, Patrick Le Boeuf, Maja Žumer; edizione italiana a cura dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche. Roma:ICCU. [https://www.iccu.sbn.it/export/sites/iccu/documenti/2020/IFLA\\_LRM\\_ITA.pdf](https://www.iccu.sbn.it/export/sites/iccu/documenti/2020/IFLA_LRM_ITA.pdf).
- McCallum, Sally.H. 2017. "BIBFRAME Development." *JLIS.it* 8, 3: 71-85. <https://doi.org/10.4403/jlis.it-12415>.

Pauman Budanović, Mihela e Maja Žumer. 2021. “Prototype Cataloging Interface Based on the IFLA Library Reference Model (LRM). Part 1: Conceptual Design.” *Cataloging & Classification Quarterly* 49,7: 619-643. <https://doi.org/10.1080/01639374.2021.1974633>.

Petruciani, Alberto. 2022. “I modelli bibliografici (da FRBR a LRM): un edificio da ricostruire dalle fondamenta?” In *Bibliografia e cultura: studi per Alfredo Serrai* a cura di Enrico Pio Ardolino e Diego Baldi, 155–173. Roma: C.N.R., Istituto di scienze del patrimonio culturale.

Peyrard, Sébastien e Roche, Mélanie. 2018. “Still Waiting for that funeral: the challenges and promises of a next-gen INTERMARC.” In *The IFLA Library* <http://library.ifla.org/id/eprint/2204/>.  
RDF Schema 1.1. 2014. <https://www.w3.org/TR/rdf-schema/>.

Tennant, Roy. 2002. “MARC must die.” *Library journal*, Oct 15. <http://www.libraryjournal.com/article/CA250046.html>. Versione disponibile anche su: <http://soiscompsfall2007.pbworks.com/f/marc%20must%20die.pdf>.

Tennant, Roy. 2017. “‘MARC Must Die’ 15 years on.” *Hanging Together*, October 15. <https://hangingtogether.org/marc-must-die-15-years-on/>.

Wallis, Richard. 2022. “Follow me to the library! Bibliographic data in a discovery driven world.” *JLIS.It* 13,1: 37–44. <https://doi.org/10.4403/jlis.it-12762>.

Wikibase/DataModel. 2020. <https://www.mediawiki.org/wiki/Wikibase/DataModel>.

## The entities of the IFLA-LRM, RiC-CM and CIDOC-CRM models in the semantic web

Carlo Bianchini<sup>(a)</sup>

a) Università di Pavia, Dipartimento di musicologia e beni culturali, <https://orcid.org/0000-0002-6635-6371>

**Contact:** Carlo Bianchini, [carlo.bianchini@unipv.it](mailto:carlo.bianchini@unipv.it)

**Received:** 28 May 2022; **Accepted:** 6 June 2022; **First Published:** 15 September 2022

### ABSTRACT

The traditional approach to cataloguing has always been based on the investigation of the characteristics of the objects in collections that are most relevant to the construction of catalogues. The evolution of cataloguing theory has therefore led to the identification and analysis of entities with those characteristics and to the construction of a logical model capable of explaining bibliographic phenomena in an increasingly refined manner. That approach also led to the definition of RiC-CM and CIDOC-CRM models. In addition to this valid point of view, a second perspective is proposed, which takes into account the entities identified in the logical models developed in the library, archive and museum models as part of the much larger, richer and more numerous ontologies of the semantic web, represented by the Linked Open Data Cloud. In this perspective, the logical models of libraries, archives, and museums can be seen as some of the possible infinite modeling of web entities, constructed in the light of the principles and tradition of each subject area. This new perspective helps to better clarify the role of data professionals, the concept of metadata, the characteristics of logical models and to take a unified view of the bibliographic, archival and museum universes.

### KEYWORDS

IFLA LRM; RiC-CM; CIDOC-CRM; Metadata; Semantic Web; Identification; Entity.

## Le entità dei modelli IFLA-LRM, RiC-CM e CIDOC-CRM nel web semantico

### ABSTRACT

L'approccio tradizionale della catalogazione si è sempre basato sull'indagine delle caratteristiche degli oggetti delle collezioni più rilevanti ai fini della costruzione dei cataloghi. L'evoluzione della teoria catalografica ha quindi portato a individuare ed analizzare le entità che possiedono quelle caratteristiche e a costruire un modello logico in grado di spiegare i fenomeni bibliografici in modo sempre più raffinato. Questo approccio è quello che ha portato anche alla definizione di RiC-CM e CIDOC-CRM. A questo valido punto di vista, si propone di affiancare una seconda prospettiva, che vede le entità individuate nei modelli logici sviluppati in ambito bibliotecario, archivistico e museale come entità facenti parte delle ontologie molto più vaste, ricche e numerose del web semantico, rappresentate dalla Linked Open Data Cloud. In questa prospettiva, i modelli logici di musei, archivi e biblioteche possono essere visti come alcune tra le possibili infinite modellazioni delle entità del web, costruite alla luce dei principi e della tradizione di ciascun ambito disciplinare. Questa nuova prospettiva aiuta a chiarire meglio il ruolo dei professionisti dei dati, il concetto di metadato, le caratteristiche dei modelli logici e ad assumere una visione unitaria dell'universo bibliografico, archivistico e museale.

### PAROLE CHIAVE

IFLA LRM; RiC-CM; CIDOC-CRM; Metadata; Semantic Web; Identification; Entity.

L'autore ringrazia Mauro Guerrini, Laura Manzoni e Alessandra Moi per i loro suggerimenti per migliorare le bozze del testo.

© 2022, The Author(s). This is an open access article, free of all copyright, that anyone can freely read, download, copy, distribute, print, search, or link to the full texts or use them for any other lawful purpose. This article is made available under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. JLIS.it is a journal of the SAGAS Department, University of Florence, Italy, published by EUM, Edizioni Università di Macerata, Italy, and FUP, Firenze University Press, Italy.

L'obiettivo di questo contributo è presentare una riflessione sul rapporto che sussiste tra le entità dei modelli logici delle discipline biblioteconomiche, archivistiche e museali (ma con un'attenzione dedicata soprattutto al modello IFLA LRM) e le corrispondenti entità in altre ontologie del web semantico, allo scopo di comprendere come e perché questo rapporto è diventato più stretto, di capire come l'analisi di questo rapporto può aiutare a comprendere meglio lo sviluppo dei nostri modelli logici e ad aprire nuove prospettive di lavoro comune tra le discipline biblioteconomiche, archivistiche e museali.

Per lungo tempo le biblioteche hanno creato i propri strumenti di indicizzazione basandosi sulle caratteristiche specifiche degli oggetti delle raccolte, come l'oggetto fisico, le sue caratteristiche materiali e le sue modalità di produzione e di circolazione. In questa prospettiva si possono inquadrare, per esempio, i cataloghi inventario delle biblioteche monastiche, nei quali l'incipit di un supporto manoscritto – che poteva contenere più testi e più opere – era invece tutto ciò che era necessario per individuare e ottenere l'oggetto fisico conservato. Nella stessa prospettiva di presentazione formale ed esplicita delle caratteristiche degli oggetti di interesse si collocano i cataloghi editoriali, come quello di Aldo Manuzio o di Regiomontano.

La consapevolezza della necessità di identificare con certezza gli autori ecclesiastici a vantaggio dei propri lettori spinge Johann Trithemius a progettare il suo *Liber de scriptoribus ecclesiasticis* come un repertorio non puramente bibliografico, ma bio-bibliografico, nel quale gli oggetti letterari sono descritti anche in relazione al loro autore, che deve essere identificato con precisione. Per questo stesso motivo il repertorio viene dotato oltre che di un ordinamento cronologico, anche di un indice alfabetico per nome dell'autore (Balsamo 1984, 24).

L'approccio basato sull'analisi e la descrizione delle caratteristiche peculiari degli oggetti delle raccolte – e su concetti come pubblicazione e documento, autore e curatore, editore – è a fondamento delle funzioni del catalogo formulate da Charles A. Cutter nel 1876 e rimaste sostanzialmente invariate per moltissimo tempo, tempo fino al dibattito fondamentale sui concetti di *libro*, *edizione* e *opera* con il contributo di Eva Verona, Seymour Lubetzky e Akos Domanovszky (Verona 1959; 1963; Lubetzky 1963; Domanovszky 1975).

È sempre a partire dai dati su queste entità che si è sviluppato il *Rapporto FRBR*, (IFLA Study Group on the Functional Requirements for Bibliographic Records 2009), il quale ha consentito di focalizzare i diversi aspetti dell'oggetto di interesse delle biblioteche, facendo emergere con maggiore chiarezza la distinzione tra il contenuto intellettuale, la forma espressiva, la forma di pubblicazione e la singola copia, ovvero tra le entità opera, espressione, manifestazione e item.

Nel modello FRBR si nota però un rilevante cambiamento di prospettiva. Da Cutter ai *Principi di Parigi* del 1961 gli oggetti di interesse – come le persone che agivano per la creazione o la pubblicazione dei libri – erano considerati in funzione del contesto specifico legato alle raccolte bibliografiche, e quindi erano designati, per esempio, con termini come “pubblicazione” e “edizione”, o “autore” e “curatore” ecc. Dal 1998 in poi, nel *Rapporto FRBR* chi agisce in relazione alla creazione o alla modifica di un testo viene definito “persona” o “ente collettivo” (e nei rapporti successivi anche “famiglia”, ovvero FRAD e FR SAR) (IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) 2009)“DOI”:"10.1080/14649055.2011.10766293", "note": "00000", "title": "Functional Requirements for Authority Data: a Conceptual Model. Final report", "author": [{"literal": "IFLA Working

Group on Functional Requirements and Numbering of Authority Records (FRANAR (IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR) 2010)”note”:”00000”,”publisher”:”K. G. Saur”,”publisher-place”:”München”,”title”:”Functional Requirements for Subject Authority Data (FRSAD. Inoltre, riguardo a queste entità che compaiono nel primo modello logico dell’IFLA, si precisa che “nell’ambito del presente studio le persone [e gli enti collettivi] sono trattate come entità solo nella misura in cui sono coinvolte nella creazione o realizzazione di un’opera o sono soggetti di un’opera” (IFLA Study Group on the Functional Requirements for Bibliographic Records 2009, 25–26). Il cambiamento terminologico segnala un cambio di prospettiva: da *autore* e *curatore* si passa a parlare di creatore,<sup>1</sup> ovvero “*persone o enti coinvolte nella creazione o realizzazione di un’opera*” e si specifica che l’interesse dell’analisi del Rapporto si limita a persone o enti con questa specifica caratteristica. Altrimenti, ne possiamo dedurre, il *Rapporto* non se ne occupa.

Il cambiamento terminologico che si osserva in FRBR pone una questione solo apparentemente secondaria nella creazione di un modello di dati per l’universo bibliografico. Gli “autori” sono solo meri dati bibliografici all’interno di uno strumento bibliografico o sono “persone” reali che ci celano, più o meno esplicitamente, dietro a un dato bibliografico? Le entità del modello esistono solo perché “sono coinvolti nella creazione o realizzazione di un’opera” o esistono a prescindere – in forma di dati e metadati del web semantico, o anche nel mondo reale – e, a certe condizioni soltanto interessano anche il modello FRBR e i bibliotecari?<sup>2</sup>

Se si rimane in un’ottica puramente bibliografica, la domanda sembra puramente retorica, e la risposta irrilevante, perché, come precisa FRBR, queste entità interessano in quanto hanno una responsabilità bibliografica.

La questione acquista maggiore interesse e maggiore rilevanza se osserviamo un altro importante cambiamento terminologico e concettuale avvenuto tra FRBR e IFLA LRM. FRBR prevedeva un terzo gruppo di entità, costituito dalle entità *concetto* (una conoscenza o un’idea astratta), *oggetto* (una cosa materiale), *evento* (un’azione o un avvenimento) e *luogo* (una posizione). (IFLA Study Group on the Functional Requirements for Bibliographic Records 2009, 17). Tuttavia, anche in questo caso, tutte e quattro l’entità condividevano la caratteristica di essere trattate come entità solo se in possesso di uno specifico requisito, ovvero in quanto “sono soggetto di un’opera” (IFLA Study Group on the Functional Requirements for Bibliographic Records 2009, 26–29).

Questa situazione rimane sostanzialmente invariata in FRAD e in FRSAR, mentre un cambiamento decisivo avviene in IFLA LRM. Nel nuovo modello le entità *concetto*, *oggetto* ed *evento* diventano esempi di categoria dell’entità *res* (IFLA 2017, fig. Tab. 4.4; 2020, fig. Tab. 4.4) e l’entità *luogo* è considerata rilevante in sé e viene descritta in modo autonomo, cioè a prescindere che sia “soggetto di un’opera”. Infatti, qualsiasi entità del modello, e in particolare l’entità *res* può avere

<sup>1</sup> Sull’evoluzione del concetto di autore verso quello di agente e creatore, a si veda anche (Sardo 2017, 341–54).

<sup>2</sup> Tiziana Possemato distingue opportunamente tra lo *stato* e i *comportamenti* di una classe. Si possono utilizzare quelle due categorie logiche anche con questo significato: lo stato è l’insieme degli attributi che caratterizzano un’entità, il comportamento sono i metodi o le procedure che si applicano a quell’entità. Una persona può avere un peso, un’altezza, un colore di occhi e di capelli, e può *agire da* autore, curatore, editore, produttore d’archivio, pittore o da amico, da cittadino ecc. Tiziana Possemato, *Entity modeling: tracce di un percorso in evoluzione*, relazione presentata al 1° Seminario Jlis.it “Modellare la conoscenza. Standard archivistici e bibliografici a confronto” tenutosi a Firenze il 19 maggio 2022, <https://www.youtube.com/watch?v=eXu8cuoQJxU>.

una relazione del tipo “è soggetto di” (LRM-R12) con una qualsiasi istanza di un’opera, per esprimere l’idea che qualsiasi cosa del mondo reale può essere soggetto di un’opera.<sup>3</sup>

La scomparsa di entità come concetto, oggetto ed evento e la loro sostituzione funzionale con una o più relazioni sono un segnale di un’evoluzione importante nel modo in cui i dati bibliografici cercano di integrarsi nel web semantico. Inoltre, sono giustificate dall’obiettivo di svolgere una funzione relativamente nuova, richiesta ai sistemi bibliografici per la prima volta da Elaine Svenonius (Svenonius 2000, 20; 2008, 29–30) e poi dai *Principi internazionali di catalogazione* (ICP) come una nuova funzione essenziale dei cataloghi: la navigazione e l’esplorazione. Questa funzione si deve svolgere “all’interno di un catalogo, tramite la disposizione logica dei dati bibliografici e d’autorità e la presentazione chiara delle relazioni tra entità [e] oltre il catalogo, *verso altri cataloghi e in contesti non legati alle biblioteche*” (IFLA Cataloguing Section e IFLA Meeting of Experts on an International Cataloguing Code 2016, 11).

Ciò che determina un punto di svolta cruciale nell’approccio alle entità dei modelli logici e nelle funzioni dei sistemi bibliografici è la necessità di garantire la navigazione verso altri cataloghi e contesti diversi da quelli delle biblioteche, e non solo all’interno del singolo catalogo.

L’interoperabilità tra i cataloghi è sempre stato un obiettivo del mondo delle biblioteche, perché costituisce il presupposto del Controllo bibliografico universale. Ma la navigazione e l’esplorazione verso contesti informativi diversi, a partire dagli archivi e dai musei per arrivare al web semantico, costituiscono un importante aspetto di novità e una caratteristica indispensabile per i sistemi bibliografici e per i loro utilizzatori. Un sistema di ricerca d’informazione che non abiliti a una navigazione e un’esplorazione a tutto tondo oggi è destinato a esaurire rapidamente il proprio interesse e a soddisfare un pubblico sempre più limitato. Non tanto perché senza questa caratteristica il sistema non possa essere valido o ben costruito, come sono sempre stati in passato i nostri cataloghi, quanto piuttosto perché se costruito e progettato in modo isolato rischia di rimanere tale: ormai navigazione ed esplorazione non sono funzioni facoltative.

Questo nuovo approccio ha avuto un impatto diretto sui modelli logici, che si sono modificati di conseguenza. L’obiettivo della navigazione tra i cataloghi e fonti di dati diverse ha reso necessaria l’introduzione della distinzione tra una specifica entità (per esempio, una persona) e il suo nome in un determinato contesto. Che un catalogo sia fatto di nomi è noto, e che i nomi scelti per costituire l’indice di un catalogo siano determinati da un contesto culturale, linguistico, sociale ecc., è altrettanto noto: i cataloghi sono linguaggi (Guerrini 1996).

Per rendere possibile il dialogo tra cataloghi e sistemi di indicizzazione diversi è necessario un sistema di traduzione interlinguistico che consenta, per esempio, di fare riferimento alla stessa ‘cosa’ usando ‘nomi’ diversi, perché propri di lingue catalografiche diverse.

In un modello relativamente semplice come FRBR (figura 1), il nome è un attributo di una Persona e non c’è spazio per fornire attributi specifici del nome (per esempio, la lingua) perché ogni attributo eventualmente inserito si riferisce all’entità – cioè la persona. Se si registra un valore per l’attributo lingua, si esprime infatti la lingua della persona e non di una particolare forma del nome. Perciò, un modello come FRBR ci fornisce uno schema più semplice, ma anche meno potente dal punto di vista espressivo.

---

<sup>3</sup> Il modello poi precisa che la forma che assume il nome di un particolare oggetto del mondo reale, una res, per essere rappresentato come soggetto dipende dalla relazione complementare LRM-R13 (res - ha come denominazione – nomen).

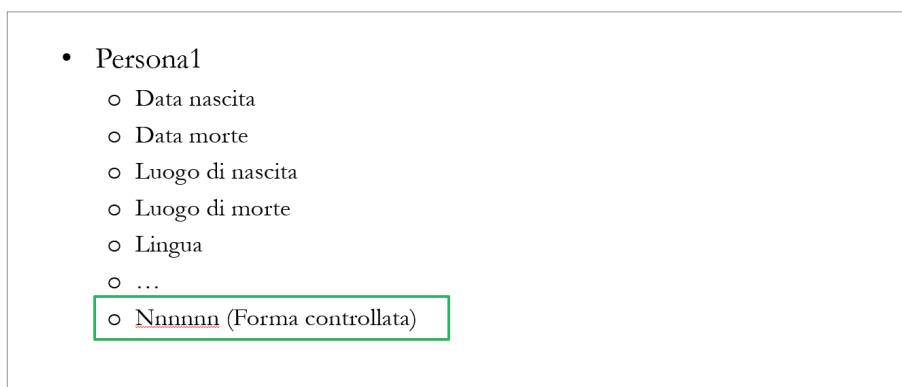


Figura 1 – Il nome come attributo dell'entità persona nel modello FRBR

La scelta di modellizzazione può essere diversa; per esempio nel modello IFLA LRM (figura 2) tra gli attributi della persona non c'è una stringa relativa al nome. Viene invece prevista una relazione «ha come denominazione» tra la persona e l'entità Nomen con la quale viene associata.

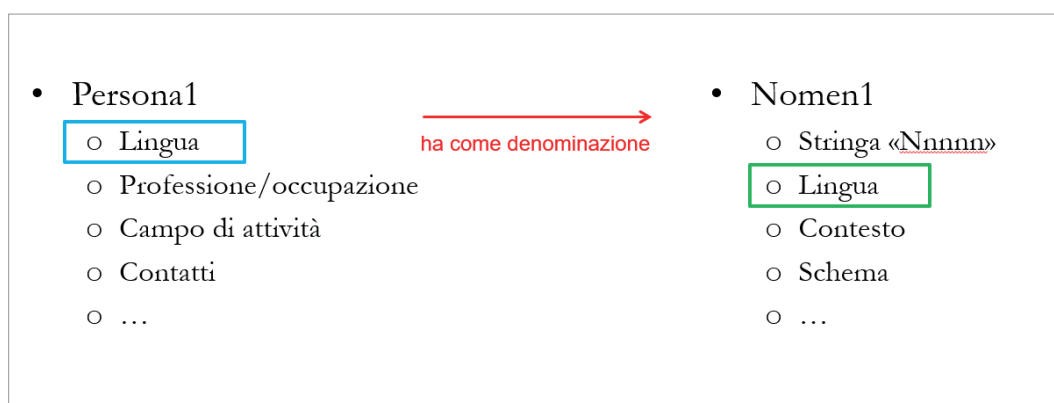


Figura 2 – L'entità Nomen in relazione con l'entità persona nel modello IFLA LRM

Dato che, in questo modo, il Nomen è un'entità, può essere caratterizzato mediante una serie di attributi. Si può quindi registrare l'attributo lingua relativo alla persona (in azzurro) ma anche la lingua in cui può essere espresso il nomen (in verde), che in teoria può essere anche una lingua mai parlata dalla persona associata. La relazione di associazione tra un'entità e il nomen è ripetibile, e quindi è possibile associare e descrivere molti nomen diversi per la stessa entità (per esempio, per una persona). Perciò, se una specifica istanza di una certa entità (per esempio, una particolare persona) è conosciuta con più forme del nome, o è addirittura conosciuta con più nomi diversi, per ciascuno di essi si costituisce un'associazione con l'istanza della specifica persona. Ogni nomen costituito secondo le regole di catalogazione applicate, che sia il punto di accesso controllato e autorizzato o un punto d'accesso controllato e variante, viene associato – mediante la relazione di denominazione (IFLA 2017, fig. 5.3) – a una precisa istanza di un'entità adeguatamente descritta e identificata.

Questo modello è più articolato e complesso – e al momento molto difficile da implementare – ma consente di esprimere molte più caratteristiche del nomen associato a un'entità. Modellare

l'entità nomen ha il vantaggio di potere esprimere nel sistema molte caratteristiche specifiche del nomen stesso; prima di tutto la stringa del nomen, cioè la sequenza alfanumerica che viene utilizzata per indicare la res collegata al nomen, poi la categoria di appartenenza del nomen, il contesto d'uso, i destinatari, la fonte di riferimento, la lingua, lo schema, la scrittura e il sistema di traslitterazione.

Il motivo per cui in IFLA LRM viene introdotta l'entità nomen è che esso è uno strumento indispensabile per l'interoperabilità tra cataloghi, che si riferiscono alla stessa cosa con nomen differenti. L'interoperabilità diventa possibile perché qualsiasi agenzia può creare un insieme di dati strutturato utilizzando i nomen costruiti secondo il proprio schema preferito e più adatto ai destinatari specifici di quel sistema bibliografico in termini di lingua, di traslitterazione e di contesto d'uso ma, allo stesso tempo, l'agenzia può fare riferimento al 'significato' universale di un proprio nomen identificando la specifica *istanza* dell'entità a cui quel nomen viene *associato*.

Tuttavia la separazione tra nomen ed entità alla quale sono associati uno o più nomen ha un'altra importante conseguenza. La descrizione dell'istanza dell'entità 'denominata' avviene con la modellazione di un insieme modulare di attributi che, essendo specifici di quella entità e non del nomen associato, potranno essere liberamente e apertamente scambiati tra tutti i produttori di dati e metadati, anche se provenienti da ambiti completamente differenti. In questo modello quindi l'entità (per esempio, la persona) è priva di un dato che è specifico e proprio di un catalogo (come il nomen) e invece ha molti attributi che sono 'tipici' e che si ritroveranno in modelli con entità simili anche di natura completamente diversa da quella bibliografica (per esempio, DBPedia, FOAF, RiCO, CIDOC-CRM ecc.). Per capirci, i dati associati solo alla persona sono gli stessi che saranno validi anche in un modello archivistico, o museale ... o nell'intero web semantico.

Il presupposto per lo scambio dei dati sulle entità associate a un nomen è che ciascuna entità sia opportunamente identificata, all'interno e all'esterno del sistema.

L'identificazione dell'istanza dell'entità 'denominata' avviene per mezzo di un identificatore, che "è un tipo di nomen destinato a essere persistente e unico nell'ambito di uno specifico dominio di applicazione, quali gli identificatori di pubblicazioni di tipo specifico o gli identificatori delle persone, in modo che le istanze dell'entità si possano specificamente identificare e vi si possa fare riferimento senza ambiguità" (IFLA 2020, 88). Per questo scopo, l'identificatore è costituito in modo tale che "il valore dell'attributo *stringa del nomen* di un identificatore non può essere identico al valore dell'attributo *stringa del nomen* di qualsiasi altro *nomen* **all'interno di un determinato sistema** (naturalmente, altri *nomen* al di fuori di quel sistema potrebbero avere lo stesso valore dell'attributo *stringa del nomen*)." (IFLA 2020, 88)

È necessario chiarire la differenza tra il processo di identificazione a livello locale e a livello globale – ed è utile introdurre una distinzione terminologica. Nel primo caso, sarebbe più corretto parlare di *individuazione*, cioè dell'indicazione precisa di un oggetto in una classe (cioè all'interno di un sistema). È la funzione svolta, per esempio, con un'intestazione uniforme in SBN redatta secondo le REICAT: con questo sistema, l'entità (persona, ente, opera, luogo ecc.) che ha come attributo l'intestazione è identificata con certezza all'interno del sistema (cioè il catalogo SBN), ma non necessariamente al di fuori del sistema stesso; ciò avviene perché "l'intestazione uniforme per una persona si basa sul nome [...] sufficiente a identificare la persona **nel catalogo**" (REICAT, 15.1), e perché eventuali qualificazioni si aggiungono al nome solo "se due o più intestazioni (compresi i rinvii) risulterebbero identiche" (REICAT, 15.3). Quindi l'intestazione uniforme così definita non



ha le caratteristiche necessarie per consentire la navigazione o l'esplorazione "verso altri cataloghi e in contesti non legati alle biblioteche", come richiesto da ICP.

È preferibile parlare di *identificazione* solo quanto è assoluta, cioè rimane valida anche al di fuori di uno specifico sistema. Solo un'identificazione intesa in questo modo consente di assicurare la correttezza della navigazione "verso altri cataloghi e in altri contesti non legati alle biblioteche".

La valenza del modello IFLA LRM per le informazioni bibliografiche si basa proprio sulla capacità di esprimere la differenza tra ciò che è rilevante, necessario e sufficiente all'interno del sistema bibliografico – ovvero i nomi costituiti secondo le regole di catalogazione applicate e identificativi a livello locale – e ciò che è rilevante e necessario all'esterno del sistema bibliografico.

Per garantire il processo d'identificazione all'esterno del sistema bibliografico, è necessario dotare l'entità di un dispositivo d'identificazione a livello globale. Questa è la fase della *riconciliazione*, ovvero della ricerca e assegnazione a specifiche istanze di un'entità degli identificatori globali disponibili in uno o più insiemi di dati esterni. Con la riconciliazione, tutti gli identificatori che, nel web semantico, si riferiscono alla medesima istanza di una certa entità vengono raggruppati insieme in un cluster (o grappolo). Maggiore è il numero degli identificatori raggruppati in un cluster di un dataset per una certa istanza di entità, maggiore è la rilevanza di quel dataset come hub – cioè come punto di snodo - del web semantico.

Quindi, nel modello IFLA LRM all'insieme dei nomen di una specifica entità definiti all'interno di un sistema locale si affida il compito del riconoscimento di un'entità all'interno di quel sistema (individuazione), mentre al cluster degli identificatori (un particolare tipo di nomen) si affida il compito di identificare un'istanza di un'entità di un sistema anche al suo esterno (identificazione). La clusterizzazione, la distinzione tra individuazione e identificazione, la distinzione tra nomen e entità associata forniscono elementi validi per rispondere al quesito posto in partenza: ovvero che alcune entità descritte da IFLA LRM – come l'entità persona – possono esistere nel web semantico, cioè essere metadate anche in base ad altri modelli e soprattutto anche indipendentemente da una loro individuazione e descrizione all'interno del sistema informativo di una biblioteca, un museo o un archivio.

Questa prospettiva rovesciata in cui alcune entità del web semantico – definite nella LOD Cloud – interessano e sono presenti anche in modelli come IFLA LRM, RiC-CM, e CIDOC-CRM, consente di fare alcune interessanti considerazioni.

La prima considerazione è che a partire da un enorme insieme di dati potenzialmente disponibili su un altrettanto ampio insieme di entità rappresentate nella Linked Open Data Cloud, per ottenere i dati utili e necessari per la creazione di un sistema bibliografico è necessario individuare le entità di interesse, le loro reciproche potenziali relazioni e le loro caratteristiche rilevanti. L'aggettivo *rilevanti* qui va inteso come l'equivalente di 'funzionale' nel contesto di FRBR. Ciò significa che, tra tutte le possibili scelte, i dati e metadati che definiscono l'ontologia del sistema che si vuole costruire devono essere determinate in base alle funzioni attribuite al sistema stesso, funzioni che nel caso dell'universo bibliografico sono definite dai *Principi internazionali di catalogazione* (IFLA Cataloguing Section e IFLA Meeting of Experts on an International Cataloguing Code 2016).

Sono le funzioni tradizionali e nuove dei cataloghi che guidano la selezione di ciò che è rilevante, ovvero di quali sono i dati e i metadati necessari per rappresentare l'universo bibliografico tramite uno strumento di indicizzazione. Le funzioni tradizionali e nuove sono l'elemento che consente di creare un modello logico corretto. La necessità di questo processo di selezione mette in chiara evi-

denza l'importanza del ruolo del catalogatore, che svolge un lavoro intellettuale complesso e viene quotidianamente posto di fronte a scelte operative, e del suo compito di trovare un equilibrio tra il rispetto dei principi internazionali e delle funzioni del catalogo e la discrezionalità necessaria per realizzare uno strumento adatto ai propri lettori, come avevano segnalato Andrew Osborne e poi Seymour Lubetzky in un dibattito di metà Novecento (Osborn 1941; Lubetzky 2001; Revelli 2001). Una seconda considerazione offerta dalla prospettiva che guarda ai dati e ai metadati delle istituzioni culturali come un sottoinsieme della LOD Cloud è che essa offre il migliore punto di vista dal quale si può comprendere la differenza che esiste tra metadattazione e catalogazione. In una prospettiva in cui le biblioteche sono solo una tra le tante agenzie che – come archivi e musei – creano e condividono dati e metadati, è evidente che l'integrazione e l'interoperabilità tecnica delle biblioteche nell'ecosistema informativo e culturale si possono realizzare solo producendo dati e metadati. La metadattazione come processo di produzione e di riuso di dati e metadati di fonti diverse diventa un'attività centrale delle biblioteche, che ha un insieme di obiettivi tradizionali e nuovi, dove gli obiettivi tradizionali sono quelli legati proprio alla catalogazione in senso stretto. In conclusione, la metadattazione è un'attività più estesa e flessibile in grado di svolgere molte funzioni possibili grazie al web semantico, ma che include e svolge tutti i compiti tradizionalmente assegnati alla catalogazione.

Un'altra suggestione che deriva da questa prospettiva è che offre una migliore comprensione delle scelte di modellazione adottate da IFLA Library Reference Model. Lo si è visto nel caso della modellazione di una relazione tra un'entità e il nomen, ma vale anche rispetto alla scelta di trasformare molti attributi che si riferiscono ad aspetti cronologici o geografici di certe entità in relazioni tra quelle entità e le nuove entità arco di tempo e luogo. Quindi non più attributi come data di nascita o data di morte, o data di inizio o data di fine della serie, o luogo della sede legale, o luogo topico, ma relazioni – debitamente qualificate – tra un'entità e un arco di tempo o un luogo. In particolare il luogo, che nei modelli logici passati era preso in considerazione esclusivamente in quanto "soggetto di un'opera", nel nuovo modello diventa un'entità pienamente autonoma e più facilmente interoperabile con dataset esterni che si occupano dei luoghi (per esempio, GeoNames).<sup>4</sup> Un altro aspetto interessante del nomen è che, oltre a rappresentare e differenziare le diverse forme del nome con il quale un'entità compare in un insieme di risorse, consente di trovare e collegare le risorse che condividono una particolare forma del nomen di quell'entità.

La figura 3 mostra un esempio di modello di registrazione dei dati che rende possibile rappresentare il modo nel quale un'istanza di un'entità è indicata su una manifestazione; nella figura, si vede che nella relazione di tipo autore tra l'istanza della persona Ranganathan e l'istanza della manifestazione della prima edizione de *The five laws of library science* è possibile registrare il modo in cui il nome di Ranganathan figura sul frontespizio. Il dettaglio che si può raggiungere qui è molto alto: non solo è possibile registrare la formulazione che compare su una risorsa quando si descrive la risorsa nel suo complesso (ciò è previsto da IFLA LRM attraverso il "manifestation statement", LRM-E4-A4) per facilitare la collazione tra la descrizione della risorsa e la risorsa che si ha in mano, ma si può anche indicare come un autore viene presentato di volta in volta.

---

<sup>4</sup> <https://www.geonames.org/>

The screenshot shows the 'Browse Primary Sources' interface for the entity 'Ranganathan'. On the left, there is a sidebar with the title 'Browse Primary Sources' and a 'back to top' link. The main content area displays three source categories:

- autore P50:** Ranganathan / Q457933. It includes a sub-label 'indicato come / P1810', a list of references (currently 0), and buttons for '+ aggiungi riferimento' and '+ aggiungi valore'. A 'modifica' button is also present.
- editore P123:** Madras Library Association / Q96807861 <sup>#inglese</sup>. It includes a list of references (currently 0) and buttons for '+ aggiungi riferimento' and '+ aggiungi valore'. A 'modifica' button is also present.
- luogo di pubblicazione P291:** Chennai / Q1352. It includes a list of references (currently 0) and buttons for '+ aggiungi riferimento' and '+ aggiungi valore'. A 'modifica' button is also present.

Figura 3 – Qualificazione della forma del nome di una entità su una specifica fonte in Wikidata

Infine, con lo stesso schema logico, diventa possibile rappresentare la presenza di una particolare forma del nome di una istanza di un'entità anche in qualsiasi altro repertorio bibliografico o strumento catalografico. Nella figura 4 si vede che si può registrare la specifica forma del nome utilizzata in SBN per Pietro d'Abano.

The screenshot shows the Wikidata entity page for Pietro d'Abano. The 'SBN author ID' property is highlighted with a red box, showing the value 'Pietro : d'Abano <ca. 1250-ca. 1315>'. Other properties include:

- FAST ID:** + add value
- National Library of Israel J9U ID:** 987007296542305171, + add value
- Library of Congress authority ID:** n65112114, + add value
- National Library of Lithuania ID:** LNB:B/H8;=wC, + add value

Figura 4 – Registrazione in Wikidata della forma verbale corrispondente all'identificatore di SBN

Un ultimo aspetto da prendere in considerazione nel modello IFLA LRM è la sua capacità di rappresentare in modo funzionale alcuni fenomeni tipici dell'ambito culturale, e bibliografico in particolare. Si tratta per esempio per l'entità persona degli anonimi, degli pseudonimi e degli pseudonimi collettivi,<sup>5</sup> per l'entità opere dei titoli<sup>6</sup> e per l'entità luogo dei luoghi non dichiarati, luoghi falsi (cioè che denotano un luogo esistente diverso dal luogo effettivo di stampa) e luoghi falsi immaginari (cioè luoghi inesistenti con toponimi inventati in sostituzione del luogo effettivo di stampa).

L'occultamento dell'identità reale di persone e luoghi sono determinati, sul piano sociale e culturale, dalla volontà degli agenti responsabili dei diversi aspetti della risorsa (per esempio, in qualità di autori o di editori) di evitare eventuali conseguenze negative legate alle restrizioni alla libertà di stampa.<sup>7</sup> L'autore sceglie di non comparire, con l'anonimato, o di comparire sotto uno pseudonimo; oppure è un gruppo di autori che adotta uno pseudonimo collettivo. Spesso a un luogo di stampa falso (cioè l'indicazione di un luogo reale in sostituzione del luogo di stampa effettivo) o a un luogo di stampa immaginario (cioè l'indicazione di un luogo inesistente in sostituzione al luogo di stampa effettivo) si accompagna una formulazione di responsabilità relativa alla risorsa sotto forma di pseudonimo.

Anche se è opportuno rinviare a uno studio specifico l'analisi di come il modello IFLA LRM riesce a fornire uno schema per la creazione di dati coerenti per rappresentare le diverse casistiche, questa riflessione consente di sottolineare un aspetto fondamentale della costruzione dei modelli: i modelli che si occupano di un certo ambito disciplinare devono avvalersi sì delle entità definite e disponibili nel web semantico, ma devono anche saper prevedere e modellare le entità e i fenomeni propri e tipici di quello specifico ambito. Questo vale per la bibliografia e la biblioteconomia, ma certamente anche per l'archivistica e la museologia. In questo esercizio si riconosce, a maggior ragione, la professionalità specifica del bibliotecario, dell'archivista e del curatore.

Per concludere, l'approccio che consente di osservare le entità descritte nei modelli bibliografici come IFLA LRM, ma si può estendere la riflessione sicuramente anche ad altri modelli come RiC-CM (International Council on Archives. Experts group on archival description 2016) e CIDOC-CRM («Definition of the CIDOC conceptual reference model» 2021) si rivela utile alla comprensione dei fenomeni di uno specifico dominio.

Tuttavia, il portato più rilevante che si ha adottando questa prospettiva è che non esistono più un universo bibliografico, uno archivistico e uno museale *totalmente* separati tra loro, perché la distinzione di ciò che è peculiare di un ambito (come un nomen) da ciò che non lo è (l'entità associata, come una persona, un luogo, un'opera ecc.), consente di vedere più chiaramente come alcune entità sono comuni a tutti e tre gli universi e con il web semantico in generale. Ci sono, e ri-

---

<sup>5</sup> L'entità persona si può riferire in questo caso a qualsiasi tipo di responsabilità (intellettuale, editoriale, materiale ecc.).

<sup>6</sup> Si veda, sul tema dei titoli, la relazione di Mauro Guerrini, *Dal titolo uniforme al titolo preferito: l'importanza del contesto culturale e linguistico*, al 1° Seminario Jlis.it "Modellare la conoscenza. Standard archivistici e bibliografici a confronto" tenutosi a Firenze il 19 maggio 2022, <https://www.youtube.com/watch?v=eXu8cuoQJxU>.

<sup>7</sup> Secondo Marino Parenti (1951, 10–11): "Il ricorso al falso luogo di stampa e a non meno falsi nomi di stampatori trova ragione genericamente, nelle restrizioni, più o meno giustificate, alla libertà di stampa. Di qui anche, gli stretti legami tra i falsi dati editoriali e tipografici e l'anonimo e lo pseudonimo. Ma più particolarmente, queste ragioni, debbono ricercarsi, in origine, nelle controversie religiose – e l'uso nasce infatti con la Riforma – quindi nella politica e nella letteratura licenziosa".

marranno sempre, alcune specificità tipiche di ciascun universo, ma sicuramente, per certe entità, questi universi specialistici sono soltanto una o più delle possibili faccette dell'universo descritto e rappresentato, nella sua totalità e completezza, nel web semantico. I modelli logici che descrivono gli universi specialistici di cui ci occupiamo – come IFLA LRM, RiC-CM e CIDOC-CRM – esaminano un sottoinsieme di entità del web semantico e – sotto la guida di principi generali definiti all'interno dei diversi ambiti disciplinari – ne prendono in considerazione un numero scelto e limitato di caratteristiche e specificano le caratteristiche di quelle entità in relazione agli scopi specifici – tradizionali e nuovi – dello strumento di ricerca che si vuole allestire.

Perciò adottare un approccio ai nostri sistemi di organizzazione e recupero della conoscenza che va dall'esterno all'interno, oltre a quello tradizionale dall'interno verso l'esterno, può essere davvero utile. Per esempio, per riconoscere con maggiore facilità che una persona può ricoprire infiniti ruoli, culturali e non. Si può affermare, inoltre, che in alcuni ruoli specifici diventa un autore; ma può diventare di volta in volta un curatore o un traduttore di un testo, ma anche un produttore d'archivio, o un pittore, o un artista, e così via.

La prospettiva che vede le entità dei nostri modelli logici come parte del web semantico e che suggerisce di seguire un percorso che va dall'esterno all'interno offre come valore aggiunto fondamentale per le nostre discipline una visione unificante. Consente di riconoscere che le nostre discipline bibliografiche, bibliotecarie, archivistiche e museali – tradizionalmente diverse – si occupano di molte entità per le quali condividono un forte interesse, ma che guardano attraverso una prospettiva specifica;<sup>8</sup> questa presa di coscienza è forse il presupposto migliore per iniziare a progettare finalmente un percorso comune.

---

<sup>8</sup> Come per i profili di Pierpaolo Pasolini nella relazione di Tiziana Possemato, *Entity modeling: tracce di un percorso in evoluzione*, presentata al 1° Seminario Jlis.it “Modellare la conoscenza. Standard archivistici e bibliografici a confronto” tenutosi a Firenze il 19 maggio 2022, <https://www.youtube.com/watch?v=eXu8cuoQJxU>.

## Riferimenti bibliografici

Balsamo, Luigi. 1984. *La bibliografia: storia di una tradizione*. Firenze: Sansoni.

«Definition of the CIDOC conceptual reference model». 2021. Version 7.1.1. ICOM/CIDOC CRM Special Interest Group. <https://www.cidoc-crm.org/version/version-7.1.1>.

Domanovszky, Ákos. 1975. *Functions and objects of author and title cataloguing*. München: Verlag Documentation.

Guerrini, Mauro, a c. di. 1996. *Il linguaggio della biblioteca scritti in onore di Diego Maltese*. Milano: Bibliografica.

IFLA. 2017. *IFLA Library Reference Model. A Conceptual Model for Bibliographic Information*. A cura di Pat Riva, Patrick Le Boeuf, e Maja Zumer. Den Haag: IFLA. [https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla\\_lrm\\_2017-03.pdf](https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla_lrm_2017-03.pdf).

———. 2020. *IFLA Library Reference Model. Un modello concettuale per le informazioni bibliografiche*. A cura di Pat Riva, Patrick Le Boeuf, e Maja Zumer. Edizione italiana. Roma: ICCU.

IFLA Cataloguing Section e IFLA Meeting of Experts on an International Cataloguing Code. 2016. *Statement of International Cataloguing Principles (ICP)*. Den Haag: IFLA. [https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp\\_2016-en.pdf](https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2016-en.pdf).

IFLA Study Group on the Functional Requirements for Bibliographic Records. 2009. «Functional Requirements for Bibliographic Records: Final Report / Approved by the Standing Committee of the IFLA Section on cataloguing. September 1997; as amended and corrected through February 2009». [http://www.ifla.org/files/cataloguing/frbr/frbr\\_2008.pdf](http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf).

IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). 2009. «Functional Requirements for Authority Data: a Conceptual Model. Final report». <https://doi.org/10.1080/14649055.2011.10766293>.

IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR). 2010. *Functional Requirements for Subject Authority Data (FRSAD). A conceptual model*. München: K. G. Saur. <https://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>.

International Council on Archives. Experts group on archival description. 2016. *Records in Contexts. A conceptual model for archival description*. Consultation draft v0.1. <https://www.ica.org/en/egad-ric-conceptual-model-ric-cm-01pdf>.

Lubetzky, Seymour. 1963. «The function of the main entry in the alphabetical catalogue: one approach». In *International Conference on Cataloguing Principles, Paris, 9th-18th October, 1961. Report*, 139–43. London: IFLA International Office for UBC.

———. 2001. *Writings on the classical art of cataloging*. A cura di Elaine Svenonius e Dorothy McGarry. Englewood, Colorado: Libraries Unlimited.

Osborn, Andrew D. 1941. «The crisis in cataloging». *The library quarterly* 11 (4): 393–411. <https://doi.org/10/bjqn7>.

Parenti, Marino. 1951. *Dizionario dei luoghi di stampa falsi, inventati o supposti in opere di autori e traduttori italiani : con un'appendice sulla data «Italia» e un saggio sui falsi luoghi italiani usati all'estero, o in Italia, da autori stranieri*. Biblioteca bibliografica italiana 1. Firenze: Sansoni antiquariato.

Revelli, Carlo. 2001. «Andrew D. Osborn. La catalogazione in crisi: un articolo che sessant'anni fa segnò una svolta. Premessa e traduzione di Carlo Revelli». *Biblioteche oggi* 19 (1): 44–51.

Sardo, Lucia. 2017. «La lingua del catalogatore (parte 1). L'autore». *Bibliothecae.it* 6 (2): 339–83. <https://doi.org/10.6092/ISSN.2283-9364/7739>.

Svenonius, Elaine. 2000. *The intellectual foundation of information organization*. Cambridge, Mass.: MIT Press.

———. 2008. *Il fondamento intellettuale dell'organizzazione dell'informazione*. Tradotto da Maria Letizia Fabbrini. Firenze: Le lettere.

Verona, Eva. 1959. «Literary Unit versus Bibliographical Unit» 9 (1–4): 79–104.

———. 1963. «The Function of the Main Entry in the Alphabetical Catalogue. A Second Approach». In *International Conference on Cataloguing Principles. Paris, 9th-18th October, 1961. Report*, a cura di A. H. Chaplin e Dorothy Anderson, 145–57. Organizing Committee of the International Conference on Cataloguing Principles.

## From the uniform title to the preferred title: the prominence of the cultural and linguistic context

Mauro Guerrini<sup>(a)</sup>

a) University of Florence, <https://orcid.org/0000-0002-1941-4575>

**Contact:** Mauro Guerrini, [mauro.guerrini@unifi.it](mailto:mauro.guerrini@unifi.it)

**Received:** 26 April 2022; **Accepted:** 13 June 2022; **First Published:** 15 September 2022

### ABSTRACT

By anchoring the title to a cultural context, it is possible that a universally valid title form (uniform title) does not exist or exist; the form used at the national or regional level is a form commonly used in that or those countries, that is, in a local or national context, not a global one. The concept of preferred title goes beyond, relativizing it, the cataloging model, considered an absolute time for every reality; it is acknowledged that what is valid for a specific cultural and linguistic context does not necessarily apply to the international context. The dissimilar historical behavior is exemplified by two examples: “One Thousand and One Nights”, and “The Chanson de Roland”.

### KEYWORDS

Uniform title; Cataloging; Controlled access point; Bibliographic control.

## Dal titolo uniforme al titolo preferito: l'importanza del contesto culturale e linguistico

### ABSTRACT

Con l'ancoraggio del titolo a un contesto culturale si prende atto che non esiste né può esistere una forma del titolo (titolo uniforme) valida universalmente perché la forma usata a livello nazionale o regionale (più paesi) è una forma comunemente usata in quel o in quei paesi, ovvero in un contesto locale o nazionale, non globale. Il concetto di titolo preferito supera, relativizzandolo, il modello di catalogazione nominale considerato un tempo assoluto per ogni realtà; si prende atto che ciò che vale per un determinato contesto culturale e linguistico non vale necessariamente per l'ambito internazionale. Il comportamento storico dissimile è esemplificato tramite due esempi: “Le mille e una notte” e la “Chanson de Roland”.

### PAROLE CHIAVE

Titolo uniforme; Catalogazione; Punto d'accesso controllato; Controllo bibliografico.



*Today we recognize the importance of context. You have to understand the context where the innovation or policy is to be introduced. The recipients must be heard, and treated with respect. [...] Cataloguing is for connecting people with resources.*

Peter Lor

## Premessa

I primordi della normalizzazione delle registrazioni bibliografiche sono riscontrabili nella struttura del catalogo della Bodleiana curato da Thomas Hyde (1636-1703) (opera preceduta dalla mirabile *Praefatio ad lectorem* del 1674) e del catalogo della Casanatense di Giovanni Battista Audiffredi (1714-1794). In particolare, il titolo di un'opera ha avuto spesso una variabilità nel tempo; ogni epoca, infatti, ha adottato criteri differenti d'identificazione e denominazione delle opere, in particolare di quelle del mondo antico, con titoli diversi attribuiti nelle edizioni a stampa a caratteri mobili e nelle varie lingue di traduzione. Il titolo delle opere, almeno di quelle della letteratura occidentale dal XIV secolo in poi, si è stabilizzato prevalentemente negli ultimi due secoli.

Occorre precisare tre punti:

1. titolo uniforme è un concetto che riguarda l'oggetto primario del catalogo: la pubblicazione o l'opera? Come non ricordare la discussione fra Eva Verona e Seymour Lubetsky (Buizza 2002) e il breve quanto fondamentale saggio *Literary unit versus bibliographical unit* (Verona 1959). Il tema è inserito all'interno del controllo d'autorità come aspetto strategico di un ragionamento ancora più ampio: la struttura sindetica del catalogo;
2. il titolo uniforme nella storia della catalogazione è, tendenzialmente, il titolo dell'opera in lingua originale così come si è stabilizzato nel tempo; ciò, tuttavia, vale solo per i titoli formulati in una delle lingue usate dal catalogo (concetto quanto mai generico e indefinito);
3. FRBR è il punto d'approdo di un dibattito iniziato nella seconda metà degli anni Ottanta del Novecento, con il ritorno del tema dell'opera considerata centrale rispetto alla pubblicazione.

Il criterio della stabilizzazione copre ambiti e processi distinti:

- i titoli delle edizioni.
- un titolo, fra quelli di edizione o di repertorizzazione, che si afferma come titolo di denominazione dell'opera;
- il titolo di riferimento per quell'opera nelle varie lingue.

## Uniformità: richiami storici e definizioni

Il tema dell'uniformità, pur oggetto di riflessione da tempo, non è mai stato perseguito come obiettivo (e strumento) unificante nei codici catalografici, a partire da quello emanato dall'ALA nel 1908. Il concetto ha avuto una pienezza elaborativa grazie a tre grandi bibliotecari e studiosi del Novecento: Seymour Lubetsky (1898-2003), Akos Domanovsky (1902-1982) ed Eva Verona

(1905-1996); tutti e tre sono stati protagonisti dell'ICCP, International Conference on Cataloguing Principles del 1961. Alla conclusione dei lavori furono emanati i *Principi di Parigi* relativi alla scelta e la forma delle parole d'ordine nel catalogo per autore, come, al tempo, si chiamava l'elemento d'accesso alla registrazione bibliografica. Il tema del titolo uniforme è poi stato ripreso dagli ICP, *International Cataloguing Principles* del 2009 (con aggiornamento del 2016 e 2017) e dai modelli concettuali emanati dall'IFLA dal 1998 al 2017, ovvero da FRBR a IFLA LRM.

Il titolo uniforme, concettualmente, ha avuto e ha la funzione di facilitare l'identificazione di un'opera nel catalogo, cioè di aggregare e differenziare le varie edizioni di un'opera, riunendo le loro registrazioni sotto un'unica denominazione; è, pertanto, un dispositivo catalografico utilizzato per evitare la dispersione dei record indicizzati sotto titoli differenti corrispondenti ai titoli delle edizioni pubblicate nella medesima lingua e soprattutto in lingue (traduzioni) diverse.

Di per sé il concetto di *titolo uniforme* non significa uniformare a una tradizione letteraria, almeno per le opere di un autore; significa piuttosto riconoscere che fra le possibili denominazioni (o *nomen*, nel linguaggio di IFLA LRM) di un'opera, nel suo testo originale, ve n'è una che si è stabilizzata maggiormente sia nell'universo editoriale (in genere, quella della prima edizione in lingua originale) sia nei repertori; per esempio, sulla base delle edizioni di *The Tragedy of Hamlet*, storicamente è il titolo *Hamlet* che si è affermato.<sup>1</sup>

Il titolo uniforme, pertanto, non deriva solo dai titoli presenti sulle risorse, bensì è il titolo prevalente nella memoria culturale di un Paese, così com'è attestato nei repertori: ciò è motivo di una problematica complessa e difficilmente risolvibile. Il titolo uniforme, ovvero quel titolo assunto come riconoscimento culturale nell'universo letterario, è, dunque, il nucleo centrale d'individuazione in un cerchio, dal quale si dipanano con distanza e relazioni mutevoli sia i titoli delle edizioni sia i titoli delle diverse memorie culturali.

La riflessione sull'argomento ha assunto un nuovo paradigma nell'era digitale, periodo in cui un record creato da un'agenzia bibliografica o da una biblioteca è consultabile da qualsiasi lettore in ogni parte del mondo;<sup>2</sup> non si tratta più di definire un punto d'accesso uniforme, bensì di valutare quali punti d'accesso creare per permettere all'utente di trovare, identificare, selezionare e ottenere (le funzioni utente di FRBR) la risorsa desiderata.

Nei codici catalografici, nei manuali e nei dizionari biblioteconomici il titolo uniforme, in un servizio di reperimento bibliografico, è definito come il titolo con cui un'opera (o una sua parte), pubblicata e conosciuta con titoli varianti, viene identificata ai fini della creazione del punto d'accesso del record. In altre parole, il titolo uniforme è il titolo distintivo selezionato per rappresentare un'opera emessa sotto più titoli in risorse diverse (Smiraglia 2001). *Wikipedia* (inglese) dà una definizione sintetica molto puntuale: "Un titolo uniforme nella catalogazione bibliotecaria è un titolo distintivo assegnato a un'opera che non ha titolo o è apparsa sotto più titoli. Stabilire un titolo uniforme è un aspetto del controllo d'autorità".<sup>3</sup>

<sup>1</sup> Si veda l'interessante caso citato da Agatha Christie circa il titolo di un suo romanzo, *And there where none*, citato dall'autrice per esemplificare un esempio di "titolo uniforme", vedi: *And then there where none*. In: *Wikipedia, the free encyclopedia*, <https://tinyurl.com/2uhn56pp/>.

<sup>2</sup> Pur in una dimensione diversa, ma nella medesima prospettiva dell'uniformità, il problema è stato affrontato in maniera puntuale da Françoise Leresche, *Découvrir l'entité Expression: son intérêt pour les utilisateurs*, <https://youtu.be/mcZXwn-gW9xw?t=15207>.

<sup>3</sup> [https://en.wikipedia.org/wiki/Uniform\\_title](https://en.wikipedia.org/wiki/Uniform_title).

La questione del titolo uniforme s’inserisce, infatti, all’interno della tematica dell’authority control, cioè della creazione del *punto d’accesso controllato* (già chiamato *intestazione* prima di ICP del 2009) per favorire il lettore nella funzione di ricerca tramite la collocazione della risorsa all’interno di un insieme di record, ovvero il raggruppamento sotto:

- a) il nome dell’autore delle registrazioni delle manifestazioni di una sua opera pubblicata con titoli diversi;
- b) il titolo uniforme di ciascuna opera anonima delle registrazioni delle varie manifestazioni delle opere pubblicate di cui sia stato dimenticato il nome dell’autore.

Il titolo uniforme, infatti, risponde alla *funzione di raggruppamento*, ovvero la funzione ideata affinché il lettore trovi, individui, selezioni e ottenga tutte le espressioni di un’opera, le cui manifestazioni si presentino con titoli differenti, compresi i titoli di traduzioni e le variazioni ortografiche, anche all’interno del medesimo contesto culturale e linguistico; come, per esempio, la stessa opera pubblicata nel tempo con titoli differenti rispetto a quello della prima edizione, per esempio, *Novellino*, oppure *Cento novelle antiche*, oppure *Libro di novelle e di bel parlar gentile*.

## Le tipologie di titolo uniforme

Il *titolo uniforme* interessa almeno cinque situazioni bibliografiche, con un grado diversificato di variabilità:

1. titoli riconducibili a un autore per “associazione costante”; non si tratta del titolo originale dell’opera, ovvero del titolo della prima edizione, bensì del titolo in lingua originale “stabilizzato” nel tempo come identificante l’opera; per esempio, *La Divina Commedia* rispetto a *Commedia*,<sup>4</sup> oppure *Le avventure di Pinocchio* rispetto a *Le avventure di un burattino* o a *Pinocchio* (titolo molto ricorrente);
2. titoli con attribuzione di responsabilità intellettuale variata nel tempo, e quindi con associazioni diversificate; per esempio, *Il Fiore*, opera dall’identità incerta del suo autore, per il quale sono stati ipotizzati i nomi di Dante Alighieri e di Dante da Maiano;
3. titoli di opere anonime, per esempio, *Chanson de Roland*;
4. titoli di testi sacri, per esempio, *Bibbia*, e di opere liturgiche;
5. titoli di opere musicali, per esempio, *Carmina Burana*.<sup>5</sup>

La gestione del titolo di un’opera anonima rappresenta la parte più complessa del problema per la forte instabilità dei titoli, seppure anche le manifestazioni delle opere di un autore possano avere pari instabilità. Non casualmente, il titolo uniforme per un’opera, come il nome di un autore, può cambiare da un’epoca a un’altra, contrassegno di convenzioni diverse in un mutato contesto culturale e letterario.<sup>6</sup>

<sup>4</sup> Titolo usato da Treccani nell’edizione Roma: Istituto della Enciclopedia italiana, 2018.

<sup>5</sup> Joan M. Reitz, *ODLIS: Online Dictionary for Library and Information Science*, [www.abc.clilo/ODLIS/odlis\\_A.aspx](http://www.abc.clilo/ODLIS/odlis_A.aspx).

<sup>6</sup> Sarebbe utile disporre di un repertorio delle opere anonime; da qualche decennio l’IFLA sostiene il gruppo di lavoro Anonymous classics che, non casualmente, procede molto lentamente; vedi: *Report 2011*, [https://cdn.ifla.org/wp-content/uploads/2019/05/assets/Cataloguing/reports/anonymous-classics-report\\_2011.pdf](https://cdn.ifla.org/wp-content/uploads/2019/05/assets/Cataloguing/reports/anonymous-classics-report_2011.pdf).

## I criteri per la scelta del *titolo uniforme*

Quali sono i criteri per la scelta del titolo uniforme? I cataloghi dei vari paesi testimoniano uno storico comportamento difforme nella scelta del titolo uniforme, mentre i principi, i modelli concettuali e la normativa catalografica corroborano da sempre l'irrisolutezza della questione, almeno dal punto di vista teorico. AACR2 25.1, per esempio, prescrive che la decisione si fondi sulla notorietà dell'opera, sulla quantità delle manifestazioni dell'opera, sulla presenza di un'altra opera con lo stesso titolo o, per un'opera post 1500, sul titolo con cui l'opera è diventata nota nelle sue manifestazioni o nelle fonti di riferimento. AACR2 25.3C3 prevede un caso specifico e sempre più diffuso nell'editoria scientifica per le opere pubblicate simultaneamente in lingue diverse e con titoli differenti, nessuna delle quali e nessuno dei quali riconoscibili come originale; come titolo uniforme prescrive di usare un ordine di preferenza, con precedenza al titolo formulato nella lingua dell'agenzia bibliografica che cataloga; viene scelto un titolo, pertanto, che varia da Paese a Paese. Se non c'è alcun titolo in una lingua usata dal catalogo, il codice prescrive di usare il titolo dell'edizione ricevuta per prima. Nessuna norma è più arbitraria di questa che contrasta proprio con il concetto di uniformità dei cataloghi.

FRBR al punto 4.2.1 legge: "Se l'opera è apparsa sotto vari titoli (diversi nella forma, nella lingua ecc.) l'agenzia bibliografica, di norma, ne sceglie uno come base per il 'titolo uniforme' al fine di nominare e fare riferimento all'opera in modo coerente".

ICP 5.3.3.2, *Scelta del titolo preferito per opere, espressioni, manifestazioni ed esemplari* introduce il termine *titolo preferito* e abbandona esplicitamente il lemma *titolo uniforme*; il punto recita: "Quando forme varianti del titolo dell'opera si trovano nelle manifestazioni, si dovrebbe attribuire la preferenza: a) al titolo che appare nella prima manifestazione dell'espressione originale dell'opera, generalmente nella lingua originale, oppure b) al titolo comunemente usato".

Le REICAT, al punto 9.1, *Scelta del titolo uniforme*, richiamano la complessità della questione: "Il titolo uniforme per un'opera si basa sul titolo con cui l'opera stessa è generalmente identificata. Può trattarsi del titolo originale, scelto dall'autore o assegnato nella prima pubblicazione dell'opera, di una sua variante o di un titolo adottato nelle edizioni successive, delle parole iniziali del testo (*incipit*), di un titolo tradizionale o convenzionale o di un'altra designazione comunemente usata. In alcuni casi può rendersi necessaria la formulazione di un titolo". Se un'opera è conosciuta con più titoli si sceglie il *titolo prevalentemente usato* (9.1.1). Il criterio della *prevalenza* è di difficile definizione e variabile da contesto a contesto.

In definitiva, per i modelli concettuali, i principi e i codici di catalogazione non esiste un criterio, ovvero, ne esistono di indefiniti e di contrastanti.

## ICP e RDA: *titolo preferito*

ICP nel paragrafo *Termini non più usati* prescrive l'abbandono della formulazione *Titolo uniforme* a favore di *Punto d'accesso autorizzato*, *Forma autorizzata del nome*, *Nome*,<sup>7</sup> definendo ciascuno dei nuovi termini. Coerentemente con lo scambio proficuo a livello terminologico tra ICP e RDA, in *RDA Glossary* non compare *uniform title* ma *preferred title*: "Un titolo selezionato come preferito

<sup>7</sup> La prescrizione compare di nuovo al punto 10 del loro aggiornamento del 2016 (con addenda del 2017).

nella costruzione di un punto d'accesso controllato". *RDA Glossary* distingue ulteriormente, offrendo una definizione specifica di *preferred title* per ognuna delle entità WEMI: *preferred title of work / of expression / of manifestation / of item*. In ciascuna definizione si corrobora l'idea dell'importanza del contesto nella scelta del titolo come punto di accesso controllato; per esempio, *preferred title of a work*: "Un nomen che corrisponde a un titolo dell'opera selezionato come preferito in una specifica applicazione o in un contesto".<sup>8</sup>

La novità è importante. Con l'ancoraggio del titolo a un contesto culturale si prende atto che non esiste né può esistere una forma del titolo (titolo uniforme) valida universalmente perché la forma usata a livello nazionale o regionale (più paesi) è una forma comunemente usata in quel o in quei paesi, ovvero in un contesto locale o nazionale, non globale.

Il concetto di titolo preferito, pertanto, supera, relativizzandolo, il modello di catalogazione nominale considerato un tempo assoluto; si prende atto che ciò che vale per un determinato contesto culturale e linguistico non vale necessariamente per l'ambito internazionale.

## Il comportamento storico dissimile in due esempi

Vediamo con due esempi – *Le mille e una notte* e la *Chanson de Roland* – il comportamento dei cataloghi di alcune biblioteche del mondo.<sup>9</sup>

*Le mille e una notte*. Il titolo originale è *Hazar Afsaneh (Mille storielle)*. Il numero delle novelle non corrisponde a quello presentato nel titolo; ciò deriva dall'abitudine persiana, presente anche in Arabia e altrove, di utilizzare il numero *mille* per indicare una quantità indefinita di cose. L'edizione nominata *Calcutta I* è la prima in lingua araba. Il testo è diviso in due parti: la prima, pubblicata nel 1814, è intitolata *فلا ةلئل ةلئل نم فلا ةلئل ةلئل* (*Racconti di cento notti delle Mille e una notte*), mentre la seconda, pubblicata nel 1818, è intitolata *ءزلء نم ىنءءءء ةلئل ةلئل* (*Il secondo volume del libro "Le Mille e una notte" che contiene i racconti delle cento notti e le novelle del Sindibad e l'Hindibad*). La prima edizione in arabo non costituisce, come si deduce dal titolo, un'edizione completa dell'opera, bensì un'antologia.<sup>10</sup> L'edizione nominata Breslavia di Habicht e Fleischer è l'unica edizione araba delle *Notti* pubblicata in Europa nell'Ottocento, nella città polacca, fra il 1825 e il 1843, ed è composta da dodici volumi. I primi otto furono pubblicati da Maximilian Habicht fra il 1825 e il 1838. Dopo la sua morte, avvenuta nel 1839, M. Heinrich Fleischer proseguì nella pubblicazione degli altri quattro volumi, che videro la luce fra il 1842 e il 1843. Si presume che questa sia la prima edizione completa delle *Mille e una notte* e per questa ragione l'opera è intitolata nel frontespizio arabo *فلا ةلئل ةلئل نم ءءءءءء ةلئل ةلئل*, ovvero, *Questo è il libro delle Mille e una notte dall'inizio alla fine*.

<sup>8</sup> [https://access.rdatoolkit.org/Glossary?externalId=en-US\\_rdaregistry.info-Elements-w-P10223&highlight=preferred&highlight=title](https://access.rdatoolkit.org/Glossary?externalId=en-US_rdaregistry.info-Elements-w-P10223&highlight=preferred&highlight=title).

<sup>9</sup> Interessante *And then there were none*. In: *Wikipedia, the free encyclopedia*, <[https://en.wikipedia.org/wiki/And\\_Then\\_There\\_Were\\_None](https://en.wikipedia.org/wiki/And_Then_There_Were_None)>. *And there were none* è un romanzo scritto da Agatha Christie, citato dall'autrice per esemplificare un esempio di titolo uniforme.

<sup>10</sup> Edizione Calcutta I. *فلا ةلئل ةلئل نم فلا ةلئل ةلئل* or *The Arabian Nights Entertainment*, on the original Arabic, published under the patronage of the College of Fort William, by Shuekh Uhmud bin Moohummud Shirwanee ool Yumunee, of the Arabic Department, author of the Nufhut-oolumun, ul-ujub-ool-oojab, hudeeque-ool-ufrah, and various Arabic works, Vol. I., *Containing the stories of 100 nights*. Calcutta: Pereira, 1814, 430 p.

برطل الامم برعلل يل اي للا رامسا امومع ي عدي ةل يلو ةل يل فلا باتك (in arabo) باءك  
بمضتي ةءاك فلا ثروي و si indica la quarta edizione araba delle *Notti* in 4 volumi.

La prima traduzione delle *Notti arabe* in italiano vide la luce fra il 1721 e il 1722 pubblicata a Venezia da Sebastiano Coleti. *Novelle arabe divise in mille e una notte tradotte dall'idioma francese nel volgare italiano*, Venezia, Sebastiano Coleti, 1721-1722, XII tomi in IV volumi. La prima traduzione italiana completa deriva dalla versione francese di J.C. Mardrus a opera di Anna Franchi, *Le mille e una notte, prima traduzione italiana completa*, Milano, V. Nougol, 1921, XII volumi.<sup>11</sup>

Alcuni titoli con cui l'opera è citata nei cataloghi:

- Le mille e una notte
- Mille e una notte
- Racconti delle Mille e una notte
- Notti arabe
- Arabian nights
- The Arabian nights
- The thousand and one nights
- One thousand and one Arabian nights
- Tales from the thousand nights and one night
- The Book of the Thousand Nights and a Night
- Les Mille et une nuits
- Mille et une nuits
- Le livre des mille nuits et une nuit
- Tausendundein Nacht
- Erzählungen aus den Tausendundein Nächten
- Alf Layla wa-Layla (arabo)
- Alf Lélah ua Lélah (arabo)
- ألف ليلة و ليلة, *ʿAlf layla wa layla*
- ابش كى و رازه, “Hezār-o yek šab” (persiano)
- Kniga tysiatchi i odnoi notchi (russo)
- De vertellingen van Duizend-en-één-Nacht (Paesi Bassi)
- Tusen og en natt (svedese – film Pasolini)
- 図説「千夜一夜物語 (アラビアン・ナイト) (giapponese)

*La Chanson de Roland* (o *Chanson de Roncevaux*). Il poema epico in endecasillabi è stato scritto nella seconda metà dell’XI secolo ed è considerato tra le opere più significative della letteratura medievale francese. Esso esalta un evento avvenuto tre secoli prima, ovvero la spedizione militare di Carlo Magno contro gli arabi di Spagna che si conclude con la battaglia di Roncisvalle del 15 agosto 778; la retroguardia di Carlo Magno, comandata dal paladino Rolando, fu attaccata e annientata dai ba-

<sup>11</sup> Altre edizioni italiane: *Le Mille ed una notte o racconti meravigliosi*, edizione condotta per uso dei giovinetti, libro di lettura e di premio. Milano: Libreria Editrice di Educazione e d’Istruzione di Paolo Carrara, 1881, ill.; 146 p. *Favole delle Mille e una notte per la gioventù italiana* con sei tavole in cromolitografie e otto incisioni, seconda edizione riveduta e corretta. Milano: Ulrico Hoepli, 1907, ill.; 287 p. *Racconti delle Mille e una notte*, versione di Arturo Jahn Rusconi, illustrazioni di Edmondo Dulac. Bergamo: Istituto italiano d’arti grafiche, 1924, ill.; XIX 139 p.

schì (alleati degli arabi) – nella riscrittura epica divenuti saraceni. Il testo ebbe subito molta fortuna, prima orale e poi scritta, testimoniata dal notevole numero di manoscritti giunti fino a noi. Nel 1832, Henri Monin scoprì nella Librairie du Roi il poema *La chanson de Roland*. All'inizio degli anni Trenta dell'Ottocento si conoscevano solo due manoscritti della *Chanson*, quello di Parigi e quello rinvenuto a Châteauroux. Francisque Michel pubblicò nel 1837 una prima edizione del testo che si basava su un altro manoscritto conservato alla Bodleian Library di Oxford da lui scoperto nel 1835. Altri manoscritti saranno rinvenuti negli anni successivi a Venezia, Versailles, Lione e Cambridge. Alcuni titoli con cui l'opera è citata nei cataloghi:

- Chanson de Roland
- Chanson de Roland: chanson de geste
- La chanson de Roland
- La chanson de Rolland
- La chanson de Roncevaux
- Chanson de Roncevaux
- La gesta de Roldán
- La légende de Roland
- Les textes de la Chanson de Roland
- Chanson de Roland = Cantar de Roldán (con titolo parallelo)
- Chanson de Roland = El cantar de Roldán (con titolo parallelo)
- Chanson de Roland et le roman de Roncevaux
- Chanson de Roland ou de Roncevaux
- La canzone di Rolando
- La canzone d'Orlando
- The song of Roland
- La canción de Rolando
- La cançó de Roland
- A chevalier of old France
- A concordance of the Chanson de Roland
- A translation of The song of Roland
- Cân Rolant
- Can Rolant: the medieval Welsh version of the Song of Roland
- Cançó de Roland: cançó de gesta
- Cantar de Roldán
- Das altfranzösische Rolandslied
- Das Rolandslied
- Das Rolandslied des Pfaffen Konrad
- Dziesma par Rolandu: senfrancu eposs
- El cantar de Roldán
- Het Roelandslied: Frans en Nederlands
- Il fragmento de Roncesvalles
- Münchener Bilderbogen
- O cavaleiro sem igual
- Pesem o Rolandu (sloveno)

- Pesn' o Rolande (russo)
- Pesn' o Rolande : starofrantsùzskiï geroicheskiï épos
- Pisnià pro Rolíanda = La chanson de Roland (con titolo parallelo)
- Pjesme o Rolandu (croato)
- Rappeler Roland
- Rencesval : édition critique du texte d'Oxford de la Chanson de Roland
- Roelantslied (néerlandais)
- Roland
- Roland le vaillant paladin
- Rolandkanto
- Rolando
- Rolandslied
- Roncesvalles
- Roncesvalles (Chanson de geste)
- Roncesvalles navarro
- Rōran no tsunobue
- The Franco-Italian Roland
- The legend of Roncevaux: adapted from "La chanson de Roland"
- The Tournament of the Lions
- 송면 / 1982 / 오딧세이-롤랑의 노래

I due esempi dimostrano che il titolo uniforme:

- realmente, non ha un criterio per la sua definizione;
- storicamente, è esistito solo all'interno di un catalogo di un Paese.

Il titolo *Le mille e una notte* è indicizzato in una forma vernacola specifica nel catalogo di ciascun Paese, mentre il titolo *Chanson de Roland* compare in forma francese in numerosi cataloghi di biblioteche europee; gli esempi di titolo variante sopra riportati dimostrano, tuttavia, che vi è disomogeneità. La *Chanson de Roland* nel corso del tempo si stabilizza, ma solo in Europa, come titolo di riferimento nell'universo letterario e repertoriale.

Perché questo comportamento dissimile nei cataloghi europei tra *Le Mille e una notte* e *Chanson de Roland*?

Si è risposto, informalmente, che il trattamento catalografico disomogeneo dipendeva da ragioni pratiche, non teoriche. Opere come *Le mille e una notte* (titolo in persiano: *آبش کی و رازہ*, "Hezār-o yek šab") non hanno il titolo uniforme in persiano in quanto il loro titolo è stato formulato originariamente in una lingua indoeuropea non usata dal catalogo delle biblioteche europee (anche in questo caso, mai motivazione fu più vaga). Per rendere più semplice ai lettori la consultazione del catalogo veniva consentito manipolare il titolo di opere anonime redatte in lingue "lontane", titolo reso, cioè tradotto (e non solo traslitterato) nella lingua d'arrivo del catalogo. Una differenza di trattamento che con la diffusione della conoscenza delle lingue e con la globalizzazione della consultazione dei cataloghi ha sempre meno forza e addirittura oggi sembra bizzarra. Antonio Panizzi aveva colto il problema quando aveva deciso di non coprire con il nuovo catalogo generale del 1841 le pubblicazioni in lingue orientali (*Bibbia* esclusa).



## **Titolo preferito e nuovo concetto di authority control**

Nell'ambiente culturale globale l'obiettivo è rispettare nelle soluzioni adottate le tradizioni culturali e linguistiche di ogni Paese e di ogni singola comunità. Il meccanismo di "riconciliazione" delle forme differenti con cui un titolo di un'opera anonima è conosciuto e identificato in un contesto globale diventa il principio per nuove modalità di condivisione delle informazioni.

La scelta del titolo preferito è una questione che è stata risolta pragmaticamente grazie alla tecnologia, con la costruzione di *cluster*.<sup>12</sup> *Cluster* è un termine tecnico, che significa un insieme di attributi utili a rappresentare un'entità, cioè un *real world object*. Il cluster contiene (o può contenere) le forme varianti di un nome e altri attributi utili a identificare l'entità. La selezione e il raggruppamento di elementi omogenei in un insieme di dati legano le forme varianti del titolo; tra queste, ciascuna agenzia bibliografica o biblioteca seleziona quella che ritiene la preferita nel contesto in cui si porranno i dati. L'affermazione dei *cluster* è il superamento della supremazia del titolo uniforme con valore assoluto in ogni contesto – principio, peraltro, lo ripetiamo, mai attuato e smentito dalla ricognizione del comportamento dei cataloghi. I due casi di studio sopracitati hanno dimostrato che il titolo uniforme è esistito solo come concetto, smentito dalla prassi catalografica; le due opere, infatti, sono state indicizzate utilizzando titoli varianti, ovvero titoli considerati funzionali al lettore e, pertanto, formulati nella lingua prevalentemente da lui usata.

A queste constatazioni si aggiunge un cambio filosofico del concetto di *authority control* in era digitale: l'elaborazione teorica e le opportunità offerte dalla tecnologia, infatti, stanno configurando in maniera diversa le modalità del controllo bibliografico. Il nuovo concetto sta sempre più evolvendo verso una forma di attività universale e collaborativa d'identificazione di entità, intese come agenti, concetti, risorse. È un'attività sulla quale convergono grandi progetti locali e internazionali, di agenzie autorevoli, come VIAF e ISNI, e collaborative, come Wikidata. L'assegnazione di identificatori diventa preponderante rispetto all'assegnazione di attributi tradizionalmente "qualificanti" l'entità come le date di nascita o la professione. La creazione di database interoperabili consente di condividere globalmente i punti d'accesso per autore, titolo, soggetto, contraddistinti da identificatori, e sganciandoli, quindi, dai contesti culturali o linguistici specifici. Il concetto base dell'UBC della necessaria omogeneità dei dati intesa come uniformità di formulazione per l'identificazione dell'entità è stato ripensato. Un'opera, infatti, proprio per essere spesso conosciuta con più *nomen* in contesti culturali, linguistici, geografici diversi, necessita di cluster in cui le forme varianti del titolo diventano più *nomen* collegati con la relazione di equivalenza; per esempio, i titoli che compaiono in manifestazioni di espressioni differenti di un'opera: *Elogio della pazzia* ed *Elogio della follia*; oppure *Novellino* e *Cento novelle antiche* o *Libro di novelle e di bel parlar gentile*. Si tratta di varianti tutte valide, utilizzabili e realmente utilizzate.

Un esempio molto chiaro è presentato da Alessandra Moi: "Un esempio tra tanti che può essere, a questo proposito, citato è l'opera conosciuta in Italia con il titolo *Dieci piccoli indiani*, di Agatha Christie. L'opera venne pubblicata nella sua primissima edizione inglese come *Ten little niggers* nel 1939 e mantenuto almeno fino agli anni Ottanta. All'edizione inglese fece seguito la prima edizione americana nel 1940 con titolo, radicalmente diverso, di *And then there were none*, sotto cui l'opera è ancora oggi conosciuta negli Stati Uniti ma a cui vennero affiancate nel corso degli

---

<sup>12</sup> <https://it.wikipedia.org/wiki/Clustering>.

anni le altre due versioni *Ten Little Indians* e *Ten little soldiers*. Si tratta evidentemente della stessa opera, a fronte però di tre diversi titoli con uguale autorevolezza, che risultano in misura minore o maggiore conosciuti e utilizzati per designare l'opera in aree territoriali e culturali ben precise".<sup>13</sup>

## Il processo di riconciliazione o clustering delle entità

Il processo di riconciliazione (o *clustering*) delle entità produce un raggruppamento delle forme varianti riconducibili alla stessa entità, siano esse formulate nella medesima lingua o in più lingue. In tutti i *clustering* dell'entità che utilizzano il meccanismo di riconciliazione è consuetudine attribuire un'identificazione all'entità riconosciuta, ovvero un identificatore che si collega ad altri identificatori assegnati alla stessa entità in domini diversi. Gli identificatori consentono di creare un puntatore a una risorsa, una sorta di gancio per richiamare quella serie di attributi che descrive l'entità. Il ruolo degli identificatori è primario.

La scelta della forma preferita del nome e la strutturazione della stringa (formulate secondo regole sintattiche note in passato solo ai catalogatori) perdono d'importanza a fronte della necessità pratica di creare canali di reperimento multipli ed equivalenti per la stessa risorsa.

Nell'ambito del controllo bibliografico universale resta la necessità di offrire una forma come risultato di una scelta nazionale, culturale o linguistica; ciò avviene tramite meccanismi di presentazione delle informazioni legati al *cluster*. Una forma può essere presentata come preferita in un dato contesto e un'altra in un contesto differente; o come target del lettore (provenienza del richiedente) che, sulla base della propria specifica esigenza di ricerca, guida la selezione della forma preferita; anche in questo caso, quindi, forma preferita nello specifico contesto della ricerca.

Il collegamento dei vari identificatori è importante strategicamente. Il meccanismo di *clustering* parte dal presupposto che tutte le forme di un nome usate nel contesto globale abbiano pari dignità; non c'è una particolare preferenza per l'una o l'altra. Il contesto d'appartenenza e la necessità d'uso (il target che richiama quel nome) definiscono di volta in volta la scelta sulla forma del nome da considerare *preferita*. Ciò è motivato dalla volontà di arricchire il dataset e di offrire al lettore più canali per raggiungere l'obiettivo: è questa la funzione di *discovery*. Si tratta di uno strumento pragmatico e funzionale per identificare, selezionare e ottenere la risorsa. Ogni attributo dell'entità si trasforma in una potenziale chiave d'accesso alla risorsa. La creazione del *cluster* di forme varianti del titolo è, dunque, passaggio fondamentale del controllo bibliografico universale: dal controllo di stringhe e punti d'accesso si passa al concetto più complesso d'identificazione di entità diverse e varianti con cui il titolo può essere formulato, nel rispetto delle variazioni linguistiche e culturali.

## Conclusioni

La tradizione della catalogazione cresce e si arricchisce con la metadattazione, nel dialogo con comunità e gruppi di lettori. Il libero riutilizzo dei dati può avvenire in contesti molto diversi da quelli originari, moltiplicando per tutti le opportunità d'accesso e di produzione di nuova co-

---

<sup>13</sup> Alessandra Moi, *Risposta 2 in Metadattazione. La catalogazione in era digitale*. Milano: Editrice Bibliografica, 2022.

noscenza. L'uso dei metadati ha aperto nuove prospettive del controllo bibliografico. L'UBC ora contempla l'interoperabilità e la flessibilità nel dialogo con le varie comunità e istituzioni della memoria registrata. Chissà cosa ci riserverà il futuro. Nel campo dei metadati e dell'authority control potremmo aspettarci sviluppi e sorprese da tecnologie alternative sul *machine learning* o sull'intelligenza artificiale, uno strumento che si preannuncia molto utile; uno strumento che nulla toglie alla valutazione del catalogatore, che resta un'attività intellettuale fondamentale del suo patrimonio genetico professionale e scientifico.

In definitiva, la scelta della forma del titolo di un'opera è stata storicamente condizionata dal contesto culturale e linguistico in cui si poneva il catalogo. Le tradizioni bibliografiche del mondo occidentale sono state considerate assolute, ma ora la dimensione globale della comunicazione cambia tutti i parametri e relativizza ogni soluzione.

## Riferimenti bibliografici

American Library Association, Library of Congress, Library Association, and Canadian Library Association. 1967. *Anglo-American Cataloging Rules*. North American text. Chicago: American Library Association.

*And then there where none*. In: *Wikipedia, the free encyclopedia*, <[https://en.wikipedia.org/wiki/And\\_Then\\_There\\_Were\\_None](https://en.wikipedia.org/wiki/And_Then_There_Were_None)>.

Audiffredi, Giovanni Battista, Gioacchino Salvioni, and Giuseppe Salvioni. 1761. *Bibliothecæ Casanatensis catalogus librorum Salvioni typis impressorum ... tomus primus (-quartus)*. Romæ: Excudebant Joachim, & Joannes Josephus Salvioni fratres, typographi pontificii vaticani in Archilyceo Romano.

Buizza, Pino. 2002. "Dai Principi di Parigi a FRBR." *Bibliotime*, 5(1). <https://www.aib.it/aib/sezioni/emr/bibtime/num-v-1/buizza.htm>.

Commissione permanente per la revisione delle regole italiane di catalogazione e Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche. 2009. *Regole italiane di catalogazione: REICAT*. Roma: ICCU.

Hyde, Thomas. 1674. *Catalogus impressorum librorum Bibliothecæ Bodleianæ in Academia Oxoniensi*. Oxonia.

IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional requirements for bibliographic records. Final report*. München: Saur, 1998. Edizione italiana: IFLA. 2000. *Requisiti funzionali per record bibliografici*. Roma: ICCU.

International Conference on Cataloguing Principles. 1961. *International Conference on Cataloguing Principles*. Paris, 9th-18th October. Report. 1963. Edited by A. H. Chaplin and Dorothy Anderson, London, Organizing Committee of the International Conference on Cataloguing Principles.

International Federation of Library Associations and Institutions. 2009. *Statement of International Cataloguing Principles 2009*. [https://www.ifla.org/wpcontent/uploads/2019/05/assets/cataloguing/icp/icp\\_2009-en.pdf](https://www.ifla.org/wpcontent/uploads/2019/05/assets/cataloguing/icp/icp_2009-en.pdf).

IFLA (International Federation of Library Associations and Institutions). 2016. *Statement of International Cataloguing Principles*. [https://www.ifla.org/files/assets/cataloguing/icp/icp\\_2016-en.pdf](https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf).

IFLA (International Federation of Library Associations and Institutions). 2017. *IFLA Library Reference Model. A Conceptual Model for Bibliographic Information*. Edited by Pat Riva, Patrick Le Boeuf, and Maja Zumer. Den Haag: IFLA. [https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla\\_lrm\\_2017-03.pdf](https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla_lrm_2017-03.pdf).

Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche. 1979. *Regole Italiane Di Catalogazione per Autori: RICA*. Roma: ICCU.

Joint Steering Committee for Development of RDA, American Library Association, Australian Committee on Cataloguing, British Library, Canadian Committee on Cataloguing, CILIP: Chartered Institute of Library and Information Professionals, Deutsche Nationalbibliothek, and Li-

brary of Congress. 2015. *Resource description and access: RDA*. Chicago: American Library Association. Edizione italiana a cura dell'ICCU. Roma. [http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/RDA Traduzione ICCU 5 Novembre REV.pdf](http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/RDA_Traduzione_ICCU_5_Novembre_REV.pdf).

Joint Steering Committee for Revision of AACR, American Library Association, Australian Committee on Cataloguing, British library, Canadian committee on Cataloguing, Library Association, Library of Congress; a cura di Michael Gorman e Paul W. Winkler. 1988. *Regole di catalogazione angloamericane*. Edizione italiana a cura di Rossella Dini e Luigi Crocetti. 1997. Milano: Editrice Bibliografica.

Smiraglia, Richard P. 2001. "Glossary." In *The nature of "A work": implications for the organization of knowledge*. Lanham, MD; London: Scarecrow.

Verona, Eva. 1959. "Literary unit versus bibliographical unit." *Libri*, 29(2):79-104.

## Conceptual models and users

Lucia Sardo<sup>(a)</sup>

a) University of Bologna - Campus of Ravenna, <https://orcid.org/0000-0001-6480-759X>

---

**Contact:** Lucia Sardo, [sardo.lucia@gmail.com](mailto:sardo.lucia@gmail.com)

**Received:** 10 May 2022; **Accepted:** 13 June 2022; **First Published:** 15 September 2022

---

### ABSTRACT

The paper presents an analysis of user tasks in the conceptual models, with an overview of the objectives and functions of the catalogue from Cutter to the International Cataloguing Principles, a critical reflection on user tasks and a proposal for a rethinking of these. The user tasks as presented in the conceptual models present critical aspects, in particular they do not bring out some important aspects concerning the users' motivations for searches and the users' competences. The presentation of some critical aspects of user tasks is followed by an initial proposal for a more complex reflection on how users use catalogues to carry out searches of different kinds.

### KEYWORDS

Conceptual models; User tasks; Cataloguing.

## I modelli concettuali e gli utenti

### ABSTRACT

La relazione presenta una analisi delle funzioni utente nei modelli concettuali, con una panoramica sugli obiettivi e le funzioni del catalogo da Cutter ai Principi Internazionali di catalogazione, una riflessione critica sulle funzioni utente e una proposta per un ripensamento di queste ultime. Le funzioni utente come presentate nei modelli concettuali presentano aspetti critici, in particolare modo non fanno emergere alcuni aspetti importanti relativi alle motivazioni delle ricerche da parte degli utenti e alle competenze degli utenti stessi. Alla presentazione di alcuni aspetti critici delle funzioni utente segue una prima proposta per una riflessione più complessa sulle modalità con cui gli utenti utilizzano i cataloghi per svolgere ricerche di diverso tipo.

### PAROLE CHIAVE

Modelli concettuali; Funzioni utente; Catalogazione.

## Premessa

La relazione intende presentare una analisi delle funzioni utente nei modelli concettuali, con una panoramica sugli obiettivi e le funzioni del catalogo da Cutter ai Principi Internazionali di catalogazione, una riflessione critica sulle funzioni utente e una proposta per un ripensamento di queste ultime.

Una premessa fondamentale è necessaria: fino a FRBR non si è mai parlato di “funzioni utente”: i teorici della catalogazione non hanno mai affrontato sistematicamente la questione della modalità con cui gli utenti effettuano le ricerche, ma si sono limitati ad affermare obiettivi e funzioni del catalogo.

Le esigenze degli utenti non sono state del tutto omesse, ma non ci sono state esplicite affermazioni, nelle regole di catalogazione, paragonabili a ciò che verrà fatto, ad esempio, in FRBR. Ciò non toglie che già con Cutter la dicotomia fra “convenience of the public” e “ease of the cataloger” sia presente e affrontata. Cutter nella prefazione all’edizione del 1904 delle Regole per un catalogo dizionario afferma infatti esplicitamente che “the convenience of the public is always to be set before the ease of the cataloger. In most cases they coincide. A plain rule without exceptions is not only easy for us to carry out, but easy for the public to understand and work by. But strict consistency in a rule and uniformity in its application sometimes lead to practices which clash with the public’s habitual way of looking at things. When these habits are general and deeply rooted, it is unwise for the cataloguer to ignore them, even if they demand a sacrifice of system and simplicity”. (Cutter 1904, 6)

L’attenzione dell’autore si focalizza sugli obiettivi e sulle funzioni del catalogo, con la convinzione di andare a definirli e individuarli sulla base delle presunte esigenze di un presunto pubblico, che a fine Ottocento mi permetto di dubitare fosse anche solo lontanamente paragonabile ai pubblici odierni.

La questione riguarda a mio avviso il fatto che questi obiettivi e funzioni sono rimasti sostanzialmente invariati nella loro essenza, anche se le formulazioni si sono via via “raffinate” e sono state meglio definite e formalizzate. La migliore definizione e formalizzazione però riguarda soprattutto la terminologia usata per riferirsi agli “oggetti di interesse” dell’utente, e non una analisi degli utenti stessi e delle loro modalità di ricerca.

La tradizione catalografica, da Cutter al XXI secolo, letteralmente, ha confermato le due funzioni del catalogo (con Domanovsky si aggiunge una terza, ma la sua riflessione rimane marginale rispetto al pensiero catalografico mainstream), localizzazione e raggruppamento.

Solo con la *Dichiarazione di principi internazionali di catalogazione* del 2009 (ICP), si è vista una modifica apparentemente radicale di queste funzioni, creando però un corto circuito logico: la terminologia utilizzata è ricavata da FRBR, cosa che non dovrebbe avvenire in quanto un modello concettuale non può essere logicamente la base di una dichiarazione di principi, quantomeno non così esplicitamente. Infatti, ogni eventuale modifica fatta al modello concettuale rende obsoleta la dichiarazione di principi.

È con la dichiarazione di principi che la tematica degli obiettivi del catalogo si intreccia con le funzioni utente, per la prima volta presenti in FRBR. Nella *Dichiarazione di principi internazionali di catalogazione* gli obiettivi del catalogo vengono declinati seguendo le funzioni utente di FRBR e utilizzando la terminologia del modello:

“Il catalogo dovrebbe essere uno strumento efficiente ed efficace che consente all’utente:

6.1 di trovare risorse bibliografiche in una raccolta come risultato di un’interrogazione basata sugli attributi o sulle relazioni delle entità [...];

6.2 di identificare una risorsa bibliografica o un agente [...];

6.3 di selezionare una risorsa bibliografica appropriata alle necessità dell’utente [...];

6.4 di acquisire o di ottenere accesso a un esemplare descritto [...];

6.5 di navigare ed esplorare:

all’interno di un catalogo, [...]

oltre il catalogo, verso altri cataloghi e in contesti non legati alle biblioteche.” (ICP 2016, 11-12)

Gli obiettivi del catalogo sono riportati al punto 6 della dichiarazione. Al punto 2, invece, sono presentati i Principi generali, dove si afferma che fra quelli elencati, “l’interesse dell’utente è il più importante, mentre i principi da 2.2 a 2.13 non sono in un ordine particolare”. (ICP 2016, 6). Con interesse dell’utente nei principi si intende che “si dovrebbe compiere ogni sforzo per rendere tutti i dati comprensibili e adatti agli utenti. La parola “utente” comprende chiunque cerchi nel catalogo e usi dati bibliografici e/o d’autorità. Le decisioni adottate nel creare le descrizioni e le forme controllate dei nomi per l’accesso dovrebbero essere prese tenendo presente l’utente” (ICP 2016, 6).

Le critiche agli obiettivi del catalogo, da Cutter in poi, si sono concentrate su questioni legate alla discrepanza fra quanto viene affermato nei principi e quanto effettivamente fanno o sono in grado di fare i cataloghi, ma soprattutto sui problemi legati alla convenienza dell’utente: come afferma Clarke, “despite the claim of user service, it seems clear from years of user studies that the traditional cataloging objectives are not as well aligned with user needs as perhaps was once thought. Users do not use library catalogs in the same way librarians assume or expect they should. Users seem to have a different mental model of how bibliographic data and links between these data are structured”. (Clarke 2015, 298).

Anche secondo Borgman gli obiettivi di Cutter non rappresentano adeguatamente le modalità di ricerca degli utenti “Cutter’s objects do not adequately represent the way people seek information. While Cutter certainly claimed to have the user in mind, his perspective and others of his time was based on a rational, positivistic approach, and not on direct study of how people formulate questions and seek information; rather, they made assumptions about the knowledge people brought to the information-seeking process” (Borgman 1996, 495); ancora più lapidario Miksa, quando definisce gli obiettivi di Cutter “simply a practical map of the entry system”. (Miksa 2012, 20).

In ogni caso, è su queste basi che la teoria catalografica ha realizzato nel corso del ventesimo secolo principi, norme e pratiche operative.

## Le “funzioni utente”

In FRBR per la prima volta si parla di “user tasks”, tradotto in italiano un po’ infelicemente con “funzioni utente”: forse sarebbe meglio utilizzare l’espressione “attività dell’utente”, ovvero ciò che fa l’utente quando cerca in un catalogo, perché più intuitiva e comprensibile.

Innanzitutto, non bisogna dimenticare che le funzioni svolte dagli utenti sono messe in relazione con i requisiti funzionali dei record bibliografici, requisiti che erano lo scopo principale del



documento, come esplicitamente conferma il titolo del rapporto: Requisiti funzionali per record bibliografici.

Riprendendo dall'originale inglese “for the purposes of this study the functional requirements for bibliographic records are defined in relation to the following generic tasks that are performed by users when searching and making use of national bibliographies and library catalogues” (FRBR 1998, 8) tradotto come “per gli scopi di questo studio i requisiti funzionali dei record bibliografici vengono definiti in relazione alle seguenti funzioni di carattere generale svolte dagli utenti nel corso della ricerca e dell'uso di bibliografie nazionali e di cataloghi di biblioteca”. (Requisiti funzionali per record bibliografici 2000, 17).

Le funzioni utente prevedono:

“utilizzo dei dati per trovare materiali che corrispondano ai criteri di ricerca definiti dall'utente [...];  
utilizzo dei dati recuperati per identificare un'entità [...];  
utilizzo dei dati per selezionare un'entità adeguata ai bisogni dell'utente [...];  
utilizzo dei dati al fine di acquisire o ottenere l'accesso all'entità descritta” (Requisiti funzionali per record bibliografici 2000, 17-18)

I modelli successivi modificano parzialmente le funzioni, le ampliano e adattano ai diversi rispettivi contesti. In FRAD (*Requisiti funzionali per i dati di autorità*), poiché gli utenti dei dati di autorità vengono definiti in senso ampio, sono previste anche le funzioni contestualizzare (“Collocare una persona, ente, opera, etc., in un contesto; chiarire la relazione fra due o più persone, enti, opere, etc.; oppure, chiarire la relazione fra una persona, ente, etc. e un nome attraverso cui quella persona, ente, etc. è conosciuta”) e giustificare (“Documentare le ragioni in base alle quali il creatore dei dati di autorità ha scelto il nome o la forma del nome su cui si basa un punto di accesso controllato”). (Requisiti funzionali per i dati di autorità 2010, 83). In FRSAD (*Functional requirements for subject authority data*) invece viene introdotta la funzione esplorare: “Explore relationships between subjects and/or their appellations (e.g., to explore relationships in order to understand the structure of a subject domain and its terminology)” (FRSAD 2010, 9), in quanto considerata importante nel contest della ricerca per soggetto.

IFLA LRM (IFLA Library reference model), il modello che ha consolidato e reso quindi obsoleti i modelli precedenti, prevede cinque funzioni utente: sinteticamente, esse sono:

“Trovare: Raccogliere informazioni su una o più risorse d'interesse tramite qualsiasi criterio di ricerca rilevante.  
Identificare: Comprendere chiaramente la natura delle risorse trovate e distinguere risorse simili.  
Selezionare: Determinare l'idoneità delle risorse trovate ed essere in grado di accettare o rifiutare risorse specifiche.  
Ottenere: Accedere al contenuto della risorsa.  
Esplorare: Scoprire risorse usando le relazioni tra di esse e dunque porre le risorse in un contesto”. (IFLA LRM 2020, 15).

In tutti i modelli si affermano esplicitamente le tipologie di utenti prese in considerazione per andare a stabilire quali siano le attività che svolgono gli utenti stessi in modo però molto generico, e inoltre nessun modello prende esplicitamente in considerazione quali potrebbero essere le esigen-

ze di coloro che interrogano i cataloghi. FRBR parla di “ampio spettro di utenze: lettori, studenti, ricercatori, personale di biblioteca, editori, distributori, venditori al dettaglio, intermediari nel settore delle informazioni, curatori di diritti sulla proprietà intellettuale, etc. (Requisiti funzionali per record bibliografici 2000, p, 17); FRAD di utenti dei dati di autorità definiti in senso ampio, in modo da comprendere “I creatori dei dati di autorità, i quali creano e curano la manutenzione dei dati di autorità. Gli utenti che utilizzano le informazioni di autorità sia attraverso l’accesso diretto ai dati di autorità, sia indirettamente, attraverso i punti di accesso controllati (forme autorizzate del nome, forme varianti del nome/rinvii, etc.) nei cataloghi, nelle bibliografie nazionali, in altre basi dati simili, etc.” (FRAD 2010, 83), FRSAD di “information professionals who create and maintain subject authority data, information professionals who create and maintain metadata, intermediaries and end users who search for information to fulfil information needs.” (FRSAD 2010, 9). IFLA LRM riprende sostanzialmente FRBR parlando di “lettori, studenti, ricercatori e altri tipi di utenti finali, da bibliotecari, da altri attori della catena informativa compresi editori, distributori, fornitori ecc.” (IFLA LRM 2020, 15).

Le principali critiche rivolte a queste definizioni riguardano soprattutto la ristrettezza della tipologia di istituzioni di provenienza dei redattori dei diversi modelli concettuali (anche se sicuramente esperti della disciplina); l’esplicitazione delle eventuali fonti utilizzate per giungere a queste definizioni; e quella che già è stata chiamata l’eterogeneità dei fini: quello che era un documento che voleva presentare i requisiti funzionali dei record bibliografici prodotti dalle agenzie catalografiche nazionali è diventato il modello di riferimento per l’universo bibliografico.

Fra le molte voci critiche, quella di Hoffman è particolarmente dura: “Yet, the chair of the study admits, ‘It did not involve studies of how actual users approach and make use of bibliographic records.’ How can FRBR call itself user-focused when users were not studied? This is significant, because this model is the basis of the new cataloging rules, Resource Description and Access (RDA). In this way, the idea of being user-focused is perpetuated. Cataloging standards claim to focus on users but are not based on an understanding of users’ needs that originates from empirical studies of real users.” (Hoffman 2009, 634; la citazione interna è da Madison 2000). La critica di Hoffman ha suscitato reazioni, ma personalmente ritengo che le critiche avanzate non siano né fuori luogo, né trascurabili. Anche Coyle sottolinea questo aspetto problematico di FRBR, in modo ancor più forte, affermando che “Non c’è una analisi degli utenti; non si fa menzione di come sia cambiata la base degli utenti della biblioteca; non si fa menzione di bambini, anziani o disabili. Invece, secondo me, il Rapporto conclusivo si legge come uno studio di catalogatori per catalogatori. Ancora più nello specifico è uno studio di catalogatori di grandi istituzioni di ricerca”. (Coyle 2017, 114). La critica di Coyle va a toccare anche la questione degli obiettivi del catalogo, quando afferma che “gli obiettivi del catalogo non sono espressi in termini di soddisfazione delle domande degli utenti, ma delineano quali domande possono essere fatte e gestire le aspettative per quali risposte saranno ottenute.” (Coyle 2017, 58)

Quanto detto sopra da Coyle tocca un aspetto a mio avviso fondamentale, ovvero quello della non esistenza di un utente platonico della biblioteca; questa banalissima considerazione vuole far riflettere sul fatto che la convenienza di un utente può non essere quella di un altro, e che in momenti e per finalità diverse anche lo stesso utente cerca e si comporta in modo differente rispetto al catalogo, e quindi la sua “convenienza” può avere livelli di variabilità notevoli. Un altro aspetto chiave è quello delle finalità per cui le ricerche vengono effettuate.

Insomma, i modelli ci dicono un generico “chi” fa le ricerche e quindi “come” le fa, ma non perché questo utente stia facendo quello che fa. Non è una questione secondaria. In queste affermazioni sugli utenti manca anche una valutazione della situazione di partenza rispetto a quanto viene ricercato.

Gli utenti cercano davvero come i bibliotecari immaginano che cerchino, e si comportano come i bibliotecari pensano si debbano comportare, oppure i radicali cambiamenti avvenuti negli ultimi decenni hanno cambiato le modalità di ricerca degli utenti (dando per scontato, e scontato non lo è, che prima lo facessero come immaginavano i bibliotecari)? Poiché le ricerche in un catalogo sono un sottoinsieme di un information seeking (rispetto al quale sono numerosi gli studi e le teorie), sarebbe importante anche avere delle risposte alla più generica domanda, che è: in quali situazioni di bisogno informativo gli utenti utilizzano il catalogo? cosa fanno davvero gli utenti quando cercano in un catalogo (che è uno degli ormai molti strumenti a disposizione per effettuare ricerche su risorse documentali)? (Dando per scontato che gli utenti li usino, i cataloghi, perché è questo il punto dolente: gli utenti li usano sempre meno.

Sono aumentate le categorie potenziali di utenti che si avvicinano a un catalogo per le più disparate esigenze, o così ci auguriamo che sia, e sono cambiati i parametri di appartenenza e le categorie comuni di riferimento intellettuale; si è avuta una progressiva frammentazione dal punto di vista culturale e il riconoscersi in una o più categorie culturali sta diventando la norma sociale piuttosto che l'eccezione. Studi sociologici confermano questo senso di straniamento e di difficile collocazione all'interno di gruppi con parametri, confini culturali, la difficoltà di sentirsi parte di un gruppo sociale, e un progressivo isolamento dell'individuo, mentre allo stesso tempo gli effetti della globalizzazione facilitano l'avvicinamento e l'approccio con realtà sociali e culturali radicalmente diverse da quelle di origine degli individui, creando un senso di straniamento. In questo senso, diventa difficile categorizzare con chiarezza e pensare di poter prevedere il comportamento e le necessità degli utenti. Dati questi elementi le ricerche sulle modalità di comportamento degli utenti dovrebbero essere uno dei punti chiave della futura ricerca in ambito biblioteconomico tout court, non solo per quanto riguarda l'uso del catalogo, ma anche della biblioteca, e soprattutto la sua percezione.

Sempre Coyle ricorda che “creiamo record che trattano tutti i tipi di risorse nello stesso modo, e per una sola definizione di ‘utente’. Ignoriamo o sottovalutiamo molte delle caratteristiche che sono importanti per gli utenti”. (Coyle 2017, 13).

Andrebbero indagate quindi non solo le modalità ma anche le finalità per cui gli utenti cercano un catalogo, il tutto analizzato in base alla tipologia degli utenti stessi, cercando di avere, almeno in una prima fase di analisi, una categorizzazione di massima che aiuti nella modellizzazione.

## **Per un ripensamento delle funzioni utente**

Un ripensamento radicale delle funzioni utente in base a livelli e modalità potrebbe portare a una migliore definizione delle funzioni del catalogo, ad avere il coraggio di ammettere che un catalogo unico per tutti gli utenti non sia una soluzione praticabile per soddisfarne veramente le esigenze, e che forse si dovrebbero realizzare strumenti di ricerca diversi per utenti diversi con finalità di ricerca diverse, con una base descrittiva e di indicizzazione comune valida per sostanzialmente la maggior parte delle situazioni.

Le modalità di ricerca degli utenti sono state studiate da molteplici punti di vista. Come riportato da Fattahi: “le fasi principali [dei processi di ricerca informativa] che sono comuni alla maggior parte dei modelli, e che implicano attività cognitive da parte dell’utente, includono: 1) Arrivare a un problema: divario informazione/conoscenza, 2) Identificare l’informazione necessaria, 3) Sottoporre una richiesta, 4) Scorrere liste e analizzare i risultati, 5) Riconoscere e distinguere, e 6) Apprendimento e meta-riconoscimento”. (Fattahi 2010, 32).

I fattori da tenere in considerazione per studiare e analizzare le modalità comportamentali degli utenti quando effettuano ricerche sono molteplici: gli studi sui comportamenti di ricerca sono ormai un filone di studi a sé stante nell’ambito della Scienza dell’informazione; l’attività di ricerca è un processo condizionato dalla struttura cognitiva dell’utente, dal contesto, dai processi mentali, dalle abitudini e da molteplici fattori personali, sociologici e psicologici, e inoltre non è un processo statico, ma evolve con l’evolversi della conoscenza dell’utente stesso; infine, le modalità di ricerca possono parzialmente variare anche se riferite allo stesso utente.<sup>1</sup>

Fatta questa sintetica premessa si può affermare che quanto presentato da IFLA LRM, pur nella sua chiarezza e indubitabilità, è sicuramente semplicistico e riduttivo e non tiene conto delle molteplici situazioni che spingono un utente a effettuare una ricerca.

Le funzioni utente previste in IFLA LRM (Trovare, Identificare, Selezionare, Ottenere, Esplorare) ci dicono solo quali sono le macro funzioni, ovvero per quali motivi generici gli utenti utilizzano un catalogo (e volendo non solo i cataloghi, per inciso).

Le ricerche però non vengono fatte per le stesse finalità, o per gli stessi motivi: limitando al minimo le possibilità si potrebbero ipotizzare le seguenti situazioni: informazione di base (questa prima situazione – già prima dell’avvento della rete e dei motori di ricerca più legata al servizio di reference, forse, che al catalogo – sta diventando marginale se non residuale in un mondo dove una percentuale altissima di popolazione ha uno smartphone connesso a Internet. La questione di come fare ricerca diventa quindi un problema di *information literacy* piuttosto che strettamente catalografico o di reference); informazione più approfondita; studio/ricerca/apprendimento; divertimento/intrattenimento.

Non può non essere fatta la distinzione fra le finalità di studio/ricerca e quelle di altro tipo, perché le esigenze degli studiosi sono radicalmente diverse, ad esempio, da quelle di coloro che cercano informazioni (più o meno strutturate), o di coloro che utilizzano il catalogo e le risorse di una biblioteca per puro intrattenimento personale; ed è anche evidente che la stessa persona può utilizzare il catalogo o altri strumenti per tutte e tre le finalità, in base alle necessità del momento.

Thomas Mann, nella sua risposta al report redatto da Calhoun per la Library of Congress, (Calhoun 2006) criticando alcuni assunti della studiosa, e in particolar modo il non aver compreso la distinzione fra i due livelli di ricerca (che potremmo anche chiamare finalità della ricerca) afferma che “naturalmente studiosi e “persone che cercano un’informazione” si sovrappongono: del resto, le stesse persone possono avere scopi di ricerca per un certo progetto e in altri campi possono aver bisogno solo di informazioni rapide. Ma le biblioteche di ricerca hanno la responsabilità specifica di soddisfare le loro esigenze di studiosi, anche perché le alternative di Google, Amazon o AltaVista, che sono ottime risorse per trovare rapidamente “qualcosa” su un argomento,

---

<sup>1</sup> Volutamente sono tralasciate le situazioni di disabilità di diverso tipo, pur con la consapevolezza che anche per queste situazioni andrebbero previsti filtri e possibilità di ricerca mirate.

non rispondono adeguatamente, invece, alle loro esigenze di ricerca, in tante specifiche (e molto importanti) maniere”. (Mann 2006, 191)

Secondo Mann, la ricerca effettuata da studiosi ha caratteristiche peculiari, che fanno sì che le differenze fra le due modalità siano radicali: “1) gli studiosi desiderano un quadro complessivo il più possibile chiaro ed esauriente di tutte le fonti pertinenti; 2) sono particolarmente attenti a non lasciarsi sfuggire fonti che siano di particolare importanza o significative nel loro campo di ricerca, o considerate di riferimento; 3) non vogliono duplicare, senza necessità, ricerche precedenti; 4) desiderano particolarmente essere al corrente di collegamenti con altre discipline che possono riguardare il loro lavoro; 5) desiderano trovare i libri recenti su un argomento insieme a quelli anteriori, così che le opere nuove possano essere valutate nel contesto di quelle precedenti, non solo di quelle (molte meno) che sono al momento in commercio (gli specialisti desiderano anche trovare i libri in inglese insieme a libri significativi in altre lingue, per avere facilmente un quadro della letteratura su un argomento a livello internazionale); 6) apprezzano particolarmente i sistemi che gli permettono di riconoscere fonti molto rilevanti che non potrebbero individuare in anticipo con parole chiave da inserire con operatori booleani in una finestrella di ricerca; 7) anche se sono ben consapevoli dell’esigenza di una ricerca prolungata e su più fonti, oltre la “prima schermata” di una ricerca in Internet, desiderano evitare di scorrere lunghissime liste, di qualsiasi genere, in cui un po’ di materiale pertinente sia sepolto fra cumuli di cose inutili selezionate in modo inadeguato perché le parole chiave “giuste” compaiono in contesti concettuali “sbagliati” (Mann 2006, 191). Esiste un’altra modalità di ricerca, quella di chi cerca qualcosa per “intrattenimento”. È una tipologia ricerca che può essere solo parzialmente assimilata alle due precedenti, e che può inoltre avere diversi livelli di esplicitazione dei bisogni. In situazioni di questo tipo non si sta cercando una informazione rapida o approfondita e neppure si sta eseguendo una ricerca mirata per finalità di studio o ricerca: questa è quella ricerca che potremmo definire tipica dell’utente di una biblioteca pubblica.

Un’altra importante categorizzazione da fare è rappresentata dalla competenza dell’utente rispetto al tema cercato. Coloro che cercano possono essere distinti in diversi livelli di competenza rispetto alla ricerca che stanno effettuando. Questo livello varia non solo da persona a persona, ma da ricerca a ricerca, e si può modificare nel tempo, anzi è auspicabile che ciò avvenga. Per semplificare si potrebbero pensare quantomeno tre livelli per gli adulti, a cui potremmo aggiungere livelli simili, ma tarati diversamente, per bambini e per adolescenti. I tre livelli sono: principiante; intermedio; avanzato.

Le situazioni di partenza della ricerca, poi, sono (anche in questo caso si sta semplificando, poiché lo spettro delle possibilità è più ampio e articolato):

si ha una idea di cosa si sta cercando (si dispone di dati bibliografici);

si ha una idea dell’argomento che si sta cercando, ma non di uno specifico o più specifici documenti (ovvero non si dispone di dati bibliografici); in altre parole non si ha una idea di cosa si sta cercando, ma di ciò che interessa sapere.

Queste situazioni vengono esemplificate da Mike Sweet come: known item search; exploratory search.<sup>2</sup> Nel primo caso, come riportato da Bianchini, abbiamo una “ricerca consapevole o incre-

---

<sup>2</sup> [http://www.casalini.it/retreat/2012\\_docs/sweet.pdf](http://www.casalini.it/retreat/2012_docs/sweet.pdf).

mentale, nella quale si ha una specifica risorsa in mente (un articolo, una rivista, un libro, ecc.) si ha familiarità con le parole chiave e i termini della ricerca e si sa da dove partire (sito web, database, ecc.)”. (Bianchini 2012, 304). Nel secondo invece abbiamo il caso della ricerca esplorativa (exploratory search), “nella quale non si hanno preferenze relative al tema, né si ha familiarità con le parole chiave. Per iniziare questa ricerca si ha quasi sempre bisogno di aiuto e lo scopo è ottenere informazioni generali”. (Bianchini 2012, 304) Le due situazioni di ricerca sono invece così esemplificate da Petrucciani: “le indagini sull’uso dei cataloghi di biblioteca hanno però individuato concordemente una situazione di ricerca dominante, quella nella quale l’utente ha già selezionato, o comunque conosce, l’oggetto particolare che desidera ed è in grado, almeno in linea di principio, di individuarlo fra gli altri. L’esito della ricerca, nella situazione che stiamo esaminando, è il reperimento di una particolare copia, e in qualche caso può essere proprio essa l’oggetto specificamente desiderato, per esempio per chi sia interessato a una nota di possesso o a varianti tipografiche. Ma nella massima parte dei casi l’utente, anche quando ha già utilizzato e quindi conosce una particolare copia, non è interessato ad essa più che a qualunque altra: l’oggetto della sua ricerca può essere, quindi, la pubblicazione in generale. Allo stesso modo, egli può essere a conoscenza dell’esistenza e dei caratteri di una o più edizioni di un’opera, ma non desiderare proprio quella o quelle in particolare. L’utente può anzi preferire il ritrovamento di una edizione diversa, per esempio più aggiornata. È certamente frequente, infine, che l’utente non conosca, prima della ricerca, alcuna particolare edizione dell’opera che lo interessa, di cui può avere avuto notizia senza dettagli bibliografici.

Alla prima situazione di ricerca si contrappone in genere soprattutto la ricerca di opere su un particolare argomento. Anche qui l’esito della ricerca consiste nella localizzazione e nel reperimento di particolari oggetti fisici, ma essi vengono selezionati fra tutti quelli, segnalati dal catalogo, che condividono certe caratteristiche, le quali sono in generale insufficienti a individuarli singolarmente, ma consentono di prevederne l’interesse. Nella ricerca di opere su un argomento possiamo includere le ricerche di opere che rispondono a caratteristiche normalmente segnalate nei cataloghi semantici anche se di tipo diverso: forma, genere, destinazione.

Queste diverse ricerche sono assimilabili fra loro in contrapposizione a quelle di un’opera o pubblicazione particolare, e possiamo perciò designarle complessivamente con l’espressione «seconda situazione di ricerca». Fra esse la ricerca per argomento è quella quantitativamente più rilevante e più facilmente individuabile”. (Petrucciani 1984, 4-5)

Considerando queste possibili situazioni si arriverebbero a definire le attività dell’utente in modo meno bidimensionale, e con una maggiore consapevolezza della complessità delle potenziali situazioni di ricerca in un catalogo. Le combinazioni possibili sono molto alte anche solo con le semplificazioni sopra presentate e senza tenere conto di ulteriori fattori di tipo sociale e culturale che potrebbero ulteriormente arricchire la matrice. Sarebbe quindi opportuno valutare quali siano effettivamente rilevanti in quale contesto bibliotecario e quali siano quelle a cui dovrebbe rispondere un catalogo, con quali dati.

Per fare alcuni esempi, possiamo supporre che per ricerche finalizzate all’intrattenimento la navigazione possa avere una discreta importanza solo in alcuni casi. Se sto cercando una risorsa per avere qualcosa da leggere o da guardare, non è detto, anzi è improbabile, che sia mio interesse navigare nel catalogo o all’esterno (ed è invece possibile o probabile che all’esterno abbia già “navigato”, per arrivare a quella determinata selezione); invece, nel caso di una ricerca più strutturata,

quale “sto cercando un buon romanzo giallo da leggere in vacanza”, la navigazione può essere utile per farmi selezionare ciò che è di mio interesse. Per queste ricerche i cataloghi non sono, se non in rari casi, in grado di fornirmi una risposta, e soprattutto sono carenti di valutazioni di tipo qualitativo, che possono sì essere fuorvianti, ma fornirmi comunque una idea di base da cui partire. Nel caso esemplificato, una forma diversa di presentazione dei dati catalografici potrebbe essere adatta per le biblioteche pubbliche. Nel caso di opere letterarie è molto probabile avere diverse espressioni e manifestazioni della stessa opera, e il catalogo deve poter esplicitare il maggior numero di relazioni possibili fra queste entità, potenziando l’offerta con una integrazione di dati di tipo qualitativo per facilitare le operazioni di selezione delle risorse, con una interfaccia che faciliti in modo intuitivo alcune ricerche di tipo formale e contenutistico. Esempi di cataloghi simili sono stati pensati e realizzati, ma ancora molta strada si potrebbe fare per renderli più funzionali ed efficaci.

Le attività di studio e ricerca possono variare molto in base ai livelli di partenza sopra elencati, e questo dovrebbe essere tenuto in considerazione nel momento in cui si realizzano strumenti catalografici. I documenti che maggiormente interessano le attività di studio e ricerca è probabile abbiano poche espressioni e manifestazioni della stessa opera, ma molti aggregati; quindi è necessario raggiungere livelli molto alti di analicità nella descrizione delle risorse, avere una indicizzazione semantica spinta e una organizzazione classificata dei materiali pensata ad hoc per i particolari ambiti disciplinari di riferimento, e infine prevedere percorsi alternativi per facilitare le ricerche multidisciplinari e il reperimento di documenti “marginali” all’ambito di interesse specifico ma potenzialmente interessanti. È in queste situazioni di ricerca che valutazioni di tipo qualitativo andrebbero a incrementare il soddisfacimento dei bisogni degli utenti.

Nel caso in cui si effettui una ricerca per cosa non nota va ricordato che quando un utente avvia una ricerca in un sistema informativo, “la formula in base a uno schema mentale, determinato da fattori quali le sue conoscenze linguistiche, la familiarità con la disciplina, l’esperienza accumulata in precedenti ricerche, la conoscenza diretta di uno o più dei documenti desiderati, nonché dalle finalità della ricerca. Le corrispondenze fra lo schema mentale dell’utente e quello applicato dalla risorsa e le loro interrelazioni sono spesso difficili da identificare e da rendere esplicite” (Weston e Sardo 2017, 11).

I fattori possono anche essere di ordine psicologico e personale, legati al momento specifico e alla situazione nella quale si effettua la ricerca, né vanno dimenticati quelli legati allo strumento che viene interrogato per fare la ricerca; tutti questi elementi vanno a incidere sulle modalità di ricerca degli utenti, sulla loro capacità di interagire col sistema e sul grado di soddisfazione o di frustrazione finale. Questa è una delle criticità insormontabili dei cataloghi per soggetto o classificati; la possibilità di avere strumenti personalizzabili in base alla tipologia di ricerca che l’utente vuole effettuare sarebbe un grande vantaggio per favorire il recupero di documenti e informazioni pertinenti, per aiutare nella scoperta di documenti non noti ma potenzialmente interessanti e per favorire il miglior utilizzo delle risorse messe a disposizione dalla biblioteca; le potenzialità offerte dal *machine learning* e dagli algoritmi potrebbero essere ulteriormente sfruttate per queste finalità.

## Conclusioni

Le cosiddette funzioni utente, così come espresse nei modelli concettuali, e di conseguenza riprese nella Dichiarazione di principi internazionali di catalogazione, possono essere considerate valide, ma sono sicuramente riduttive e non sono in grado di esprimere compiutamente la complessità delle situazioni di partenza che spingono gli utenti a effettuare ricerche nei cataloghi. Allo stesso tempo, va sottolineato come questa categorizzazione presupponga un utente in grado di capire e quindi utilizzare al meglio la complessità delle strutture catalografiche realizzate. Uno studio fatto sulle parole di ricerca utilizzate dagli utenti del discovery tool dell'Alma Mater Studiorum Università di Bologna ha portato alla sostanziale conclusione, ad esempio, che lo strumento viene utilizzato principalmente per trovare documenti di cui si conosce già l'esistenza con lo scopo di raggiungere il full text della risorsa.

Per un ripensamento delle funzioni utente sarebbe però necessario intraprendere ulteriori studi sull'interazione degli utenti con i cataloghi, e soprattutto sarebbe di fondamentale importanza riuscire a individuare potenziali utenti che non li utilizzano e cercare di capire perché lo strumento non venga utilizzato.

Ma prima ancora di domandarci cosa facciano gli utenti quando interagiscono con i cataloghi, dovremmo chiederci cosa vogliamo che i cataloghi facciano in un ecosistema digitale sempre più ricco di modalità alternative di ricerca e di reperimento di risorse nonché di risorse non indicizzate nei cataloghi stessi, perché a mio avviso sta mancando una riflessione seria e molto laica sulla natura e sulle funzioni dei cataloghi agli inizi del XXI secolo.



## Riferimenti bibliografici

Bianchini, Carlo. 2012. “Dagli OPAC ai library linked data: come cambiano le risposte ai bisogni degli utenti”. *AIB Studi* 52 (3): 303-23, <http://dx.doi.org/10.2426/aibstudi-8597>.

Borgman, Christine. 1996. “Why are Online catalogs still hard to use?”. *Journal of the American Society for Information Science* 47 (7): 493-505.

Calhoun, Karen. 2006. *The changing nature of the catalog and its integration with other discovery tools*, prepared for the Library of Congress by Karen Calhoun, <https://www.loc.gov/catdir/calhoun-report-final.pdf>.

Clarke, Rachel Ivy. 2015. “Breaking records: the history of bibliographic records and their influence in conceptualizing bibliographic data”. *Cataloging & Classification Quarterly* 53 (3-4): 286-302, <http://dx.doi.org/10.1080/01639374.2014.960988>.

Coyle, Karen. 2017. *FRBR: prima e dopo. Un esame dei modelli bibliografici*, traduzione di Lucia Sardo. Roma: Associazione italiana biblioteche.

Cutter, Charles Ammi. 1904. *Rules for a Dictionary catalog*, 4<sup>th</sup> ed. Washington: Government printing office.

Fattahi, Rahmatollah. 2010. *From information to knowledge: superworks and the challenges in the organization and representation of the bibliographic universe = Dall'informazione alla conoscenza: le super-opere e le sfide dell'organizzazione e rappresentazione dell'universo bibliografico*, Fiesole (Firenze): Casalini Libri. Traduzione di Maria Letizia Fabbrini. <https://www.torrossa.com/resources/an/2413720>.

Hoffman, Gretchen L. 2009. “Meeting users’ needs in cataloging: what is the right thing to do?”, *Cataloging & Classification Quarterly* 47 (7): 631-41, <http://dx.doi.org/10.1080/01639370903111999>.

IFLA (International Federation of Library Associations and Institutions). 2000. *Requisiti funzionali per record bibliografici. Rapporto conclusivo*. Edizione italiana a cura dell’Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche. Roma: ICCU.

IFLA (International Federation of Library Associations and Institutions). 2020. *IFLA Library Reference Model. Un modello concettuale per le informazioni bibliografiche*. Pat Riva, Patrick Le Boeuf, and Maja Zumer. Edizione italiana a cura dell’Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche. Roma: ICCU.

IFLA Cataloguing section and IFLA meeting of experts on an International cataloguing code. 2016. *Dichiarazione di principi internazionali di catalogazione (ICP)*. Edizione 2016 con piccole correzioni 2017 di Agnese Galeffi (chair), María Violeta Bertolini, Robert L. Bothmann, Elena Escolano Rodríguez, and Dorothy McGarry, traduzione italiana, a cura del Gruppo di lavoro tecnico per la traduzione degli ICP 2016, [https://www.ifla.org/files/assets/cataloguing/icp/icp\\_2016-it.pdf](https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-it.pdf).

IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). 2010. *Requisiti funzionali per i dati di autorità. Un modello concettuale*. A cura di Glenn E. Patton. Edizione italiana a cura dell’Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche. Roma: ICCU

IFLA Working Group on the Functional Requirements for Subject Authority Records (FRSAR). 2010. *Functional Requirements for Subject Authority Data (FRSAD). A conceptual model*, editors Marcia Lei Zeng, Maja Žumer, Athena Salaba; Approved by the Standing Committee of the IFLA Section on Classification and Indexing, June 2010, <https://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>.

Madison, Olivia M. A. 2000. "The IFLA Functional Requirements for Bibliographic Records: International Standards for Universal Bibliographic Control". *Library Resources & Technical Services* 44 (3): 153–59.

Mann, Thomas. 2006. "Il catalogo e gli altri strumenti di ricerca: un punto di vista dalla Library of Congress". *Bollettino AIB*, 46 (3): 186-206, <http://bollettino.aib.it/article/view/5154/4921>.

Miksa, Francis. 2012. "The legacy of the library catalogue for the present". *Library Trends*, 61 (1): 7-34, <http://dx.doi.org/10.1353/lib.2012.0023>.

Petruciani, Alberto. 1984, *Funzione e struttura del catalogo per autore*. Firenze: Giunta regionale toscana, La Nuova Italia Editrice.

Weston, Paul Gabriele, e Lucia Sardo. 2017. *Metadati*. Roma: Associazione italiana biblioteche.

## A new edition of the “Guida al sistema italiano di indicizzazione per soggetto”\*

Anna Lucarelli<sup>(a)</sup>

a) National Central Library of Florence

Contact: Anna Lucarelli, [anna.lucarelli@cultura.gov.it](mailto:anna.lucarelli@cultura.gov.it)

Received: 20 June 2022; Accepted: 5 July 2022; First Published: 15 September 2022

### ABSTRACT

This contribution reproduces the text of the speech that was held in the framework of the presentation of the second edition of the volume *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto*, that took place in Florence on June 7th, 2022. As happened in these years with the subject indexing tools of important national libraries, the National Central Library of Florence, too, has updated the *Nuovo soggettario*, that saw the light in 2007, today employed by the Italian National Bibliography (BNI) and by most libraries of the National Library System (SBN). The intervention recalls the context factors that led to the project and to the elaboration of the second edition of the *Guida*, fully revised and updated: the evolution of the international scene related to indexing (new standards, new conceptual models), the semantic web and the linked open data; the development of the *Nuovo soggettario* system as a whole (increase of terminology, connections with equivalents in other languages, with resources of archives and museums, availability of the Thesaurus in exchange formats and standard protocols, etc.). The second edition's novelties are then illustrated: a better organized exposure of the standards, more functional to their use, a totally new first part dedicated to the principles and to the general aspects of the indexing, the confrontation with IFLA LRM, the references to the indexing of non-bibliographical resources (figurative works, photographs) of ancient works, etc. Finally, it ends by stressing the importance that the cooperation with other partner institutions and with ICCU has had in this work, and shall continue to have in the future, for a greater and greater sharing of the semantic data on behalf of the Italian libraries.

### KEYWORDS

Nuovo soggettario; Thesauri; Subject indexing; National Central Library of Florence.

## Una nuova edizione della “Guida al sistema italiano di indicizzazione per soggetto”

### ABSTRACT

Questo contributo riproduce il testo dell'intervento tenuto nella Giornata di presentazione della seconda edizione del volume *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto*, svoltasi a Firenze il 7 giugno 2022. Come avvenuto in questi anni per gli strumenti di indicizzazione per soggetto di importanti biblioteche nazionali, anche la Biblioteca nazionale centrale di Firenze ha aggiornato il *Nuovo soggettario*, che ha visto la luce nel 2007 ed è impiegato oggi dalla Bibliografia nazionale italiana (BNI) e dalla maggior parte delle biblioteche del Servizio bibliotecario nazionale (SBN). L'intervento richiama i fattori di contesto che hanno portato alla progettazione e all'elaborazione della seconda edizione della *Guida*, interamente rivista e aggiornata: l'evoluzione del panorama internazionale relativo all'indicizzazione (nuovi standard, nuovi modelli concettuali); il web semantico e i linked open data; lo sviluppo del sistema *Nuovo soggettario* nel suo complesso (incremento della terminologia, collegamenti con equivalenti in altre lingue, con risorse di archivi e musei, disponibilità del Thesaurus in formati di scambio e protocolli standard, ecc.). Vengono poi illustrate le principali novità della seconda edizione: un'esposizione delle norme meglio organizzata e più funzionale alla loro applicazione, una prima parte interamente nuova dedicata ai principi e agli aspetti generali dell'indicizzazione, il raffronto con IFLA LRM, i riferimenti all'indicizzazione di risorse non bibliografiche (opere figurative, fotografie), di opere antiche, ecc. Si conclude sottolineando l'importanza che in questo lavoro ha avuto – e dovrà continuare ad avere anche in futuro – la cooperazione con le istituzioni partner e con l'ICCU, per una sempre maggiore condivisione dei dati semantici da parte delle biblioteche italiane.

### PAROLE CHIAVE

Nuovo soggettario; Thesauri; Indicizzazione; Biblioteca nazionale centrale di Firenze.

\* Testo dell'intervento tenuto alla Giornata di presentazione del volume *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto* (seconda edizione interamente rivista e aggiornata), svoltasi presso la Sala Galileo della Biblioteca nazionale centrale di Firenze il 7 giugno 2022.

Per la Biblioteca nazionale centrale di Firenze (BNCF), per il nostro gruppo di lavoro, oggi è un giorno importante quasi quanto l'8 febbraio del 2007, quando del *Nuovo soggettario* venne presentata la prima edizione.

Parlare di un nuovo libro è sempre un momento particolare. In questo caso lo è ancora di più perché siamo davanti a uno strumento di lavoro finalizzato a rendere efficace la ricerca, l'accesso dei cittadini all'informazione, e quindi a migliorare il servizio pubblico che le nostre biblioteche, in quanto tali, devono principalmente rendere: descrivere e mettere a disposizione i prodotti editoriali, analogici e digitali, che hanno il compito di tutelare e, al tempo stesso, di far conoscere. Sappiamo quanto la rivoluzione del web e l'enorme quantità di informazioni disponibili rendono ancora più necessario disporre di strumenti di ricerca efficaci. I motori di ricerca, per offrire risposte mirate, elaborano algoritmi sempre più complessi e intelligenti. Con la nascita del web semantico e dei linked open data si perseguono metodi di codifica dei dati e percorsi di ricerca razionalmente fondati e utilizzabili dalle macchine.

In questo quadro, l'allestimento di strumenti di indicizzazione basati su processi intellettuali non ha perso la sua importanza. L'indicizzazione semantica praticata tradizionalmente in biblioteca ha una forza che le deriva dall'adesione alle strutture profonde del linguaggio e della comunicazione, ma può costituire anche un supporto per sistemi e procedure nuove, come quelle di tipo automatico.

Con questi temi complessi, con quelli che riguardano gli sviluppi dei nostri cataloghi elettronici, si sono confrontate negli ultimi anni tutte le Biblioteche nazionali del mondo, cercando di progettare strumenti adeguati e aderenti agli standard, di sperimentare forme di indicizzazione compatibili con le mutate esigenze da parte dell'utenza e con quelle delle stesse istituzioni spesso gravate, non solo in Italia, da una continua riduzione delle proprie risorse umane.

Mi piace ricordare che negli ultimi cinque anni, solo per fare esempi noti, sono state pubblicate nuove edizioni degli strumenti di indicizzazione per soggetto nazionali da parte della Deutsche Nationalbibliothek (nel 2017 è uscita la quarta edizione aggiornata di *Regeln für die Schlagwortkatalogisierung*<sup>1</sup>, con innovazioni ulteriori nel 2019), da parte della Bibliothèque nationale de France (a partire dal 2019 con la cosiddetta *Réforme de Rameau*<sup>2</sup>); da parte della Biblioteca Nacional de España (sempre del 2019 è la seconda edizione aggiornata del *Manual de indización de Encabezamientos de Materia*<sup>3</sup>). Analoghi aggiornamenti riguardano i vocabolari controllati e i dati di autorità terminologici che queste biblioteche sviluppano con aggiornamenti continui, come avviene, solo per fare un esempio, con le note *Library of Congress subject headings* (LCSH)<sup>4</sup>.

Tutto ciò non significa che queste stesse Biblioteche (in primis quella tedesca ma, ad esempio, anche quella finlandese) non abbiano sperimentato forme nuove e diverse di indicizzazione.

Nell'organizzare questa giornata abbiamo voluto privilegiare un approccio concreto; le relazioni sono tutte presentate da bibliotecari che hanno esperienza del *Nuovo soggettario* o che hanno dato un contributo intellettuale o pratico al lavoro svolto in questi anni in seno alla BNCF. Non sono gli unici ad averlo fatto, ma sono senz'altro fra i più rappresentativi.

<sup>1</sup> <https://d-nb.info/1126513032/34>.

<sup>2</sup> <https://rameau.bnf.fr/syntaxe>.

<sup>3</sup> <http://www.bne.es/es/Micrositios/Publicaciones/MEMBNE/>.

<sup>4</sup> <https://id.loc.gov/authorities/subjects.html>.

Come è noto, il *Nuovo soggettario* è un “sistema”<sup>5</sup>, è il linguaggio di indicizzazione ufficiale impiegato dalla Bibliografia nazionale italiana (BNI) e dalla maggior parte delle biblioteche del Servizio bibliotecario nazionale (SBN).

È utilizzabile per indicizzare il contenuto di opere di ogni campo disciplinare e di qualsiasi natura (testi, immagini, suoni, ecc.), su supporti vari (cartacei, digitali, ecc.), nell’ambito di biblioteche, archivi, mediateche e altri istituti dell’eredità culturale.

Realizzare uno strumento che possa avere una valenza per ogni tipo di biblioteche è stata sin dall’inizio la nostra scommessa, per biblioteche con utenze diverse, con tecnologie diverse di recupero dell’informazione.

Il volume che presentiamo – e che, un po’ sbrigativamente, chiamiamo *Guida* – è il prodotto di un lavoro che si è concluso negli ultimi mesi del 2021. Ha visto la luce in formato digitale, sul sito della BNCF, all’inizio del 2022 ed è fruibile in modalità totalmente gratuita<sup>6</sup>. Come già per la prima edizione pubblicata dall’Editrice Bibliografica nel 2007, contiene le norme inerenti al sistema, il metodo per l’analisi concettuale dei documenti e per l’elaborazione di accessi per soggetto di tipo verbale.



<sup>5</sup> <https://thes.bncf.firenze.sbn.it/index.html>.

<sup>6</sup> [https://www.bncf.firenze.sbn.it/wp-content/uploads/2020/01/Nuovo-soggettario\\_Guida.pdf](https://www.bncf.firenze.sbn.it/wp-content/uploads/2020/01/Nuovo-soggettario_Guida.pdf).

Contiene anche le regole che presiedono alla scelta della terminologia da impiegare in fase di indicizzazione, in base alle modalità con le quali essa viene organizzata dalla BNCF nel Thesaurus, il vocabolario multidisciplinare online ad essa dedicato<sup>7</sup>.

Non dimenticando le raccomandazioni IFLA sull'opportunità di aggiornare con una certa frequenza gli strumenti di indicizzazione, ci eravamo ripromessi di realizzare una seconda edizione già da alcuni anni; ce l'abbiamo fatta a dispetto del crollo del personale della BNCF (fra la prima e la seconda edizione della *Guida* il settore si è ridotto a un terzo), ce l'abbiamo fatta a dispetto della pandemia, del *lockdown*, del lavoro agile e delle riunioni a distanza.

## Il Gruppo di lavoro e il metodo

L'esperienza che ha portato a questa edizione si inserisce in quella tradizione di lavoro collaborativo che ha contraddistinto sin dall'inizio il progetto del *Nuovo soggettario*, sin da quando Antonia Ida Fontana, in quegli anni direttrice della BNCF, affidò lo studio di fattibilità per il rinnovamento del *Soggettario* del 1956 a un gruppo di esperti, coordinato da Luigi Crocetti, con il quale la Biblioteca nazionale lavorò fra il 2001 e il 2006 in un confronto continuo, un confronto aperto anche ad altri esperti, a Diego Maltese, a Carlo Revelli.

La cifra del "lavoro comune" (come lo chiamava Luigi Crocetti, riferendosi anche alla tradizione inaugurata da Casamassima) da quel momento non è mai venuta a mancare e ha contraddistinto tutta l'evoluzione del *Nuovo soggettario* fino alla seconda edizione del volume, realizzata grazie a studio, a scambi di idee, elaborazioni e revisioni incrociate di parti di testo, la maggior parte delle quali sono il risultato di una redazione collettiva.

Il valore aggiunto di questa seconda edizione è dato dal fatto che l'impegno per realizzarla è stato portato avanti in parallelo e in aggiunta sia al lavoro ordinario del settore che in BNCF sviluppa e cura la manutenzione continua del Thesaurus, sia in parallelo al progressivo impiego del *Nuovo soggettario* da parte della BNI e delle biblioteche italiane, un impiego che nel corso degli anni ha messo in luce aspetti da chiarire, da approfondire, da spiegare meglio.

Questo gruppo di coraggiosi che si sono dedicati alla redazione della seconda edizione della *Guida* è costituito dalle colleghe del settore "Ricerche e strumenti di indicizzazione semantica" della BNCF, da consulenti ed esperti della materia. Hanno dato un contributo coloro che operano all'interno di istituzioni e enti con cui sono state stipulate convenzioni e accordi e con i quali la Biblioteca nazionale ha stabilito negli anni collaborazioni preziose. Ma vorrei dire che anche colleghi di biblioteche che non sono partner ufficiali<sup>8</sup> hanno partecipato ad uno scambio che ha favorito l'avanzamento dello strumento, un esempio per tutti, i catalogatori vicentini della Biblioteca Bertoliana, i colleghi ravennati e molti altri.

## I ringraziamenti

Dunque i miei più grandi ringraziamenti per il contributo dato alla redazione del testo vanno alle colleghe del settore che coordino, Elisabetta Viti ed Eleonora Marzocca, ai consulenti ester-

<sup>7</sup> <https://thes.bncf.firenze.sbn.it/ricerca.php>.

<sup>8</sup> <https://thes.bncf.firenze.sbn.it/enti.htm>.

ni, i maggiori esperti italiani di indicizzazione, Pino Buizza e Alberto Cheti al quale dobbiamo, come era già avvenuto per la prima edizione, una partecipazione e un contributo intellettuale speciale, generosamente prolungato nel tempo.

Hanno dato un grandissimo apporto anche Luciana Franci, Maria Chiara Giunti, Marta Ricci, ex colleghe che hanno lavorato per anni in BNI: oltre alle loro conoscenze è stata fondamentale la loro esperienza. Grazie anche a Elisabetta Francioni che si è occupata dell'editing.

Prima di licenziare il testo abbiamo diffuso le bozze finali per una consultazione pubblica. Le hanno lette e hanno comunicato i loro consigli: Paolo Wos Bellini e Maria Chiara Iorio della BNI (rispettivamente il responsabile e la referente per l'indicizzazione semantica) ed inoltre Carlo Bianchini, Laura Gasparini, Claudio Gnoli, Mauro Guerrini, Nicola Palazzolo, Federica Paradisi, Alberto Petrucciani.

I miei ringraziamenti vanno anche all'Associazione italiana biblioteche (AIB): ad Anna Galluzzi, a Palmira Barbini e a Giovanna De Pascale per il grande lavoro di editing che ha svolto in tempi brevissimi sulla versione digitale. Ma soprattutto vanno alla Presidente Rosa Maiello perché ha accettato una proposta che rispondeva ai nostri "ideali": realizzare un'edizione aperta, gratuita, messa a disposizione delle biblioteche italiane senza oneri finanziari a loro carico.

Un grazie anche a tutti coloro che avevano partecipato alla prima edizione e agli allora "giovannissimi del *Nuovo soggettario*", alcuni dei quali stanno seguendo questa presentazione.

A Maria Grazia Pepe va il merito di gestire in BNCf la parte informatica che riguarda il *Thesaurus*.

Non posso infine non ricordare l'Editrice Bibliografica che nel 2007 fu nostra compagna di avventura per la pubblicazione della prima edizione.

## La progettazione della seconda edizione, elementi di contesto e di altra natura

Ma veniamo alle caratteristiche di questo volume, perché credo che oggi sia questa la cosa interessante di cui parlare.

In quale ottica ci si deve mettere quando si progetta la nuova edizione di uno strumento di questo tipo? Che la *Guida* si sarebbe dovuta evolvere l'avevamo già dichiarato nell'*Introduzione* della prima edizione.

Quando abbiamo iniziato ad occuparci di cosa doveva essere la seconda non potevamo non allargare lo sguardo e, come si fa tutte le volte che si vogliono elaborare proposte di miglioramento, con il gruppo di lavoro abbiamo deciso di partire:

1. dall'analisi di fattori "di contesto"
2. dalla rilevazione delle criticità (ad esempio, alcune ridondanze; alcune parti sulle norme sintattiche troppo complesse; un Indice finale un po' scarno)
3. dalla valutazione di elementi di rilevanza (ad esempio, la diffusione e l'impiego che la *Guida* aveva avuto fino a quel momento, la sua efficacia a livello didattico per l'apprendimento del metodo, ecc.)
4. dalla valutazione di elementi di fattibilità (ad esempio, quali soggetti e professionalità coinvolgere nella redazione della seconda edizione).

È attraverso questo percorso che siamo arrivati alla progettazione di ciò che volevamo fare.

Certamente la *Guida* attuale riflette la convergenza fra più elementi:

- l'evoluzione del sistema di indicizzazione nel suo complesso, nel senso delle componenti, sempre più evolute, messe a disposizione in questi anni dalla BNCF;
- l'evoluzione delle modalità che hanno contraddistinto il suo impiego;
- l'evoluzione del panorama internazionale relativo all'indicizzazione.

Ai tanti cambiamenti nei prodotti del sistema *Nuovo soggettario* e nel contesto ne avrebbero corrisposto altrettanti per la versione aggiornata della *Guida*.

Per quanto riguarda i cambiamenti che potremmo definire “esterni” al sistema, si doveva partire dal fatto che gli anni che ci separavano dalla prima edizione non erano stati anni qualunque.

Avevano visto la luce nuovi standard (fondamentale ISO 25964:2011-2013 *Thesauri and interoperability with other vocabularies*), erano stati pubblicati nuovi modelli concettuali (IFLA LRM), implementate nuove funzionalità degli OPAC, sperimentate e avviate, a livello internazionale, forme di indicizzazione diverse, come quella per Generi/Forme o quella automatica.

Ma soprattutto erano stati anni in cui si erano chiarite le opportunità offerte da:

- una visione rinnovata del web: l'affermazione del web semantico e dei linked open data;
- la convinzione che non si può prescindere dal multilinguismo dei cataloghi e degli strumenti impiegati per allestirli perché ciò favorisce l'accesso.

Anni in cui si è preso atto che gli istituti della memoria (nella cosiddetta ottica GLAM) dovrebbero valutare la possibilità di condividere standard, dovrebbero lavorare per integrare e rendere interoperabili i propri dati e sviluppare progetti che li liberino in un'ottica di condivisione, apertura e, non ultimo, di risparmio economico.

Di tutto ciò dovevamo tener conto.

Ma è anche vero che molti di questi fattori che potremmo definire di “contesto esterno” erano stati previsti dalla BNCF. Giovanni Bergamin già nel 2006, e con la sua consueta lungimiranza, aveva indicato il futuro del *Nuovo soggettario* nella direzione dei servizi web e dei protocolli standard; sapevamo benissimo che Zthes non sarebbe stato l'unico formato per i suoi dati, tant'è che dal 2010 abbiamo implementato SKOS/RDF. Inoltre, la macrostruttura del vocabolario, per come era già concepita, ne garantiva un accrescimento rigoroso; il software scelto per il gestionale del Thesaurus era stato un adattamento di quello di Agrovoc, il noto Thesaurus multilingue della Food and Agriculture Organization of the United Nations (FAO)<sup>9</sup>, il che avrebbe significato possibilità di gestire, come poi è di fatto avvenuto, il nostro multilinguismo. Il linguaggio del *Nuovo soggettario*, infine, era sin dall'inizio concepito per essere applicato anche per materiali non bibliografici.

Da un punto di vista “interno” al sistema, il *Nuovo soggettario* ha avuto dal 2007 a oggi un'evoluzione per sua naturale necessità di crescita, per la rete di cooperazione che la BNCF ha sempre più allargato, per il dialogo con specialisti e gruppi di ricerca e con organizzazioni italiane e straniere. È stato presentato in sedi IFLA e si è evoluto grazie a confronti internazionali di vario tipo.

Le biblioteche italiane, e in particolare quelle di SBN, lo hanno via via adottato, seppur in fasi e modalità differenti; l'ICCU lo contempla fra gli strumenti catalografici ufficiali.

---

<sup>9</sup> <https://agrovoc.fao.org/browse/agrovoc/en/>



## I fenomeni e i processi che hanno favorito la seconda edizione

I fattori che hanno influito sulla seconda edizione sono di varia natura; potremmo raggrupparli in base ai principali fenomeni e processi a cui sono legati.

### 1. Lo sviluppo del Thesaurus

- ad accesso gratuito già dal 2010;
- in costante incremento (dai 13.000 termini del prototipo ai quasi 70.000 attuali) anche grazie alla BNI e alle biblioteche che inviano proposte di terminologia, offrono consulenze in ambiti specialistici; un incremento che ha riguardato i termini, le relazioni semantiche, le note, i collegamenti;
- dal 2010 con metadati nel formato SKOS/RDF che li rende aperti, scaricabili, riutilizzabili, e che gli ha fatto conquistare le cinque stelle di Tim Berners-Lee fra i data set della linked open data cloud<sup>10</sup>. Il Thesaurus è disponibile anche sulla piattaforma <<https://dati.beniculturali.it>> del Ministero della Cultura<sup>11</sup>;
- il multilinguismo e il numero sempre maggiore di collegamenti (in certi casi reciproci) con LCSH: 15.170 / RAMEAU: 13.720 / GND: 4.220 / EMBNE: 2.730 (dati numerici riferiti all'aggiornamento di aprile 2022);
- i sempre più frequenti contatti con altre Biblioteche nazionali per condividere multilinguismo e *policies* sull'interoperabilità; contatti con progetti internazionali e partecipazione a workshop come quello recente di *European common data space*, un'iniziativa della Commissione europea<sup>12</sup>;
- le migliaia di collegamenti, tramite il campo Fonti, con repertori online, con l'Enciclopedia Treccani, con altri sistemi di organizzazione della conoscenza (KOS);
- i collegamenti a Wikipedia (reciproci da maggio 2013) e con Wikidata: quello del *Nuovo soggettario* è stato il primo thesaurus multidisciplinare a collegarsi a Wikipedia;
- la mappatura con la Classificazione Dewey sviluppata costantemente e, dal 2014, riferita alla WebDewey italiana<sup>13</sup>;
- i collegamenti con database di archivi e di musei, in un'ottica che mira a integrare descrizioni del patrimonio bibliografico con quelle di risorse di altra natura. Una funzionalità, inizialmente favorita anche dalla collaborazione con il MAB Toscana, che ha consentito un'apertura verso altri mondi che lavorano per la tutela e la valorizzazione del patrimonio culturale; e così il catalogo della BNCF, tramite il Thesaurus è collegato, solo per fare qualche esempio, con i database delle Gallerie degli Uffizi, del Museo Galileo, della Soprintendenza archivistica Toscana, dell'Archivio Ricordi.

### 2. L'impiego del sistema

Un secondo fattore riguarda direttamente l'impiego del sistema *Nuovo soggettario* con:

- l'uso da parte della BNI dal 2007 e, via via, di altre biblioteche e istituzioni e centri di ricerca;

<sup>10</sup> <https://lod-cloud.net/dataset/bncf-ns>.

<sup>11</sup> [https://dati.beniculturali.it/descrizione\\_dataset/](https://dati.beniculturali.it/descrizione_dataset/).

<sup>12</sup> Si veda ad esempio, [https://www.european-language-grid.eu/wp-content/uploads/2021/11/Slides\\_S1\\_Philippe-Gelin.pdf](https://www.european-language-grid.eu/wp-content/uploads/2021/11/Slides_S1_Philippe-Gelin.pdf).

<sup>13</sup> <https://www.aib.it/pubblicazioni/webdewey-italiana/>.

- il collegamento dei soggetti della BNI alle voci del Thesaurus;
  - il collegamento fra i termini del vocabolario e le stringhe di soggetto tramite il quale si naviga sia nel catalogo della BNCF sia in quello dell'Indice SBN;
  - l'impiego per l'indicizzazione di materiali "non testuali" come fondi fotografici, risorse sonore, fonti iconografiche (ad esempio, da parte di ICBSA, del Museo Galileo, ecc.);
  - l'impiego per materiali archivistici: un esempio per tutti, il progetto Performart<sup>14</sup>.
3. Lo sviluppo della componente sintattica e applicativa
- Un terzo fattore riguarda la componente sintattica e applicativa con:
- l'incremento e la validazione di norme sintattiche verificate in anni di catalogazione;
  - gli adattamenti costanti del *Manuale applicativo* online che si è progressivamente arricchito di nuovi capitoli dedicati a problematiche e casistiche differenti.
4. I gruppi di lavoro e le sperimentazioni
- Sono stati decisivi anche:
- il tentativo di una prima sperimentazione di indicizzazione automatica in BNCF (condotta su tesi di dottorato di ricerca in formato digitale);
  - il gruppo di lavoro sulla soggettazione delle opere antiche, in collaborazione con l'Accademia della Crusca;
  - quello sulla soggettazione di documenti archivistici, in collaborazione con MAB Toscana;
  - il gruppo di lavoro avviato grazie all'ICCU per un coordinamento della catalogazione semantica in SBN.
5. La didattica
- Sono stati innumerevoli i corsi di formazione tenuti in questi anni, soprattutto dal personale della BNCF, per bibliotecari di varie istituzioni, corsi che hanno offerto stimoli e suggerimenti per la revisione di specifiche parti della *Guida* e per l'approfondimento di particolari aspetti.

Tutti insieme questi fattori hanno spinto verso la seconda edizione, si riflettono nei suoi contenuti e hanno influito anche su "come" elaborarla.

Non è un caso che nel sottotitolo si dica che è "interamente rivista e aggiornata": ci sono parti completamente riscritte e parti interamente nuove.

Vi rimando all'*Introduzione* della *Guida* per i dettagli ma sintetizzo alcune delle sue principali novità.

## Le novità della seconda edizione della Guida

Nonostante le norme siano rimaste sostanzialmente invariate e alcune parti del testo non abbiano subito rilevanti cambiamenti, la struttura della *Guida* è stata totalmente ripensata con l'obiettivo di offrire uno strumento aggiornato e di più facile consultazione.

Ora è organizzata in tre sezioni (non più in cinque), dedicate rispettivamente al linguaggio di indi-

---

<sup>14</sup> <https://performart-roma.eu/it/>.

cizzazione (*Il sistema*), ai criteri di allestimento e alle caratteristiche del Thesaurus (*Il Thesaurus*), alla costruzione delle stringhe di soggetto (*Le stringhe di soggetto*).

Queste sezioni sono precedute da una parte totalmente nuova dedicata all'indicizzazione per soggetto in generale. Ne consiglio la lettura, vi sono esposti i principi e gli aspetti generali dell'indicizzazione. Può essere utile anche a chi non impieghi necessariamente il linguaggio del *Nuovo soggettario*. L'utilità della sua redazione è scaturita anche dalle esigenze didattiche di cui parlavo. Nel testo ci sono riferimenti agli ultimi standard e modelli; sono state specificate, in ordine all'accesso per soggetto, le funzioni dell'utente descritte in IFLA LRM, analizzando i modi con cui le componenti di sistema (thesaurus e stringhe di soggetto) e le conseguenti procedure di ricerca nei cataloghi soddisfano le esigenze informative dell'utente. Una nuova Appendice contiene una tabella che verifica, per confronto, la compatibilità generale del *Nuovo soggettario* con il modello IFLA LRM, secondo quanto auspicato nel modello stesso e in ragione del suo livello di astrazione. Un'altra importante novità consiste nel fatto che tutto il testo della seconda edizione privilegia maggiormente il punto di vista di chi deve impiegare le norme per indicizzare e non solo quello di chi cura, sviluppa e implementa il sistema.

Sono stati notevolmente ridotti confronti e riferimenti retrospettivi al vecchio *Soggettario*.

Le procedure sono riferite, più ampiamente che nella prima edizione, anche a risorse non bibliografiche; sono state inserite parti che alludono all'indicizzazione di opere figurative, di fotografie, di opere antiche.

Un paragrafo nuovo è dedicato ad altre forme di indicizzazione che ci auguriamo possano essere implementate, come quella per Generi/Forme, e alle prospettive aperte dall'indicizzazione automatica.

L'organizzazione e la numerazione di capitoli e paragrafi risultano, quindi, del tutto nuove.

Un'altra novità: in alcune parti, il testo è organizzato su due livelli, il livello della trattazione generale (della lettura essenziale) e un livello più specifico, segnalato con corpo tipografico minore e indentazione, destinato ad approfondimenti, a confronti con i precedenti strumenti o a considerazioni su cambiamenti intervenuti nel sistema.

Del Thesaurus sono illustrate tutte le nuove funzionalità implementate negli anni e da esso sono tratti gli esempi di termini, aggiornati e arricchiti rispetto alla prima edizione. Analogamente, gli esempi di stringhe, ritenuti illustrativi di soluzioni coerenti con le norme, sono basati su casi bibliografici reali e documentati nell'OPAC della BNCF o in quello di SBN.

Il *Manuale applicativo*, già online da molti anni<sup>15</sup>, eredita ora parti pubblicate originariamente nella prima edizione della *Guida*, mentre ne cede altre all'attuale edizione. Per esempio, la trattazione sulla forma dei nomi propri impiegati nell'indicizzazione (che era nell'Appendice A della prima edizione), si trova ora nel *Manuale applicativo*; al contrario, la trattazione dei ruoli che i toponimi possono assumere nelle stringhe di soggetto (in passato nel *Manuale applicativo*) ora si trova nella *Guida* (nel paragrafo in cui si illustra il luogo come ruolo complementare).

E ancora a proposito di Appendici, è stata eliminata l'Appendice B che nella prima edizione elencava le fonti individuate per il prototipo del Thesaurus: l'elenco delle fonti, in continuo aggiornamento, è ora consultabile direttamente online nell'interfaccia di ricerca del Thesaurus.

---

<sup>15</sup> [https://thes.bncf.firenze.sbn.it/Manuale\\_applicativo.pdf](https://thes.bncf.firenze.sbn.it/Manuale_applicativo.pdf).

L'*Indice analitico* è completamente cambiato, è frutto di un grosso lavoro concepito non solo per facilitare il rinvio a particolari paragrafi ma anche per visualizzare possibili rinvii, forme alternative, concettualmente associabili, ecc. Vi si sono dedicate soprattutto Marta Ricci ed Elisabetta Viti, già impegnate anche nella redazione del testo. I criteri con cui l'*Indice* è stato allestito sono esplicitati nella sua prima pagina.

Questa nuova edizione, infine, accoglie quasi interamente quanto – nel corso degli anni e fino alla data della sua pubblicazione – era stato comunicato agli utenti tramite il tasto Novità del Thesaurus nel link “Integrazioni, aggiornamenti, modifiche”.

Solo qualche parola ancora, in aggiunta a quanto già detto dal Direttore, dottor Luca Bellingeri, sulla scelta di pubblicare online la seconda edizione: ricordiamo che grazie a ciò sarà più semplice aggiornarla ancora, farne altre edizioni, potenziarne il collegamento al *Manuale applicativo*, collegarla alle Linee guida che produrrà l'ICCU e ad altri strumenti, così come alle analoghe guide di altre Biblioteche nazionali straniere, tutte già disponibili gratuitamente online.

## Il futuro e le conclusioni

A voi bibliotecari, utenti, studiosi di indicizzazione, va ora la parola su questa nuova edizione, sulla sua utilità e tenuta sia teorica che pratica. Sulla sua capacità di offrire un metodo, spiegarlo, evidenziarne le capacità predittive.

Vorrei concludere ricordando ancora il valore che in questo lavoro ha avuto la cooperazione, un valore così grande che oggi interverranno proprio rappresentanti di alcune delle istituzioni che sono da anni nostri partner. Ma vorrei anche ricordare il grande contributo che negli ultimi anni l'ICCU ha dato per favorire in SBN una condivisione di dati semantici sinora mai davvero così supportata. Tengo a sottolineare, però, che il lavoro da fare insieme non finisce qui; molto si deve ancora sviluppare e migliorare sul versante dell'integrazione con gli OPAC; sull'impiego dello strumento per tipologie particolari di opere (come le opere del passato; alcuni tipi di *fiction*); sull'avvio di forme alternative di indicizzazione. Ad esempio, non possiamo non riprendere il discorso sull'indicizzazione automatica con modalità che utilizzino – oltre che adeguati software – anche le potenzialità del nostro Thesaurus. Ma per fare questo ci vogliono competenze, risorse umane.

Insomma si può ancora fare molto per valorizzare l'adozione di uno strumento che, per le sue caratteristiche, è troppo avanti per non poter sperare di avere buone “prospettive di vita”...

Al momento le risorse umane per garantirne la tenuta sono davvero esigue. E non basterà che la BNCF abbia il personale per continuare a curarlo e svilupparlo, servirà che anche le biblioteche italiane abbiano le forze da dedicare all'indicizzazione per soggetto.

Abdicare a queste attività, inibire l'accesso alle opere tramite i loro contenuti concettuali, significherebbe interrompere un servizio fondante del ruolo di mediazione che svolgiamo fra cittadini, studiosi, ricercatori e il nostro patrimonio culturale.

Concludo davvero con una nota di ottimismo, leggendo una frase in cui ci riconosciamo molto; non a caso è pubblicata nell'Home page del *Nuovo soggettario*. La ricondivido oggi con tutti i presenti perché riflette l'atmosfera che aveva distinto la realizzazione della prima edizione ma che – nonostante il tempo trascorso – ha rispecchiato anche la seconda e mi pare rispecchi anche il clima festoso di questa giornata.

Diceva Carla Guiducci Bonanni, durante la presentazione del *Nuovo soggettario* (in Palazzo Vecchio a Firenze, l'8 febbraio 2007), riferendosi al gruppo che aveva lavorato al *Soggettario* del 1956 e al quale lei stessa, giovanissima, aveva partecipato:

È [quella del *Soggettario*] una storia bella, una storia che va al di là delle biblioteche perché è una storia che parla di passioni, di interesse, di vero piacere di lavorare, perché credo che il miglior premio che si possa avere sia quello del piacere di lavorare insieme, di ottenere dei risultati, di poter partecipare con gli altri ai risultati e di poter in qualche modo capire che quello che si fa è un passo avanti per il futuro (Guiducci Bonanni 2007: 81).

## **Riferimenti bibliografici**

Guiducci Bonanni, Carla. 2007. “[Intervento di apertura della giornata di presentazione del *Nuovo soggettario*]”. *Biblioteche oggi* 25 (6): 80-81.

## Rule following Rules and practice in subject indexing\*

Alberto Cheti

Contact: Alberto Cheti, [chetialberto@infinito.it](mailto:chetialberto@infinito.it)

Received: 5 June 2022; Accepted: 2 July 2022; First Published: 15 September 2022

### ABSTRACT

The importance attached to subject indexing in library services, librarian awareness of its social role and practical purposes, attention to indexing practices and the use of catalogs: How are these elements reflected in the system of rules governing subject indexing in Italy? What does “rule following”, to use Wittgenstein’s well-known expression, mean in our field? Why and how do we follow a rule? In which sense is following a rule a practice? How and in which conditions does the rule relate to the practice and vice versa? Beginning with these questions, the main features of the Italian subject indexing rules are illustrated: explicit rules, based on linguistic criteria, contain justifications for why a given rule should be followed and should be sensitive to the use-context. All three of these features imply a relationship with practice; at the same time, the rules presuppose a social practice in which they are employed and express their pragmatic meaning. Thus, the indexing rules are manifested in the language of indexing through the means by which this language is used; that is, in the indexing activity and in the use of catalogs. Hence, the call for libraries to attend to the pragmatic dimension of subject indexing.

### KEYWORDS

Subject indexing; Rules; Nuovo soggettario.

## Seguire una regola. Regole e prassi nell’indicizzazione per soggetto

### ABSTRACT

Il saggio illustra le caratteristiche principali delle regole italiane dell’indicizzazione per soggetto: regole esplicite, basate su criteri di natura prevalentemente linguistica (grammaticale, semantica, sintattica o testuale), accompagnate da giustificazioni pratiche (i motivi, gli scopi, le conseguenze dell’applicazione di una determinata regola). Tutte e tre queste caratteristiche implicano una relazione con la prassi; al tempo stesso, le regole presuppongono una pratica sociale nella quale esse sono impiegate ed esprimono il loro significato pragmatico. All’interno di questa pratica, le regole si manifestano nel linguaggio di indicizzazione tramite gli strumenti attraverso i quali il linguaggio è praticato, ossia nell’attività di indicizzazione e nell’uso dei cataloghi. Da qui il richiamo alla cura da parte delle biblioteche degli aspetti pragmatici dell’indicizzazione per soggetto. Questo contributo è anche un invito a riflettere, nell’ambito dell’indicizzazione per soggetto, sulla connessione tra regola e prassi e sulla prassi come pratica sociale, comunitaria, a partire da alcune osservazioni di Wittgenstein sulle regole, che hanno avuto una eco anche in campo biblioteconomico. Che cosa significa ‘seguire una regola’? Perché e come si segue una regola? In che senso seguire una regola è una prassi?

### PAROLE CHIAVE

Indicizzazione per soggetto; Regole; Nuovo soggettario.

\* Intervento in occasione della presentazione del libro *Viaggi a bordo di una parola*, Firenze, Biblioteca nazionale centrale, 3 ottobre 2019, con l’aggiunta di note e con alcune integrazioni.

Alberto Petrucciani, nel suo contributo in *Viaggi a bordo di una parola* (Lucarelli, Petrucciani e Viti a c. di 2019), conclude la disamina sullo stato dei cataloghi e sulle pratiche di indicizzazione per soggetto con queste parole:

Ma, tornando alla domanda iniziale, il presupposto perché l'indicizzazione per soggetto d'ambito bibliotecario possa avere un futuro, prima ancora dell'esigenza di una seria base teorica e di riflessioni concrete sullo stato di fatto dei cataloghi e sulle pratiche anche di dettaglio, è che le biblioteche considerino quest'attività come rilevante, magari anche qualificante, invece che come del tutto marginale (Petrucciani 2019, 173).

E, qualche pagina prima:

Non perdere di vista gli scopi o le finalità che l'indicizzazione per soggetto nei cataloghi di biblioteca dovrebbe perseguire e insieme tenere un occhio attento sulle caratteristiche effettive che hanno oggi i cataloghi potrebbe servire anche ad affrontare meglio le questioni concrete e a volte minute (Petrucciani 2019, 170).

Dunque, Petrucciani indica tre macro-condizioni necessarie a garantire un futuro all'indicizzazione per soggetto, in un mutato contesto tecnologico e culturale. La prima è la considerazione da parte delle biblioteche dell'importanza dell'indicizzazione nell'ambito dei servizi bibliotecari, della sua specificità rispetto ad altre forme di recupero dell'informazione. Questa condizione ne presuppone un'altra, ossia la consapevolezza delle finalità generali dell'indicizzazione e degli scopi pratici che essa deve perseguire nei cataloghi di biblioteca. La terza è l'occhio vigile sulle caratteristiche effettive che hanno oggi i cataloghi, per verificarne la rispondenza agli scopi dell'indicizzazione, affrontando le questioni concrete e adottando le soluzioni pratiche conseguenti.

Possiamo ricostruire il ragionamento in questo modo: la rilevanza da attribuire all'attività di indicizzazione per soggetto, condizione prima perché essa abbia un futuro, presuppone una consapevolezza delle sue finalità e del suo ruolo, da cui dipendono le modalità di azione e le pratiche catalografiche più efficaci. Abbiamo, dunque, tre momenti, tre aspetti interconnessi: l'importanza che si riconosce all'attività di indicizzazione per soggetto; le funzioni che questa attività deve svolgere, gli scopi che essa deve perseguire; le pratiche di indicizzazione e l'uso ordinario dei cataloghi.

Come si riflettono questi elementi nel sistema di regole che governano in Italia l'indicizzazione per soggetto? Che cosa significa, nel nostro ambito, "seguire una regola", per usare la nota espressione del filosofo Ludwig Wittgenstein? Perché e come si segue una regola? In che senso, come osserva ancora Wittgenstein, "seguire una regola è una prassi" e "non si può seguire una regola *privatim*"?<sup>1</sup>

---

<sup>1</sup> La citazione di alcune espressioni di Wittgenstein relative alla nozione 'seguire una regola', tratte dalle sue *Ricerche filosofiche* (Wittgenstein 1967, parte prima, par. 198-242), non risponde qui ad alcun intento interpretativo, per mancanza di spazio e, soprattutto, per incompetenza di chi scrive. Essa ha il solo scopo di ricavarne delle suggestioni per una riflessione, nell'ambito dell'indicizzazione per soggetto, sulla connessione tra regola e prassi e sulla prassi come pratica sociale, comunitaria. Le osservazioni di Wittgenstein sulle regole hanno avuto una eco anche in campi disciplinari diversi dalla filosofia, come quello giuridico e artistico. In biblioteconomia segnalo come punto di riferimento imprescindibile il saggio di Bernd Frohmann *Rules of indexing: a critique of mentalism in information retrieval theory* (Frohmann 1990), dove si trova un'ampia e articolata esposizione delle osservazioni di Wittgenstein sulla nozione 'seguire una regola' e delle loro implicazioni per la scienza dell'informazione, a cui si sono richiamati successivamente molti degli studiosi che hanno affrontato questo tema



Al di là delle loro implicazioni filosofiche, queste espressioni possono aiutarci a comprendere meglio alcuni fatti relativi alle regole che ci riguardano da vicino. Veniamo da un'esperienza quarantennale di costruzione di un sistema di regole per l'indicizzazione per soggetto: dagli studi, le proposte, le sperimentazioni degli anni Ottanta alla *Guida GRIS* (GRIS 2001) e al *Nuovo soggettario*, di cui a breve la Biblioteca nazionale di Firenze presenterà alla comunità bibliotecaria una guida interamente rivista e aggiornata.<sup>2</sup> Si è trattato, in secondo luogo, di un lavoro collettivo, collaborativo, da parte di alcune generazioni di bibliotecari con esperienze, provenienze, profili professionali, ruoli diversi; di un lavoro articolato, nel quale le regole sono state oggetto di una varietà di interventi: studio, elaborazione, sperimentazione, insegnamento, applicazione. Il risultato di questo lavoro è frutto di un processo caratterizzato da una sorprendente continuità di intenti, di riferimenti scientifici e culturali, di approcci metodologici. Eppure – e questa è la terza considerazione – non siamo soddisfatti del modo in cui queste regole si manifestano nella pratica di indicizzazione e nell'uso ordinario dei cataloghi; desidereremmo una loro implementazione più ricca e articolata, come mostrano, anche in questo volume, i rilievi di Petrucciani (Petrucciani 2019), di Andrea Fabbrizzi (Fabbrizzi 2019) e di Claudio Gnoli (Gnoli 2019).

Questi fatti inducono a riflettere sul significato e le implicazioni del “seguire una regola”, da cui ricavare alcune indicazioni di lavoro. Il tema è, mi sembra, la relazione tra la regola e la prassi: in che modo e a quali condizioni la regola si connette alla prassi e viceversa?

Occorre, innanzitutto, evitare alcune scorciatoie. Ne indico tre, limitandomi a enumerarle, senza approfondire:

1. La regola come codificazione di una pratica. Considerare la regola come codificazione di una pratica non equivale a una prassi governata da regole. E senza regole la pratica si riduce a comportamenti individuali delle persone impegnate in una determinata attività, senza poter distinguere se in una determinata circostanza una persona stia seguendo una regola e quale regola.
2. La regola come istruzione. Considerare la regola come un'istruzione da seguire pedissequamente può derivare da una pregiudiziale diffidenza verso regole basate su procedimenti logici, verso regole di applicabilità generale, ritenute di per sé a-contestuali e a rischio di mentalismo.
3. La regola come un procedimento algoritmico. Considerare l'applicazione di una regola come un procedimento algoritmico, un calcolo che produce sempre lo stesso risultato, a prescindere dal contesto in cui la regola opera, può essere erroneamente indotto dall'esigenza di dare alle regole una base oggettiva e universale.

---

(Hjørland 1998, Andersen and Christensen 1999, Anderson and Pérez-Carballo 2001, Ventura 2010, Cavaleri 2013, per citarne alcuni). Aggiungo soltanto che mi pare discutibile la critica di Frohmann al mentalismo che sarebbe insito nelle macro-regole di sommarizzazione, da lui intese come “internally realised and tacitly known mental rules”. Va precisato che: a) queste regole non operano isolatamente, ma nel contesto di un linguaggio (il linguaggio di indicizzazione) e di una pratica (l'analisi concettuale dei documenti) attraverso processi interattivi testo-contesto; b) è in questo ambito che le regole acquistano il loro significato pragmatico, ossia possono fornire un aiuto nell'individuare i temi di un'opera, ricostruirne progressione e organizzazione, accertarne il grado di pertinenza; c) tuttavia, esse non determinano di per sé il risultato finale dell'analisi, ossia la scelta del/dei soggetti di un'opera, sulla quale influiscono anche valutazioni di rilevanza dovute a fattori contestuali. È con questo tipo di impiego che le macro-regole sono state inserite nella *Guida GRIS* e richiamate nella *Guida al Nuovo soggettario* (Cheti 2016).

<sup>2</sup> La 2ª edizione della *Guida al Nuovo soggettario* è stata pubblicata nel dicembre 2021 (BNCF 2021).

Dobbiamo, dunque, pensare a una relazione più complessa tra regola e prassi, adeguata al valore sociale e agli scopi di un'attività come l'indicizzazione per soggetto.

Le regole di indicizzazione che la comunità bibliotecaria italiana si è data presentano almeno tre componenti: un procedimento logico, una giustificazione pratica, un contesto operativo.

Primo: il procedimento logico. La regola è costruita di solito su un criterio (caratteristica, proprietà, funzione, relazione) di natura prevalentemente linguistica (grammaticale, semantica, sintattica o testuale), applicabile alla generalità dei casi: per esempio, il criterio della *numerabilità* per stabilire la forma plurale o singolare dei termini; oppure, le *relazioni sintattiche* che intercorrono tra le parole di un termine composto nelle regole di scomposizione; o ancora, le nozioni di *tema*, *categoria*, *faccetta*, *ruolo* nell'analisi testuale, semantica, sintattica dei concetti. Si verifica la rispondenza del singolo caso al criterio – ossia, se il termine esprime un concetto numerabile o non numerabile; oppure, se la relazione tra i componenti del termine composto propende o meno per la scomposizione; o ancora, quale ruolo svolge un concetto nella formulazione del soggetto, ecc. – e si applica al caso la regola. L'insieme dei casi che soddisfano la medesima condizione avrà una soluzione analoga.

È forse opportuno osservare che l'applicabilità generale della regola non dipende da nessun assunto circa la sua natura psicologica o mentale, né giustifica di per sé un giudizio sulla sua presunta a-contestualità. L'applicabilità generale della regola dipende unicamente dal grado di generalità del criterio adottato. Né essa implica che il passaggio dalla regola alla sua applicazione corretta debba intendersi come un procedimento algoritmico, una conseguenza logica della formulazione della regola. Il passaggio è accompagnato da esempi, spesso da eccezioni, opzioni preferenziali, avvertenze cautelative, variabili contestuali, ecc. Per esempio, un termine composto, sebbene scomponibile in base alle regole di scomposizione, è mantenuto nella forma composta se si accerta che ragioni semantiche o linguistico-culturali ne sconsigliano la scomposizione. Oppure, la traduzione di un soggetto nel linguaggio di indicizzazione può avere, in qualche caso, più di una soluzione sintattica corretta: quale scegliere? Ancora, il tema di un'opera può essere pertinente e rilevante anche se, in base alle regole di sommarizzazione, non è individuato come il tema di base o il tema principale.

Questi esempi mostrano che la formulazione di una regola non basta per assicurare una regolarità nella prassi: il passaggio dalla regola alla sua applicazione corretta dovrebbe essere accompagnato, oltre che dall'insegnamento e dall'addestramento, da risultati autorevoli dell'applicazione delle regole, rappresentati dai termini e dalle stringhe di soggetto prodotte nell'ambito di una cooperazione tra le biblioteche su scala nazionale, come auspicato fin dall'avvio del progetto del *Nuovo soggettario*.

Secondo: la giustificazione pratica. La regola contiene le giustificazioni del perché seguire quella determinata regola: a cosa serve, qual è lo scopo che vuole raggiungere, quale il suo ruolo all'interno dell'attività di indicizzazione, quali conseguenze ci si aspettano dalla sua applicazione, quali impegni richiede a coloro che partecipano a questa attività. Dare una spiegazione della regola in termini di ragioni, di motivazioni è rendere esplicito il suo significato pragmatico. Che cosa faccio e a quali mosse induco l'interlocutore (indicizzatore o utente) quando scelgo un termine, lo impiego per indicizzare il soggetto di un'opera, costruisco una stringa di soggetto seguendo le regole? Se la regola è un mezzo per raggiungere lo scopo, allora la giustificazione, la motivazione è ciò che manifesta, rende trasparente il legame tra la regola e lo scopo. Se la normatività delle regole non

deriva da un punto di vista esterno, da una posizione autonoma e superiore rispetto all'attività a cui le norme si riferiscono; se la loro ragione ultima è collettiva, sociale, se il loro fondamento è nell'accordo, nella condivisione da parte di una comunità, allora la giustificazione ne diventa una componente essenziale.

Quarant'anni fa constatava Alfredo Serrai:

Per evitare di lacerarsi perennemente la coscienza, i codificatori di regole hanno accantonato la preoccupazione principale, quella cioè di realizzare *lo scopo del catalogo*: che non è altro se non la mediazione tra libri e lettori.

Le regole di catalogazione, invece, rimangono delle istruzioni in codice per un discorso all'interno del mondo dei catalogatori (Serrai 1980, 100).

E più avanti:

Senza precise finalità generali e, in primo luogo, al di fuori dell'assillo della rispondenza del meccanismo catalografico ai bisogni dell'utenza, assillo che si deve tradurre in procedure limpide enunciate e in obiettivi chiaramente ed esaurientemente articolati, l'accettazione o il rifiuto di un nuovo codice di catalogazione perde significato e diventa una operazione, appunto, futile.

I codici si sono sostituiti l'uno all'altro per ragioni interne, ossia per la incessante aspirazione dei teorici della catalogazione a sussumere la casistica sotto pochi principi, ed a perfezionare la compatibilità e la coerenza tra questi principi (Serrai 1980, 106).

A quest'ultima affermazione fa eco l'ammonimento di Luigi Crocetti:

La biblioteconomia fornisce gli strumenti per maneggiare la cultura: è una disciplina tecnica, ma guai se i suoi principî cominciano a svilupparsi autonomamente, a dedursi l'uno dall'altro, senza tener conto dei cambiamenti culturali e delle necessità del pubblico (Crocetti 2014a, 507).<sup>3</sup>

Ora, se si vuole che le regole non si sviluppino per ragioni interne, non vengano dedotte l'una dall'altra senza tenere conto dei bisogni dei lettori, non perseguano unicamente obiettivi di coerenza, occorre che ad esse si accompagnino le giustificazioni in termini di scopi e obiettivi da realizzare, anche se la loro realizzazione non dipende in ultima istanza dalla formulazione delle regole, ma dall'efficacia dagli strumenti catalografici che le regole aiutano a costruire. Nella nuova edizione della *Guida al Nuovo soggettario* alcune tabelle mostrano sinteticamente come le principali entità del linguaggio, termini del thesaurus e stringhe di soggetto, concorrano a soddisfare gli obiettivi dell'utente del catalogo e suggeriscono modalità di ricerca per soggetto ad essi funzionali. La connessione tra la regola e l'obiettivo riduce il rischio di incorrere nell'equivoco di cui parla

---

<sup>3</sup> Crocetti aveva affrontato questo tema nella sua relazione introduttiva – dal titolo significativo *Tradizione biblioteconomica e necessità di aggiornamento periodico delle norme di catalogazione* (Crocetti 2014b) – a un corso di aggiornamento per bibliotecari organizzato dalla Provincia di Trento (Arco, 19-22 maggio 1981), giustificando la necessità di rinnovamento dei codici di catalogazione come conseguenza di una catena di cambiamenti: i cambiamenti sociali, culturali, tecnologici e organizzativi che coinvolgono il mondo delle biblioteche inducono una nuova biblioteconomia e, dunque, nuove norme di catalogazione e il loro aggiornamento periodico. Questa preoccupazione sarà una costante della riflessione e dell'insegnamento di Crocetti (Cheti 2011).

Serrai, ossia di ritenere che il “perfezionamento logico” equivalga di per sé a un “perfezionamento funzionale”, che la “ricerca della coerenza” possa bastare ad assicurare l’efficacia e l’efficienza del catalogo. Per esempio, nel capitolo della *Guida* dedicato alle stringhe di soggetto sono individuate le caratteristiche generali della forma di pre-coordinazione e del metodo di analisi adottato, e indicati i vantaggi che ne derivano. Gli obiettivi sono poi specificati in obiettivi che potremmo definire di *risultato* (richiamo e precisione nel recupero, trasparenza del linguaggio di indicizzazione) e obiettivi che riguardano le *prestazioni* e i *processi* per arrivare al risultato finale (quali requisiti devono avere le stringhe di soggetto, come facilitare l’analisi di un enunciato, in base a quali criteri assegnare i ruoli e ordinare i termini nella stringa, ecc.). Per ultimo, anche il requisito della coerenza dell’organizzazione delle norme, perché, come scrive Serrai, “i bibliotecari non vanno ritenuti quali nemici naturali dei lettori” e domandarsi “se il catalogo potrà essere la migliore architettura logica che il bibliotecario è in grado di costruire o il miglior sistema per soddisfare le richieste dei lettori è tendenziosa e ha valore polemico piuttosto che euristico” (Serrai 1980, 106).

La connessione tra la regola e l’obiettivo pone le biblioteche e i sistemi bibliotecari nella condizione di adottare soluzioni consapevoli e condivise. D’altra parte, è sulla base della rispondenza della regola all’obiettivo che le regole dovrebbero essere sviluppate, modificate, migliorate, oppure sostituite da altre regole capaci di servire meglio agli obiettivi stabiliti.

Terzo: il contesto operativo, il contesto d’uso. Il contesto è presente diffusamente nelle regole, in una varietà di forme che dipendono dall’oggetto della regola, dall’operazione a cui la regola si riferisce. Per esempio, quando si analizza un’opera per individuarne il soggetto, il contesto è il significato di una parola o di un enunciato all’interno dell’opera, la sua pertinenza con un tema rilevante in essa identificato, la sua corrispondenza o meno con il significato registrato nel thesaurus. Oppure, quando si tratta di stabilire significato e forma di un termine di indicizzazione, il contesto è l’uso del termine nel linguaggio corrente o nell’ambito specialistico di una disciplina, e così via. La relazione tra significato, forma e uso di un termine è alla base della costruzione e dell’impiego del thesaurus del *Nuovo soggettario*. In questo senso, il contesto è parte integrante della regola, è l’elemento che fa sì che l’applicazione della regola non sia un procedimento algoritmico, ma un procedimento dinamico, dialettico, euristico.

Tutte e tre le componenti delle regole (il criterio, la giustificazione pratica, il contesto d’uso) implicano una relazione con la prassi, intesa come regolarità nell’applicazione delle regole, come obiettivi che esse intendono perseguire, come contesto documentario, linguistico, culturale all’interno del quale operano. C’è poi un significato più ampio di ‘contesto’ che le regole presuppongono: è la pratica sociale nella quale il sistema di regole si iscrive, l’attività nella quale le regole sono impiegate e in cui esprimono pragmaticamente il loro significato. Se qualunque regola si manifesta nella sua applicazione corretta e costante per gli scopi di una determinata attività, le regole dell’indicizzazione si manifestano nel linguaggio di indicizzazione, tramite gli strumenti attraverso i quali il linguaggio è praticato, ossia nell’attività di indicizzazione e nell’uso ordinario dei cataloghi.

Questo è del tutto evidente nel comportamento dell’utente, “che non è tenuto a sapere nulla del linguaggio di indicizzazione che gli viene proposto” (Fabbrizzi 2019, 89), che “interagisce con il catalogo non conoscendo né i criteri secondo i quali sono stati scelti e creati i dati né le tecniche di information retrieval o la logica delle interfacce” (Galeffi 2017, 248), che non apprende il linguaggio attraverso la conoscenza esplicita delle sue regole, ma attraverso l’uso che se ne fa nei cataloghi, nell’interazione tra l’utente e il catalogo, in quella pratica discorsiva attraverso la quale l’utente è

messo in condizione di usare correttamente ed efficacemente il linguaggio del catalogo, effettuando inferenze probabilistiche sui temi di interesse trattati nei documenti.

Lo stesso vale per gli indicizzatori. Costruire una stringa di soggetto seguendo le regole significa produrre un'espressione linguistica nella quale si riflette il significato pragmatico della regola, che un altro indicizzatore può comprendere, consentendogli di agire di conseguenza, di fare la mossa giusta, per esempio, associando alla stessa opera un'altra stringa di soggetto ugualmente regolare nel metodo di costruzione, ugualmente pertinente e rilevante nella sua connessione con l'opera e con l'utenza. Allora troveremmo nel catalogo le stringhe di soggetto assegnate a un'opera non come il risultato, talvolta confuso e incoerente, di un assemblaggio meccanico di stringhe di soggetto prodotte dalle singole biblioteche, ma come il frutto di una pratica sociale condivisa, il cui risultato è una descrizione chiara ed esaustiva dei soggetti di ciascuna opera, tale da consentire all'utente di identificare le opere corrispondenti al soggetto cercato.

È in questa pratica discorsiva, in cui indicizzatori e utenti usano in modo corretto, ordinato ed efficace il linguaggio di indicizzazione, in questa pratica intersoggettiva, collettiva, sociale, in cui è impegnata la comunità bibliotecaria, che la regola diventa 'usuale', 'ordinaria'.

Principi, regole, motivazioni, uso del linguaggio nelle pratiche di indicizzazione e nella costruzione e consultazione dei cataloghi costituiscono uno 'spazio di ragioni' che dà senso all'agire di coloro che partecipano a questa attività, bibliotecari e utenti. In questo spazio di ragioni il luogo decisivo, cruciale è negli scambi comunicativi, nelle pratiche linguistiche in cui si impiegano enunciati riguardanti i temi di un'opera, le opere su un determinato tema, il significato associato a un termine, le relazioni che intercorrono tra i termini e tra i concetti, ecc., ossia nell'uso del linguaggio di indicizzazione attraverso gli strumenti interattivi del catalogo.

Dunque, il rapporto tra la regola e la prassi, da cui è iniziata questa riflessione, è duplice e reciproco: la regola determina una prassi nel corso del suo impiego; dunque, presuppone una prassi nella quale trova applicazione, entro cui si iscrive, si manifesta, che sono le pratiche di indicizzazione e l'uso ordinario dei cataloghi. Le regole rendono possibile la nostra attività e, al tempo stesso, acquistano significato nel contesto di questa attività. C'è un'interdipendenza tra regola e prassi, tra comprensione della regola e produzione di un linguaggio governato da regole. In altre parole, le regole necessitano di essere comprese nell'ambito di una determinata prassi: a esse dobbiamo guardare entro un contesto di atti, relazioni, vincoli, strumenti, scopi. Allora, forse non è del tutto improprio arrischiare qui l'osservazione che la concordanza nel linguaggio usato da bibliotecari e utenti sia una concordanza della "forma di vita", per usare un'altra espressione di Wittgenstein. Ne scaturisce una direzione di lavoro che si colloca in una prospettiva pragmatica. Al termine del lavoro di revisione e aggiornamento del sistema delle regole, con l'uscita della seconda edizione della *Guida al Nuovo soggettario*, si presenta una nuova occasione per occuparci con maggiore cura, anche in dettaglio, degli aspetti pratici, ossia dell'uso del linguaggio di indicizzazione da parte di indicizzatori e utenti, come suggeriscono i saggi di Petrucciani, di Fabbrizzi e di Gnoli. Questo lavoro dovrebbe essere improntato a quello che Fabbrizzi, in un suo precedente intervento, chiama "principio di trasparenza dell'indicizzazione", da cui discende la raccomandazione di rendere visibili e comprensibili agli utenti, nella presentazione dei dati catalografici, le entità e le relazioni del linguaggio di indicizzazione, per non depotenziare il catalogo e vanificare gran parte del lavoro dei catalogatori (Fabbrizzi 2014). Ricordo che l'accento sulla pragmatica fu posto anche da Stefano Tartaglia nel suo intervento in occasione del Seminario internazionale dell'aprile 2001,

dove furono presentate le proposte di rinnovamento del *Soggettario*, dunque nella prima fase di gestazione del *Nuovo soggettario*. Insomma, l'importanza della pragmatica è proporzionale alla posta in gioco, che è l'uso corretto e significativo del linguaggio nelle pratiche catalografiche.

Curare gli aspetti pratici significa creare le condizioni per un uso efficace del linguaggio nella prassi: l'insegnamento delle regole, l'esercizio di una responsabilità collettiva nella loro applicazione, l'implementazione del linguaggio nei cataloghi, la pratica quotidiana di indicizzazione in un lavoro comune, il riscontro della comprensibilità e dell'efficacia delle regole, del linguaggio e del catalogo con bibliotecari e utenti. Significa anche tenere conto delle potenzialità offerte dalle nuove tecnologie, come i modelli di descrizione della conoscenza sul web, e delle abitudini degli utenti sia nella ricerca che nella produzione di espressioni linguistiche per analoghi scopi, come nelle pratiche di *social tagging*.

Il 'contesto', nelle sue varie forme e articolazioni, è stato richiamato più volte in questo intervento. I saggi contenuti in *Viaggi a bordo di una parola* mostrano una sorprendente varietà di contesti nei quali usare il linguaggio: viaggi delle parole entro strutture classificatorie e thesaurali, nel web semantico, tra lingue diverse, attraverso la "lunga durata" delle opere del passato, in particolari contesti culturali o documentari.

Allora, dare un futuro a una pratica sociale così ricca, articolata, qualificata come l'indicizzazione per soggetto, su cui s'interroga Petrucciani, è anche un mezzo per dare un futuro alla comunità bibliotecaria che vi è impegnata e, forse, alle biblioteche stesse.

## Riferimenti bibliografici

- Andersen, Jack, and Frank Sejer Christensen. 1999. "Wittgenstein and indexing theory". In *Proceedings of the 10th ASIG SIG/CR Classification Research Workshop*, Hanne Albrechtsen Workshop chair and editor, Jens-Erik Mai editor, 1-25. Silver Spring (MD): American Society for Information Science.
- Anderson, James D., and José Pérez-Carballo. 2001. "The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing". *Information Processing and Management* 37 (2):231-254.
- BNCF (Biblioteca nazionale centrale di Firenze). 2021. *Nuovo soggettario: guida al sistema italiano di indicizzazione per soggetto*. Roma: Associazione italiana biblioteche; Firenze: Biblioteca nazionale centrale di Firenze.
- Cavaleri, Piero. 2013. *La biblioteca crea significato: thesaurus, termini e concetti*. Milano: Editrice Bibliografica.
- Cheti, Alberto. 2011. "La lezione camaldolese di Luigi Crocetti". In «*Books seem to me to be pestilent things*». *Studi in onore di Piero Innocenti per i suoi 65 anni*, promossi da Varo A. Vecchiarelli, raccolti, ordinati, curati da Cristina Cavallaro, 863-873. Manziana (Roma): Vecchiarelli Editore.
- Cheti, Alberto. 2016. "I processi cognitivi nell'analisi concettuale dei documenti: una postilla tra biblioteconomia e linguistica". *AIB Studi* 56 (1):7-39.
- Crocetti, Luigi. 2014a. "La tradizione culturale italiana del Novecento". In Id., *Le biblioteche di Luigi Crocetti: saggi, recensioni, paperoles (1963-2007)*, a cura di Laura Desideri e Alberto Petrucciani, presentazione di Stefano Parise, 501-507. Roma: Associazione italiana biblioteche (pubblicato precedentemente in *L'automazione delle biblioteche nel Veneto: tra gli anni '90 e il nuovo millennio. 10. Seminario Angela Vinay, Venezia, Palazzo Querini Stampalia, 29-30 gennaio 1999*, a cura di Chiara Rabitti, 76-81. Venezia: Fondazione scientifica Querini Stampalia, 2000).
- Crocetti, Luigi. 2014b. "Tradizione biblioteconomica e necessità di aggiornamento periodico delle norme di catalogazione". In Id., *Le biblioteche di Luigi Crocetti*, cit., 307-309 (pubblicato precedentemente in *Quaderno RICA-ISBD(M): esempi di catalogazione bibliografica*, a cura di Fabrizio Leonardelli e Luisa Pedrini, 7-9. Trento: Provincia autonoma di Trento, Servizio delle attività culturali, 1981).
- Fabbrizzi, Andrea. 2014. "An Atlas of classification. Signage between open shelves, the Web and the catalogue". *JLIS.it* 5 (2):101-122.
- Fabbrizzi, Andrea. 2019. "I contesti dell'indicizzazione". In Lucarelli, Anna, Alberto Petrucciani, e Elisabetta Viti, a c. di, 2019. *Viaggi a bordo di una parola*, 77-89.
- Frohmann, Bernd. 1990. "Rules of indexing: a critique of mentalism in information retrieval theory". *The Journal of Documentation* 46 (2):81-101.
- Galeffi, Agnese. 2017. "Se il catalogo parlasse, lo capiremmo? Cinque assiomi della mediazione catalografica". *AIB Studi* 57 (2):239-252.

Gnoli, Claudio. 2019. “Rilevanza nel documento e rilevanza per l’utente: l’ordine di citazione dei temi”. In Lucarelli, Anna, Alberto Petrucciani, e Elisabetta Viti, a c. di. 2019. *Viaggi a bordo di una parola*, 105-117.

GRIS (Gruppo di ricerca sull’indicizzazione per soggetto). 2001 (ed. orig. 1996). *Guida all’indicizzazione per soggetto*. Roma: Associazione italiana biblioteche.

Hjørland, Birger. 1998. “Information retrieval, text composition, and semantics”. *Knowledge Organization*, 25 (1/2):16-31.

Lucarelli, Anna, Alberto Petrucciani, e Elisabetta Viti, a c. di. 2019. *Viaggi a bordo di una parola. Scritti sull’indicizzazione semantica in onore di Alberto Cheti*. Presentazione di Rosa Maiello. Roma: Associazione italiana biblioteche.

Mai, Jens-Erik. 1998. “Organization of knowledge: an interpretive approach”. In *Information Science at the Dawn of the Next Millennium*, edited by Elaine G. Toms, D. Grant Campbell and Judy Dunn, 231-242. Toronto: Canadian Association for Information Science.

Serrai, Alfredo. 1980. *Sistemi bibliotecari e meccanismi catalografici*. Roma: Bulzoni.

Ventura, Roberto. 2010. “La biblioteca tra linguaggio e rappresentazione. Per un’indagine filosofica della biblioteconomia”. *JLIS.it* 1 (1):41-61.

Wittgenstein, Ludwig. 1967 (ed. orig. 1953). *Ricerche filosofiche*, edizione italiana a cura di Mario Trinchero. Torino: Einaudi.



## Identifiers within conceptual models and standards

Laura Manzoni<sup>(a)</sup>

a) University of Florence, <https://orcid.org/0000-0002-2883-7916>

**Contact:** Laura Manzoni, [lauramanzoni1993@live.it](mailto:lauramanzoni1993@live.it)

**Received:** 26 April 2022; **Accepted:** 16 June 2022; **First Published:** 15 September 2022

### ABSTRACT

Identifiers are numeric or alphanumeric strings of characters permanently and uniquely associated with a resource to ensure stable access to it and its metadata. They play a fundamental role in the semantic web as they guarantee interoperability between different systems. Since their emergence, libraries and other institutions have been aware of their importance in data quality control processes, for the realization of authority control and universal bibliographic control, and have dedicated specific chapters for their registration within conceptual models, cataloguing standards and have provided specific fields in cataloguing formats.

### KEYWORDS

Identifiers; Standards; Conceptual Models; Cataloguing Standards.

## Gli identificatori all'interno di modelli concettuali e standard

### ABSTRACT

Gli identificatori sono stringhe numeriche o alfanumeriche di caratteri associate in modo permanente e univoco a una risorsa per garantire l'accesso stabile a essa e ai suoi metadati. Svolgono un ruolo fondamentale nel web semantico poiché garantiscono l'interoperabilità tra sistemi diversi. Fin dalla loro comparsa le biblioteche e le altre istituzioni della memoria registrata sono state consapevoli della loro importanza nei processi di controllo di qualità dei dati, per la realizzazione dell'authority control e del controllo bibliografico universale e hanno dedicato a essi capitoli specifici per la loro registrazione all'interno di modelli concettuali, standard catalogafici e hanno previsto campi appositi nei formati di catalogazione.

### PAROLE CHIAVE

Identificatori; Standard; Modelli concettuali; Standard di catalogazione.

## Gli identificatori

Gli identificatori sono delle stringhe numeriche o alfanumeriche di caratteri associati in maniera univoca alle risorse per garantire l'accesso stabile a esse. Sono normalmente attribuiti e registrati da agenzie bibliografiche nazionali e da istituzioni commerciali. Essi devono essere costruiti sulla base di alcuni principi quali l'unicità e la permanenza affinché possano svolgere le loro funzioni in modo efficace. L'unicità determina che a ogni identificatore si riferisca una sola risorsa mentre la permanenza garantisce la stabilità di un identificatore nel tempo. Fondamentale è, inoltre, l'univocità che rende un'entità inequivocabilmente riconoscibile all'interno di uno specifico contesto di produzione, di un dominio.

Gli identificatori hanno sempre svolto un ruolo fondamentale in ambito informatico in tutti i sistemi di elaborazione delle informazioni e, in particolare, nelle basi dati relazionali. A partire dagli anni Settanta del Novecento hanno cominciato a essere utilizzati dagli editori per identificare in maniera univoca le proprie pubblicazioni o parti di esse. Nel 1967 venne introdotto in Inghilterra e negli Stati Uniti SBN (Standard Book Number) da cui derivò l'ISBN (International Standard Book Number) oggi utilizzato per l'identificazione dei libri e di altre risorse come le carte geografiche, i cd, i dvd, gli audiolibri ecc. Nel 1975 nacque l'ISSN (International Standard Serial Number) per le pubblicazioni seriali. Successivamente vennero introdotti identificatori rivolti ad altre tipologie di prodotti editoriali come l'ISRC (International Standard Recording Code) e l'ISMN (International Standard Music Number) che identificano oggetti definibili su base materiale e consolidati sul mercato. Già nel corso degli anni Novanta la comunità dell'informazione decise che era arrivato il momento di identificare anche le opere intellettuali indipendentemente dalla loro forma di presentazione. Nacquero così identificatori come l'ISWC (International Standard Musical Work Code) e l'ISTC (International Standard Text Code). Tuttavia, fu soprattutto in seguito allo sviluppo del digitale che si avvertì l'urgenza di identificare nuove tipologie di risorsa, i loro metadati e i loro link (Vitiello 2004, p. 70). Vennero quindi elaborati identificatori come il DOI (Digital Object Identifier), l'NBN (National Bibliography Number) e l'ARK (Archival Resource Key). Accanto a essi, infine, sono stati sviluppati progetti specifici volti all'identificazione di persone, famiglie ed enti pubblicamente conosciuti nella filiera della produzione intellettuale, culturale e scientifica. Tra i più famosi vanno ricordati il VIAF (Virtual International Authority File), l'ISNI (International Standard Name identifier) e ORCID (Open Research Contributor ID) (Concedda 2015).

## L'utilizzo degli identificatori in ambito catalografico

Nel corso dell'ultimo decennio gli identificatori creati dalle diverse agenzie bibliografiche si sono moltiplicati e hanno assunto un ruolo fondamentale nella condivisione dei dati bibliografici. L'identificazione, infatti, costituisce l'attività caratterizzante il web semantico all'interno del quale ogni oggetto è rappresentato mediante un URI (Uniform Resource Identifier) che rende la risorsa disponibile come link utilizzabile e dereferenziabile. La presenza delle biblioteche nel web e lo sviluppo incontrollato dell'editoria digitale, quindi, hanno reso gli identificatori strumenti fondamentali nei processi di controllo della qualità dei dati, per la realizzazione dell'authority control e del controllo bibliografico universale. Il controllo d'autorità è sempre stato re-

alizzato attraverso la scelta di una forma autorizzata del nome dell'entità descritta (una persona, un ente, un'opera ecc.), ma da quando le procedure di catalogazione sono state automatizzate, a partire dagli anni Sessanta, è apparso chiaro il vantaggio di utilizzare numeri di controllo anziché stringhe testuali vincolate a convenzioni linguistiche e culturali. Progetti internazionali come VIAF e ISNI sono una chiara testimonianza di questa trasformazione (Moi 2021, p. 376). Essi consentano l'aggregazione di dati provenienti da fonti diverse in cui ogni etichetta rappresenta una forma del nome associata a un'entità in una determinata lingua e scrittura ed è valida in uno specifico contesto di applicazione. Le etichette vengono riunite mediante l'uso di identificatori numerici univoci utilizzati dalle macchine per arrivare alla descrizione della risorsa e consentendo al lettore di scegliere una delle forme varianti del nome dell'entità comprese nel medesimo cluster. Se, per esempio, facciamo una ricerca sul VIAF con il nome Giacomo Leopardi viene restituito il seguente cluster (un insieme, un aggregato) di forme varianti del nome dell'entità utilizzate da diverse agenzie bibliografiche seguite dal VIAF ID, ossia l'identificatore univoco associato a questa entità:

- Leopardi, Giacomo, 1798-1837
- Leopardi, Giacomo
- Leopardi, Giacomo, conte, 1798-1837
- Γζάκομο Λεοπάρντι
- יִדְרֵפּוּאֵל, וּמוֹקֵג, 1798-1837
- Леопарди, Дж. 1798-1837 Джакомо
- Leopardi, G. 1798-1837 Giacomo
- VIAF ID: 12311353 ( Personal )

Per sfruttare al meglio le potenzialità degli identificatori le agenzie bibliografiche nazionali stanno impegnandosi nel definire delle *policy* specifiche per la loro gestione, tenendo conto del quadro generale delle attività da compiere sia all'interno della biblioteca, integrando la gestione degli identificatori nel lavoro svolto quotidianamente dai catalogatori, sia all'esterno in relazione all'uso degli identificatori da parte degli utenti finali. In particolare, è necessario tener conto della varietà degli identificatori che ogni biblioteca deve gestire e del tipo di responsabilità che può esercitare su di essi: se può occuparsi direttamente della loro attribuzione o se deve limitarsi alla loro registrazione inviando i dati necessari per chiederne l'attribuzione all'agenzia internazionale che li gestisce (Boulet 2022).

## **Gli identificatori all'interno dei modelli concettuali e negli standard di catalogazione**

Gli identificatori, a partire dagli anni Settanta del secolo scorso, hanno acquisito un'importanza sempre maggiore in ambito catalografico. Nella prima edizione di ISBD(M), *International Standard Bibliographic Description for Monographic Publications* del 1974 era prevista un'area specifica, l'area 7, dedicata alla registrazione dell'ISBN. Nel 1977 con la pubblicazione di ISBD(S), *International Standard Bibliographic Description for Serials* venne introdotta una nuova area dedicata alle diverse tipologie di risorsa, che divenne l'area 3, mentre la precedente area 3 divenne l'area 4 e tutte le altre

scalarono di una cifra modificando la struttura dello standard. L'identificatore, che in questo caso era l'ISSN, veniva trattato in area 8. Sempre nel 1997 uscì ISBD(G), *International Standard Bibliographic Description General* con l'obiettivo di costituire una guida generale a cui avrebbero dovuto uniformarsi tutti gli standard successivi relativi alle diverse tipologie di materiale (Guerrini 2020, p. 143). Essendo uno standard generale ISBD(G) non definì l'area 8 come *area dell'ISBN* o *area dell'ISSN*, ma utilizzò la denominazione: *area del numero standard e delle condizioni di disponibilità*, intendendo con numero standard qualsiasi tipo di identificatore che poteva comparire sulla risorsa. La locuzione *numero standard* continuò a essere utilizzata fino alla pubblicazione di *ISBD Edizione consolidata* del 2011, in cui la denominazione dell'area 8 si modificò significativamente in *Area dell'identificatore della risorsa e delle condizioni di disponibilità*, con la dichiarazione esplicita della funzione dell'identificatore; l'area divenne obbligatoria, aspetto che dimostra ulteriormente l'importanza di questo tipo di informazione.

Anche le AACR2, *Anglo American Cataloguing Rules 2nd Edition* pubblicate nel 1978 e riviste nel 2002 e nel 2005, erano strutturate seguendo il modello ad aree di ISBD e prevedevano un'area specifica per la registrazione del *numero standard*.

Il concetto di identificatore in ambito catalografico, tuttavia, fece la sua comparsa consapevole tra la fine degli anni Novanta del Novecento e i primi anni Duemila all'interno di FRBR, *Functional Requirements for Bibliographic Records* e degli altri modelli della famiglia *Functional Requirements* (FR). FRBR venne pubblicato nel 1998 e considerava gli identificatori importanti attributi per l'identificazione della manifestazione e dell'item. FRAD, *Functional Requirements for Authority Data*, pubblicato nel 2009, considerava l'*identificatore* come un'entità che, analogamente al *nome*, può essere associato a un'altra entità allo scopo di differenziarla da tutte le altre appartenenti a un determinato dominio. A differenza di FRBR, quindi, non concepiva più l'identificatore come un semplice attributo, ma come un'entità che poteva essere collegata a entità bibliografiche, ma anche a persone, famiglie ed enti.

Con la pubblicazione di FSRAD, *Functional Requirements for Subject Authority Data*, nel 2010, l'identificatore venne considerato come una sottoclasse dell'entità *Nomen* che rappresenta una designazione costituita una sequenza di segni utilizzata per riferirsi a un *Thema*, ossia "ogni entità utilizzata come soggetto di un'opera" (FRSAD 2010, p. 15).

Lo stesso anno venne pubblicata la prima edizione delle linee guida RDA, *Resource Description and Access* basate sui modelli concettuali della famiglia FR e in particolare FRBR e FRAD. Come FRBR, RDA considera gli identificatori degli attributi. Essi però non vengono registrati solo per la manifestazione e l'item, ma per tutte le entità descritte all'interno dello standard e quindi anche l'opera, l'espressione, la persona, la famiglia e l'ente. La prima edizione di RDA, inoltre, considera l'identificatore come un *elemento essenziale* della descrizione riconoscendo l'importanza che esso ha acquisito all'interno del contesto del web semantico. Gli elementi essenziali, infatti, devono sempre essere presenti al fine di considerare una descrizione conforme allo standard. Essi sono stati individuati in base alla loro capacità di svolgere le principali funzioni utente corrispondenti a:

- identificare e scegliere una manifestazione;
- identificare e scegliere le opere e le espressioni contenute in una manifestazione;
- identificare il creatore di un'opera;
- trovare una persona, un ente o una famiglia associati a una risorsa.

Nel 2017 viene pubblicato il nuovo modello concettuale IFLA *Library Reference Model* (LRM) che integra e sostituisce i modelli precedenti. Analogamente a FRAD e FRASD considera gli identificatori come delle entità e, in particolare, come dei *nomen*:

un identificatore è un tipo di *nomen* destinato a essere persistente e unico nell'ambito di uno specifico dominio di applicazione, quali gli identificatori di pubblicazioni di tipo specifico o gli identificatori delle *persone*, in modo che le istanze dell'entità si possano specificamente identificare e vi si possa fare riferimento senza ambiguità. Ciò che distingue un identificatore da altri *nomen* è che il valore dell'attributo *stringa del nomen* di un identificatore non può essere identico al valore dell'attributo *stringa del nomen* di qualsiasi altro *nomen* all'interno di un determinato sistema (naturalmente, altri *nomen* al di fuori di quel sistema potrebbero avere lo stesso valore dell'attributo *stringa del nomen*). Gli identificatori sono generalmente assegnati in base a regole stabilite da agenzie autorizzate quali, ad esempio, le agenzie di registrazione di registrazione degli identificatori ISO e gli uffici della pubblica amministrazione per l'identificazione univoca di cittadini e residenti. L'ambito di un sistema basato su identificatori può essere generico (come per l'URI) o altamente specializzato (numeri di catalogo delle opere di un compositore) (IFLA 2020, p. 88).

In IFLA LRM, dunque, gli identificatori sono una tipologia specifica di *nomen* che si caratterizza per l'uso di una *stringa del nomen* univoca all'interno di un determinato contesto. La *stringa del nomen* consiste nella "combinazione di segni che forma una denominazione associata a un'entità tramite il *nomen*" (IFLA 2020, p. 53). Per esempio, la stringa di cifre 20150601 può costituire un attributo *stringa del nomen* di un ISSN. Il collegamento tra la *stringa del nomen* e l'entità a cui si riferisce da origine al *nomen* che, quindi, si presenta contemporaneamente come un'entità e come una relazione reificata tra entità.

IFLA LRM ha determinato una profonda revisione delle linee guida RDA che sono state modificate per allinearsi al nuovo modello concettuale. La seconda edizione dello standard è stata pubblicata nel dicembre 2020. Le trasformazioni introdotte hanno influito sul modo di registrare gli identificatori che non costituiscono più degli attributi per le entità ma dei *nomen* associati a esse mediante relazioni di denominazione. Il seguente schema mostra come può essere registrata la relazione tra un'entità persona e l'identificatore a essa associato.

The image shows a screenshot of a library system interface displaying an IRI (Internationalized Resource Identifier) element reference. The interface is light blue and contains the following information:

- Element Reference**: <http://rdaregistry.info/Elements/a/P50094>
- Domain**: Person →
- Range**: Nomen →
- Alternate labels**: has identifier for person
- IFLA LRM**
- MARC 21 Authority**

L'entità persona rappresenta il dominio da cui parte la relazione 'ha identificatore per persona', mentre il *nomen*, in questo caso un identificatore, rappresenta il codominio, ossia il punto di arrivo della relazione.

La normativa catalografica italiana al pari di quella internazionale riuscì a cogliere immediatamente l'importanza degli identificatori nella descrizione bibliografica. Le RICA, *Regole italiane di catalogazione per autori* del 1979 prevedevano la registrazione dell'ISBN o dell'ISSN nell'area delle note. Le successive REICAT, *Regole italiane di catalogazione* del 2009, affermano nell'introduzione che "devono costituire elementi di accesso i numeri e i codici identificativi attribuiti alle pubblicazioni e alle entità (p. es. a opere, a esemplari, a biblioteche, etc.) sia quando sono riconosciuti da standard internazionali o nazionali sia quando sono comunque largamente utilizzati. Quando esistono più codici, anche errati, la ricerca deve essere assicurata da tutti" (REICAT 2009, 0.4.2). A differenza di RDA in cui ogni identificatore è associato alla specifica entità a cui si riferisce REICAT concentrano la trattazione degli identificatori nel capitolo 4.8 *Area dei numeri identificativi*. Il paragrafo 4.8.1 è dedicato alla registrazione degli identificatori associati alla pubblicazione e, in particolare, ai numeri standard definiti dall'ISO quali: ISBN, ISSN, ISMN, ISRN, ISRC e ISAN. Il paragrafo 4.8.2 si concentra sui numeri e i codici attribuiti dall'editore, dal distributore o dal tipografo per identificare univocamente la pubblicazione o una delle sue parti. Infine, il paragrafo 4.8.3 riguarda le pubblicazioni antiche per le quali viene considerato come elemento identificativo l'impronta. Essa consiste in quattro gruppi formati da quattro caratteri ciascuno rilevati in punti predeterminati della pubblicazione seguiti da un numero o da una lettera che indica la fonte del terzo e del quarto gruppo di caratteri o dalla data di pubblicazione e da una lettera che indica come la data stessa si presenta.

## Identificatori e formati di catalogazione

Nonostante i modelli concettuali e gli standard catalografici abbiano compreso già da tempo l'importanza degli identificatori "ben più complessa è la loro applicazione a flussi e pratiche catalografiche tradizionali, ancora fortemente vincolate alla strutturazione del record e all'utilizzo del formato MARC [MACHINE REABLE CATALOGING]" (Moi 2021, p. 374). In questa fase di transizione, in attesa che i dati bibliografici vengano definitivamente prodotti in linked data, un ruolo fondamentale è svolto dal Program for Cooperative Cataloging che ha avviato una serie di sperimentazioni guidate dal PCC Task Group on URIs in MARC. Il Gruppo di lavoro si è posto l'obiettivo di definire delle modalità di arricchimento dei record MARC attraverso l'inserimento di URI autorevoli provenienti da fonti esterne al fine di garantire nel prossimo futuro una transizione dal MARC ai linked data più agevole possibile. Il MARC21 prevedeva già da tempo campi specifici dedicati alla registrazione di identificatori quali ISBN (020), ISSN (022) e altri numeri standard come l'ISTC e l'ISMN (024). Il sottocampo \$0, *Authority record control number or standard number* viene utilizzato dal 2007 per la registrazione del numero di controllo associato all'entità descritta o un URI contenente il numero di controllo come token (Hieh; Reese 2015, p. 212). Per esempio, nel record:

- 100 1# \$a Stravinsky, Igor,
- \$d 1882-1971,
- \$e composer.
- \$4<http://id.loc.gov/vocabulary/relators/cmp>
- \$0 <http://id.loc.gov/authorities/names/n79070061>

\$0 rappresenta la fonte da cui è stata ripresa la forma autorizzata del nome dell'entità persona registrata nel sottocampo \$a.

Nel 2019 il PCC Task Group on URIs in MARC ha approvato il sottocampo \$1 che è andato ad arricchire numerosi campi MARC con lo scopo di accogliere i *real world object (RWO) URIs*, ossia degli URI che identificano l'entità in se come oggetto del mondo reale. Per esempio:

- 257 ## \$a Korea (South)
- \$2 naf
- \$0 <http://id.loc.gov/authorities/names/n79126802>
- \$1 <http://vocab.getty.edu/tgn/7000299-place>

Il sottocampo \$1 contiene un URI che rimanda direttamente all'entità luogo riportata nel sottocampo \$a, mentre l'URI registrato in \$0 fa riferimento alla fonte autorevole da cui è tratta la forma del nome dell'entità.

La distinzione tra \$1 e \$0 consente di differenziare gli oggetti del mondo reale dalla loro descrizione. Per evitare confusione, infatti, è necessario utilizzare URI differenti in modo che l'oggetto e la descrizione siano chiaramente identificati. Per l'occhio umano è impossibile riconoscere la differenza tra un URI che rappresenta un RWO e un URI che rimanda alla sua descrizione. Per poterli distinguere è necessario utilizzare un programma che renda evidente lo scambio di messaggi tra client e server come Web Sniffer,<sup>1</sup> oppure è possibile utilizzare due strategie riportate nel documento *Cool URIs for the semantic web* del World Wide Web Consortium (W3C):

- il codice di stato *303 See Other*
- *Hash URI*

Il primo sistema utilizza il meccanismo della redirectione per restituire i dati RDF come descrizione della risorsa. A un client interessato a un oggetto del mondo reale il server risponde con un codice *HTTP 303 See Other* e con l'URI di un documento web che descrive l'oggetto richiesto. Per esempio, se il client chiede <https://www.wikidata.org/entity/Q200092> il server risponde *303 See Other* e invia al client l'identificatore URI relativo alla rappresentazione della risorsa, ovvero <https://www.wikidata.org/wiki/Special:EntityData/Q200092>. In questo modo il server dichiara che <https://www.wikidata.org/entity/Q200092> è un URI relativo a un RWO e che <https://www.wikidata.org/wiki/Special:EntityData/Q200092> è un URI relativo alla sua rappresentazione. Se il client avesse richiesto <https://www.wikidata.org/wiki/Special:EntityData/Q200092> il server avrebbe risposto *200 ok* comunicando così al client che la sua richiesta era stata ricevuta con successo e veniva fornito in risposta il contenuto richiesto. Quindi, il browser che usiamo quotidianamente quando incappa in un codice di stato 303 in maniera trasparente invia all'utente il contenuto dereferenziato del secondo indirizzo URI.

Il sistema *Hash URI* consiste nell'utilizzo di URI che contengono un frammento ossia una parte relativa alla specifica risorsa descritta separata dal resto dell'URI mediante il simbolo *hash* (#), per

---

<sup>1</sup> <https://websniffer.cc>

esempio, [https://en.wikipedia.org/wiki/Bonnie\\_and\\_Clyde#Clyde\\_Barrow](https://en.wikipedia.org/wiki/Bonnie_and_Clyde#Clyde_Barrow) o [https://en.wikipedia.org/wiki/Bonnie\\_and\\_Clyde#Bonnie\\_Parker](https://en.wikipedia.org/wiki/Bonnie_and_Clyde#Bonnie_Parker). Quando il client vuole recuperare un *Hash URI* il protocollo http prevede che la parte preceduta dal simbolo # sia rimossa prima di chiedere l'URI al server (in questo caso verranno rimosse le parti #Clyde\_Barrow e #Bonnie\_Parker). Questo significa che un URI che include un *Hash* non può essere recuperato direttamente, ma vengono restituite tutte le descrizioni delle risorse appartenenti a uno specifico dominio indicato nell'URI stesso ([https://en.wikipedia.org/wiki/Bonnie\\_and\\_Clyde](https://en.wikipedia.org/wiki/Bonnie_and_Clyde)).

La scelta tra i due sistemi non può dipendere da fattori oggettivi quanto piuttosto da considerazioni di opportunità dipendenti da specifici contesti di utilizzo: “il limite più rilevante dell'Hash URI è la quantità di dati che l'operazione di dereferenziazione restituisce. Infatti, di solito questa tecnica prevede la restituzione di un documento che contiene, al suo interno, una porzione (la parte locale o relativa) di informazione che corrisponde all'URI dereferenziato. È preferibile utilizzare URI 303 nel caso d'interrogazione di dataset corposi e utilizzare Hash URI per l'interrogazione di dataset poco consistenti” (Guerrini, Possemato 2015, p. 42).

Accanto all'arricchimento dei record MARC attraverso gli URI, nel corso degli ultimi anni sono stati sviluppati altri progetti per favorire il passaggio verso una descrizione interamente basata sui linked data. In particolare, va ricordata la Bibliographic Framework Initiative (BIBFRAME) che ha preso avvio nel 2011 e ha definito tra i suoi obiettivi principali quello di consentire il riutilizzo di milioni di record MARC prodotti dalle biblioteche nel contesto del web semantico attraverso i linked data e favorendo il riassetto dei dati delle attuali registrazioni MARC in nuove architetture coerenti che consentano una catalogazione collettiva a livello più granulare. Nel novembre 2012 la Library of Congress ha rilasciato la versione 1.0 di BIBFRAME, successivamente aggiornato nel 2016 con la pubblicazione della versione 2.0. BIBFRAME consiste di classi e proprietà RDF. Le classi principali sono *opera*, *istanza* e *item* a cui si aggiungono altre classi e sottoclassi. Le proprietà descrivono le caratteristiche della risorsa analizzata e le relazioni con altre risorse. Il nuovo modello attribuisce agli identificatori un ruolo centrale poiché riconosce che solo attraverso l'identificazione delle risorse mediante identificatori stabili, univoci e comprensibili da una macchina è possibile garantire la partecipazione delle biblioteche alla costruzione del web dei dati. Inoltre, il loro uso favorisce la condivisione dei dati tra le biblioteche riducendo i tempi necessari al lavoro d'autorità che ha sempre costituito uno degli aspetti più costosi della descrizione bibliografica:

BIBFRAME's use of controlled identifiers over the MARC format's reliance on controlled text strings for entity description will lessen considerably the time and costs associated with maintenance of authority data. One controlled string of data may appear in thousand of MARC records. If that controlled string of data changes, a fairly common occurrence, all MARC records containing that controlled string of data need to be changed as well. Maintenance of MARC records can be costly and time-consuming. A controlled identifier does not change, even if the controlled string of data associated with that identifier changes; BIBFRAME thus reduces the time and costs of bibliographic maintenance.<sup>2</sup>

---

<sup>2</sup> <https://www.librarianshipstudies.com/2017/12/bibframe.html>



Inoltre, l'ampio uso degli identificatori all'interno di BIBFRAME aumenterà la visibilità delle collezioni delle biblioteche di tutto il mondo. Essi, a differenza delle stringhe di testo adottate nel MARC non dipendono dalla lingua e ciò facilita la più ampia condivisione dei dati bibliografici. Il trattamento degli identificatori all'interno di BIBFRAME è stato modificato con il passaggio dalla versione 1.0 alla 2.0. Mentre la versione 1.0 definiva un'unica classe di identificatori considerando le singole tipologie di identificatore come proprietà, la versione 2.0 definisce un'unica proprietà per gli identificatori "bf:identifiedBy" e considera le diverse tipologie di identificatore come delle sottoclassi della classe principale identificatore. Quindi tipologie specifiche come ISBN, ISAN, DOI, LCCN ecc. sono delle sottoclassi di "identificatore" (bf:Isbn; bf:Isan; bf:Doi; bf:Lccn). La rappresentazione di queste tipologie come classi piuttosto che proprietà determina dei vantaggi in termini di riusabilità ed efficienza della ricerca (rende i dati più facilmente interrogabili).

## Gli identificatori all'interno dei modelli e degli standard archivistici e museali

Accanto alle biblioteche anche gli archivi e i musei dimostrano di aver colto l'importanza degli identificatori per lo scambio e il riuso delle informazioni dedicando particolare attenzione al loro trattamento all'interno di standard e modelli concettuali. In ambito archivistico i principali standard adottati a livello internazionale come ISAD(G), *General International Standard Archival Description*; ISAAR(CPF), *International Standard Archival Authority Records for Corporate Bodies, Persons and Families*; ISDF, *International Standard for Describing Function* e ISDIAH, *International Standard for Describing Institution with Archival Holdings*, elaborati tra la fine degli anni Ottanta del secolo scorso e i primi anni Duemila prevedevano specifiche aree della descrizione destinate ad accogliere gli identificatori dell'entità descritta. ISAD(G), per esempio, consentiva di registrare l'identificatore all'interno dell'*Area dell'identificazione*. Esso poteva essere rappresentato dalla segnatura archivistica del documento o da un altro codice identificativo. Inoltre, lo standard precisava che l'identificatore insieme ad altri elementi della descrizione come la segnatura, la denominazione o titolo, il soggetto produttore, la data, la consistenza dell'unità di descrizione e il livello di descrizione era considerato essenziale per lo scambio di informazioni a livello internazionale. L'importanza degli identificatori nella descrizione archivistica è stata ulteriormente ribadita all'interno di RiC-CM, *Record in Context. A conceptual model for archival description*, un modello concettuale e un'ontologia per la descrizione archivistica elaborata dall'Expert Group on Archival Description (EGAD) tra il 2012 e il 2016. Gli identificatori vengono definiti come delle proprietà condivise da qualsiasi entità riconosciuta dal modello. In particolare, vengono distinte due categorie di identificatori:

- RiC-P1<sup>3</sup> identificatori globali persistenti: sono degli identificatori globali unici e risolvibili in modo persistente per le entità e si presentano come degli URI, per esempio: <http://n2t.net/ark:/99166/w6v1266v> è un identificatore ARK associato a un record; <http://isni.org/0000000073572182> è un identificatore ISNI per una persona;
- RiC-P2 identificatori locali: sono degli identificatori assegnati localmente a un'entità, per esempio: BUD-01-F002 è un numero di classificazione assegnato a un file aziendale; NAS1/A/1.1 è un identificatore locale associato a un record.

---

<sup>3</sup> Denominazione prevista da RiC-CM per le proprietà.

In ambito museale l'ontologia maggiormente diffusa a livello internazionale è CIDOC CRM, *Conceptual Reference Model* elaborata a partire dal 1994 dall'International Committee for Documentation (CIDOC) dell'International Council of Museums (ICOM) con l'obiettivo di fornire un modello di riferimento e uno standard che musei e altre istituzioni culturali possono utilizzare per descrivere le loro collezioni e favorire la condivisione delle informazioni. Il modello prevede delle classi principali che rappresentano le entità a cui sono associate delle proprietà che le descrivono e stabiliscono delle relazioni tra esse. Gli identificatori sono una sottoclasse della classe principale *Denominazione* (E41) e comprendono stringhe e codici assegnati alle istanze e alle entità previste dal modello per identificarle in modo univoco e permanente nel contesto di una o più organizzazioni. Essi possono coincidere con numeri di inventario, codici di registrazione ecc. Indirizzi postali, numeri di telefono, URL e indirizzi di posta elettronica sono altresì considerati identificatori utilizzati nei servizi di trasporto tra clienti (CIDOC CRM 2022, p. 87).

## Conclusioni

Le diverse istituzioni della memoria registrata hanno saputo cogliere l'importanza degli identificatori per la descrizione e l'accesso alle risorse fin dalla loro comparsa. Inoltre, in seguito allo sviluppo del web e delle nuove tecnologie hanno messo in atto specifiche politiche di gestione al fine di sfruttare le loro potenzialità per lo scambio, la condivisione di dati e per il controllo d'autorità. Il peso attribuito agli identificatori si coglie, inoltre, dall'analisi dei modelli concettuali e degli standard prodotti in ambito biblioteconomico, archivistico e museale in cui gli identificatori vengono spesso considerati come elementi essenziali per la descrizione delle risorse sia che vengano considerati come attributi o proprietà di una risorsa, sia come sottoclassi di entità principali, per esempio, il *nomen* in IFLA LRM. In questo caso è necessario stabilire una relazione di denominazione tra l'identificatore e l'entità a cui si riferisce. Gli standard prevedono la possibilità di registrare sia identificatori globali supportati da standard ISO e diffusi a livello internazionale sia identificatori locali, ma riconoscono la superiorità dei primi per la condivisione e il riuso dei dati. Per arrivare a sfruttare a pieno le potenzialità degli identificatori, tuttavia, non è sufficiente riconoscerne l'importanza nell'ambito della descrizione bibliografica e documentale, ma necessario che i formati di catalogazione e metadattazione oggi a disposizione delle varie comunità si trasformino al fine di allinearsi ai linguaggi e alle tecnologie del web semantico.

## Riferimenti bibliografici

Boulet, Vincent. 2022. "How to build an 'Identifier's policy': the BnF use case". *JLIS.it* 13(1): [177]-184. DOI: <https://doi.org/10.4403/jlis.it-12768>.

Commissione permanente per la revisione delle regole italiane, ed. 2009. *Regole italiane di catalogazione REICAT*. Roma: ICCU.

Concedda, Flavia. 2015. "ISSN e I suoi fratelli: la lunga marcia dei codici internazionali: prima parte: UNESCO, ISO e il primogenito". *AIB Studi* 55(2): 295-308.

Concedda, Flavia. 2015. "ISSN e i suoi fratelli: la lunga marcia dei codici internazionali: parte seconda: il CNR e la storia italiana dell'ISSN. Gli altri sistemi identificativi". *AIB Studi* 55(3): 439-457.

Guerrini, Mauro. 2020. *Dalla catalogazione alla metedatazione. Tracce di un percorso*. Roma: Associazione italiana biblioteche.

Guerrini, Mauro, Tiziana Possemato. 2015. *Linked data per biblioteche, archivi e musei*. Milano: Editrice Bibliografica.

Heih, Jackie and Terry Reese. 2015. "The importance of identifiers in the new web environment and using the Uniform Resource Identifier (URI) in subfield zero (\$0): a small step that is actually a big step". *Journal of Library Metadata* 15(3-4): 208-226. <https://doi.org/10.1080/19386389.2015.1099981>.

IFLA (International Federation of Library Association and Institutions). (2010). *Functional Requirements for Subject Authority Data (FRSAD). A conceptual model*. Edited by Marcia Lei Zeng, Maja Žumer, Athena Salaba. Berlin: De Gruyter.

IFLA (International Federation of Library Association and Institutions). 2020. *IFLA Library Reference Model. Un modello concettuale per le informazioni bibliografiche*. Edited by IFLA Functional Requirements for Bibliographic Records (FRBR) Review Grup, Pat Riva, Patrick Le Boeuf, Maja Žumer, italian edition edited by Istituto centrale per il catalogo unico (ICCU). Roma: ICCU.

Moi, Alessandra. 2021. "Da universale a globale. Il controllo bibliografico nel web semantico". In *La trasmissione della conoscenza registrata*, a cura di Carlo Bianchini e Lucia Sardo, 369-378. Milano: Editrice Bibliografica.

Vitiello, Giuseppe. (2004). "L'identificazione degli identificatori". *Biblioteche oggi* 2: 67-80.

W3C Interest Group. 2008. *Cool URI for the semantic web*. <https://www.w3.org/TR/cooluris/>.

## Conceptual and linguistic changes: document, resource, and item

Denise Biagiotti<sup>(a)</sup>

a) <https://orcid.org/0000-0003-2815-8996>

**Contact:** Denise Biagiotti, [denisebiagiotti79@gmail.com](mailto:denisebiagiotti79@gmail.com)

**Received:** 5 June 2022; **Accepted:** 30 June 2022; **First Published:** 15 September 2022

### ABSTRACT

The bibliographic universe has radically changed in a short time; the major innovations, mainly related to the digital environment, have had consequences in terms of terminology. This contribution offers an analysis of three specific international cataloguing terms: *document*, *resource*, and *item*. The terms *document* and *resource* constitute an example of terminological evolution; *item*, introduced in cataloguing by FRBR as the fourth entity of Group I, highlights the difficulties of transposition from one linguistic system to another, and the multiple terminological meanings linked to different disciplinary contexts.

### KEYWORDS

Terminological evolution, Document, Resource, Item.

## Cambiamenti concettuali e linguistici: documento, risorsa e item

### ABSTRACT

L'universo bibliografico si è radicalmente modificato in poco tempo; le maggiori innovazioni, legate soprattutto all'ambiente digitale, hanno comportato conseguenze sul piano terminologico. Questo contributo offre un'analisi dell'evoluzione di tre specifici termini della catalogazione internazionale: *documento*, *risorsa* e *item*. I lemmi *documento* e *risorsa* costituiscono un esempio di evoluzione terminologica; *item*, introdotto in ambito catalografico da FRBR come la quarta entità del I Gruppo, consente di evidenziare le difficoltà di trasposizione da un sistema linguistico a un altro e le molteplici accezioni terminologiche legate a contesti disciplinari diversi.

### PAROLE CHIAVE

Evoluzione terminologica, Documento, Risorsa, Item.

Ringrazio Mauro Guerrini per aver seguito l'evoluzione del saggio fornendo preziosi suggerimenti.

© 2022, The Author(s). This is an open access article, free of all copyright, that anyone can freely read, download, copy, distribute, print, search, or link to the full texts or use them for any other lawful purpose. This article is made available under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. JLIS.it is a journal of the SAGAS Department, University of Florence, Italy, published by EUM, Edizioni Università di Macerata, Italy, and FUP, Firenze University Press, Italy.



“I limiti del mio linguaggio costituiscono i limiti del mio mondo. Tutto ciò che io conosco è ciò per cui ho delle parole”.

Ludwig Wittgenstein, *Tractatus logico-philosophicus*, 1921

## Premessa

L'universo bibliografico si è radicalmente modificato in poco tempo; le novità, legate soprattutto all'ambiente digitale, hanno comportato conseguenze sul piano terminologico. Poiché la terminologia è uno degli elementi fondamentali di ogni scienza, è importante definire il significato delle parole utilizzate; ciò significa partecipare attivamente all'intenso dibattito in corso in ambito internazionale nel dominio LIS volto a delinearne, de facto, un glossario della disciplina, con la volontà di promuovere l'adeguata definizione di concetti e termini, nell'ottica della standardizzazione e della condivisione dei dati bibliografici.

Questo contributo si occupa dell'evoluzione di tre termini specifici della catalogazione internazionale: *documento*, *risorsa* e *item*. Le fonti principali sono costituite dai modelli concettuali FRBR e IFLA LRM, lo standard per la descrizione bibliografica ISBD, la *Dichiarazione di Principi internazionali di catalogazione* (ICP) e lo standard di contenuto RDA; l'ordine seguito nell'analisi è quello cronologico dei documenti internazionali che li definiscono, nell'arco di tempo compreso tra la loro redazione e pubblicazione. Nel caso specifico di *item* sono state affrontate le problematiche legate alla sua resa traduttiva e le possibili motivazioni della scelta originaria del lemma. Sono state consultate, inoltre, fonti delle scienze biblioteconomiche e dell'informazione (manuali, saggi e norme ISO) e risorse linguistiche, generaliste e specialistiche, di area italiana e anglosassone. I modelli FRBR e IFLA LRM hanno fortemente sollecitato l'impiego della nuova terminologia, mentre ICP ha deliberatamente escluso termini usati per molti decenni, come *unità bibliografica*, *intestazione*, *rinvio* e *titolo uniforme*, ritenuti inadeguati al nuovo contesto e perciò sostituiti rispettivamente da *manifestazione*; *punto d'accesso autorizzato*; *punto d'accesso controllato*; *forma variante del nome*; *forma autorizzata del nome*; *nome*. La pubblicazione di ISBD Preliminary e Consolidated Edition ha portato a un aggiornamento terminologico legato alla necessità di fondere le precedenti ISBD; ad esempio, la formula *non-book material* non è stata mantenuta in quanto generica. RDA dialoga con il web anche sul piano terminologico.

## Da documento a risorsa

I lemmi *documento* e *risorsa* costituiscono un esempio di evoluzione terminologica. Il termine “documento” deriva dal verbo latino *docere* (radice comune di termini come “docente” e “dottrina”) e, anche se nel contesto digitale ha perso la connotazione testuale, copre un'area semantica legata ai significati di comunicare, insegnare, dimostrare e trasmettere elementi informativi. Il *documento* possiede un'essenza poliedrica e transdisciplinare, visto che per anni è stato oggetto di studio di numerose discipline. Nel senso più generale, un documento è una rappresentazione registrata

di un pensiero, “qualsiasi mezzo, soprattutto grafico, che provi l’esistenza di un fatto, l’esattezza o la verità di un’asserzione”;<sup>1</sup> il concetto risulta ampiamente sfruttato a partire dalla seconda metà del XX secolo.

L’*Oxford English Dictionary* (OED) lo definisce come «something written, inscribed, etc., which furnishes evidence or information upon any subject, as a manuscript, title-deed, tomb-stone, coin, picture, etc.»;<sup>2</sup> la definizione, a differenza della precedente, non entra nel merito dell’eventuale esattezza o veridicità dell’asserzione, ma con *evidence* si fa solo riferimento alla prova dell’esistenza. La voce risulta aggiornata al 1989 e presenta un elenco piuttosto ricco di formati: un manoscritto, un atto legale, una pietra tombale, una moneta, un’immagine o una fotografia; tuttavia, non prende in considerazione il mondo digitale. Nel giugno 2015 è stata inserita un’integrazione che definisce il documento come “a collection of data in digital form that is considered a single item and typically has a unique filename by which it can be stored, retrieved, or transmitted (as a file, a spreadsheet, or a graphic)”;<sup>3</sup> viene così ampliata la definizione e, allo stesso tempo, si pone l’accento sull’unità del documento, unità che viene richiamata anche nella definizione del lemma in ambito informatico data nell’*Enciclopedia Treccani*.<sup>4</sup>

Michael K. Buckland, nel saggio *What is a “Document”?*, si pone la domanda su che cosa possa o non possa essere definito *document*. La risposta è importante per qualsiasi ulteriore definizione nell’ambito dell’Information Science. “Ordinarily the word “document” denotes a textual record”. Normalmente i sistemi di archiviazione e reperimento delle informazioni si sono occupati di documenti testuali e simili (ad esempio, nomi, numeri e codici alfanumerici). L’attuale interesse per la “multimedialità” ci ricorda che non tutti i fenomeni d’interesse per la scienza dell’informazione sono di tipo testuale; possiamo avere a che fare con qualsiasi fenomeno che possa essere osservato: eventi, processi, immagini, oggetti e testi (Buckland 1997, 804).

La norma ISO 5127:2017 definisce *document* «recorded information or material object which can be treated as a unit in a documentation process».<sup>5</sup>

La voce “document theory”,<sup>6</sup> curata da Buckland e inserita nella *ISKO Encyclopedia of Knowledge Organization*, ha definito il documento come qualunque oggetto o segno atto a rappresentare un fenomeno fisico o concettuale. I documenti risultano la materializzazione dell’informazione; essi sono i mezzi grazie ai quali la conoscenza umana viene espressa, e le loro forme possono includere diversi supporti. L’inserimento del concetto di supporto, di fisicità, richiama la distinzione tra *informazione* e *documento*; l’*informazione* è un’astrazione, un pensiero, un contenuto intellettuale, mentre *i documenti* ne sono la realizzazione materiale attraverso un mezzo, un supporto appunto, che funge da veicolo dell’informazione.<sup>7</sup> Elaine Svenonius, nel saggio *The intellectual foundation*

<sup>1</sup> «Documento», *Vocabolario Treccani online*, aprile 16, 2011, <<http://www.treccani.it/vocabolario/documento>>.

<sup>2</sup> «Document», *OED Online*, Oxford University Press, 1989, <<https://www.oed.com/view/Entry/56328?rskey=eusKz-Z&result=1#eid>>.

<sup>3</sup> Ivi.

<sup>4</sup> «Documento», *Enciclopedia Treccani Online*, giugno 11, 2011, <<http://www.treccani.it/enciclopedia/documento>>.

<sup>5</sup> <<https://www.iso.org/standard/59743.html>>.

<sup>6</sup> <<https://www.isko.org/cyclo/document#3>>.

<sup>7</sup> Grazie alla rivoluzione digitale i veicoli informativi si sono moltiplicati e hanno reso maggiormente accessibile l’informazione.

*of information organization*, definisce *information* “the content of a message”, mentre “the term *document* is used to refer to an information-bearing message in recorded form” (Svenonius 2000, 7-8); ovvero, l'*informazione* è “il contenuto di un messaggio” o “qualche cosa che viene comunicata” e il *documento* è “un messaggio che contiene informazioni in forma registrata” (Svenonius 2008, 10). Inoltre, precisa Svenonius, “a document is a particular space-time embodiment of information” (Svenonius 2000, 107). In biblioteconomia, dalla fine degli anni Sessanta fino all'inizio degli anni Novanta, si parlava di documenti in un'accezione estesa che andava oltre la risorsa cartacea. ISBD(G) esplicitava, infatti, che: “The term *document* is used here in its widest sense” (International Federation of Library Associations and Institutions 1977, 2). L'*Online Dictionary for Library and Information Science* (ODLIS) definisce *document* “a generic term for a physical entity consisting of any substance on which is recorded all or a portion of one or more works for the purpose of conveying or preserving knowledge”.<sup>8</sup> Sulla base delle definizioni riportate, il *documento*, nel senso più ampio, sarebbe quindi il supporto che permette la trasmissione nel tempo e nello spazio di informazioni e conoscenza.

Tuttavia, il termine confligge col significato che ha in archivistica. Il documento è ciò di cui si compone un archivio, anche se si tratta di documentazione informale, lettere private, testi a stampa, fotografie eccetera, “testimonianza scritta di un fatto di natura giuridica, compilata con l'osservanza di determinate forme che conferiscono al documento pubblica fede e forza di prova”;<sup>9</sup> si è reso quindi necessario ricercare un termine alternativo nel rispetto delle differenze disciplinari. Il lemma *risorsa* ha, pertanto, sostituito il termine *documento*, ampiamente usato in passato (Bianchini e Guerrini 2019, 25).

*Risorsa* in italiano è un termine generico (si parla di risorse energetiche, risorse umane, gestione risorse nei PC, etc.); entrato nel vocabolario biblioteconomico internazionale dall'inizio del XXI secolo, comprende qualsiasi entità parte del nuovo ecosistema bibliografico. Una risorsa è il motivo di ricerca dell'utente, è qualsiasi istanza portatrice di un messaggio. *Risorsa* è un'entità, un'unità distinta, analogica o digitale, che veicola un contenuto, intellettuale o artistico o entrambi; è un termine omnicomprensivo per tutto ciò che è conservato nelle biblioteche, ma anche negli archivi e nei musei. Sono risorse i manoscritti, i libri antichi e moderni, gli e-book, i documenti d'archivio e le opere d'arte in genere. Il lemma ha avuto ampio impiego nei principi, negli standard, nei modelli concettuali e anche negli acronimi, per esempio, in URI, Uniform Resource Identifier; URL, Uniform Resource Locator e RDF, Resource Description Framework (Guarasci e Guerrini 2022, 11).

FRBR non prevede esplicitamente l'entità risorsa, intesa nel senso di risorsa bibliografica (si parla soltanto di *electronic resources* e *information resources*). In FRBR la risorsa è trattata implicitamente tramite le quattro entità del I Gruppo 1 (opera, espressione, manifestazione e item); *risorsa*, nelle sue molteplici connotazioni, è da intendere come l'oggetto di interesse dell'utente. ISBD Consolidated Glossary registra l'accoglimento del termine *resource* al posto di *document*, *item* o *publication*, utilizzati nelle precedenti ISBD.<sup>10</sup> Gli ICP del 2009 evitano il lemma do-

<sup>8</sup> <[https://products.abc-clio.com/ODLIS/odlis\\_d](https://products.abc-clio.com/ODLIS/odlis_d)>.

<sup>9</sup> <<http://www.archivi.beniculturali.it/index.php/abc-degli-archivi/glossario>>.

<sup>10</sup> ISBD era nato e si era modellato sulla struttura e le caratteristiche del libro, e ne era rimasto fortemente condizionato sia a livello concettuale sia terminologico.

*cumento* sia nel testo sia nel glossario, sostituito da *risorsa bibliografica*. Nel Glossario di ICP *risorsa bibliografica* è:

Un'entità nell'ambito della biblioteca e di raccolte analoghe che consiste di prodotti dell'attività intellettuale o artistica. Nel modello FRBR le risorse bibliografiche sono le entità del I Gruppo: opera, espressione, manifestazione ed esemplare (International Federation of Library Associations and Institutions 2009, 12).

In RDA il lemma *risorsa*, adottato per rappresentare tutte le tipologie di mezzi informativi utili a soddisfare i bisogni degli utenti, serve a identificare un'opera, un'espressione, una manifestazione o un item; può riferirsi, quindi, a un'entità tangibile o intangibile. *Risorsa*, inoltre, perde l'aggettivo *bibliografica*; si prende coscienza che non è più adeguato, poiché la catalogazione si occupa di qualsiasi entità che risponda agli interessi dell'utente, non soltanto del libro; peraltro, in area anglosassone *bibliographic* aveva un'accezione più ampia rispetto a quella italiana legata solo a esso. Il modello IFLA LRM considera generalmente d'interesse per le biblioteche le informazioni bibliografiche relative a tutti i tipi di risorsa e dichiara, nel prospetto delle funzioni utente (3.2), che il termine *risorsa* è usato in senso molto ampio; esso include le istanze di qualsiasi entità definita nel modello, così come le risorse di biblioteca propriamente dette, a dimostrazione che queste ultime sono rilevanti dal punto di vista dell'utente finale.

Il passaggio dal *documento* alla *risorsa*, supportato dall'adozione del nuovo termine a livello internazionale nei modelli, principi e standard emanati dall'IFLA appena analizzati, si allinea ai contemporanei mutamenti avvenuti nel panorama informativo.

## Un caso di studio: Item

Il termine *Item* ha subito un'evoluzione semantica significativa, che permette di evidenziare sia le difficoltà di trasposizione da un sistema linguistico a un altro sia le molteplici accezioni terminologiche legate a contesti disciplinari diversi. Rilevabile oggi in ambito biblioteconomico, istituzionale, giuridico e archivistico, in latino non compariva come sostantivo ma soltanto come avverbio utilizzato nel linguaggio notarile con il significato di “ancora, inoltre”, ad esempio nell'enumerazione degli oggetti di un lascito o di una donazione. In inglese risulta invece sostantivato, il cui utilizzo è oggi legato a contesti come quello informatico e commerciale con il significato di “articolo”. Introdotto in ambito catalografico da FRBR, indicava la quarta entità del I Gruppo dopo *Work*, *Expression* e *Manifestation*, ed era inizialmente considerato come l'unità concreta, oggettiva presente in biblioteca. FRBR dichiara che “the fourth entity defined in the model is *item*: a single exemplar of a manifestation. The entity defined as *item* is a concrete entity”. Dopo IME ICC di Francoforte (Guerrini e Genetasio 2012) le quattro entità WEMI del I Gruppo di FRBR, identificanti i prodotti della creazione intellettuale e artistica, hanno assunto una valenza astratta in quanto entità del modello.

Non esiste in italiano un corrispondente di *Item*, per questo si è ampiamente discusso se mantenere in testi ufficiali il termine originale. Nella traduzione di FRBR curata dall'ICCU (2000) *Item* viene reso con *documento*. Questa scelta, però, entra in contrasto con il significato che il termine ha in archivistica. *Item* e *document*, inoltre, sono due termini che tendevano ad alternarsi



nel lessico consolidato per indicare ciò che intendiamo oggi con *risorsa* bibliografica, ma, mentre *document* veicolava immediatamente una valenza bibliografica, *item* acquisiva tale valenza solo in ambito biblioteconomico. In altri contesti *item* indicava l'*elemento* indifferenziato caratterizzante un determinato insieme; *item* era innanzitutto il *pezzo* visto nella sua unità. L'indeterminatezza di un termine come *item* ne veicolava l'idea di varietà e di differenziazione. Anche in altri paesi oltre l'Italia la resa traduttiva di *item* ha suscitato molteplici riflessioni. La prima traduzione francese di FRBR (2001),<sup>11</sup> come l'edizione italiana del 2000, traduceva *item* con *document*; l'aggiornamento del 2012<sup>12</sup> invece mantiene *item* inalterato. In FRBR traduzione spagnola (2004)<sup>13</sup> si ritrova *item*, successivamente mutato in *ejemplar* nell'edizione del 2016. La traduzione tedesca del 2009<sup>14</sup> presentava *exemplar*. In RDA francese e spagnolo viene mantenuto *item* mentre la traduzione tedesca presenta *exemplar/en*.

La traduzione di ICP, sia nella prima edizione del 2009 sia nell'aggiornamento del 2016, ha reso *item* con *esemplare*.<sup>15</sup> ICP definisce *item*: "a single exemplar of a manifestation". La resa con *esemplare* era sembrata corretta ai traduttori visto che è il termine utilizzato per indicare la copia di un libro; in realtà, *item* non può essere tradotto con il termine *esemplare* poiché il concetto di *esemplare* nella tradizione bibliografica italiana possiede almeno due accezioni: quella di *copia in mano* rappresentativa della manifestazione e l'altra di *unità concreta*. *Esemplare* connota una risorsa vista nella sua unicità di appartenenza, di contrassegni e di relazioni (ad esempio, ex libris, note, commenti a margine eccetera). Inoltre, in una traduzione corretta il termine utilizzato nella lingua di destinazione non dovrebbe avere un corrispettivo diverso in quella di partenza. Nel caso specifico l'italiano *esemplare* potrebbe invece trovare il suo sostituto linguistico nell'inglese *exemplar* il quale, seppur presente nel lessico inglese, risulta però un termine ambiguo che unisce le diverse accezioni di modello, modello archetipico, esemplificazione. In ambito catalografico *exemplar* non è stato impiegato in ISBD, ICP e RDA proprio a causa della sua ambivalenza semantica. Anche *item* ha confini semantici non ben definiti ma ha il pregio in lingua inglese di contemplare una molteplicità di aspetti che rinviano alla singola unità elementare; Webster's Third International Unabridged Dictionary definisce *Item*:<sup>16</sup>

1. a distinct part in an enumeration, account, or series: article
2. an object of attention, concern, or interest
3. a separate piece of news or information

La scelta originaria di utilizzare *item* anziché *exemplar* ne ha sottolineato la novità concettuale: il latino si è inaspettatamente trasformato nella lingua "nuova" utile per esprimere nuovi concetti che non possono essere resi con termini di uso comune. *Exemplar*, infine, si applica bene a un manoscritto ma non alle risorse elettroniche o alla musica a stampa.

ISBD Consolidata (2012) ha imposto nuovi concetti anche tramite scelte terminologiche; si è de-

<sup>11</sup> <<https://www.ifla.org/files/assets/cataloguing/frbr/frbr-fr.pdf>>.

<sup>12</sup> <<https://repository.ifla.org/bitstream/123456789/811/2/ifla-functional-requirements-for-bibliographic-records-frbr.pdf>>

<sup>13</sup> <<https://www.ifla.org/files/assets/cataloguing/frbr/frbr-es.pdf>>.

<sup>14</sup> <[https://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2009\\_de.pdf](https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2009_de.pdf)>.

<sup>15</sup> Nell'*Introduzione alla traduzione italiana* si legge "Nella tradizione catalografica italiana l'entità *item* viene resa sia con il termine *esemplare* sia con il termine *item*; si è deciso qui di mantenere il termine *esemplare* per continuità con la traduzione della Dichiarazione 2009" (ICP2016, p.4).

<sup>16</sup> <<https://www.merriam-webster.com/dictionary/item>>.

ciso di utilizzare il lemma *risorsa* al posto di *item* per evitare l'ambiguità legata al diverso uso che del termine aveva fatto FRBR; il modello entità-relazione era stato concepito a un livello di astrazione più alto rispetto allo standard. Infine, RDA (2015) e IFLA LRM (2017) hanno mantenuto *item* inalterato. *Item*, pertanto, è da intendere come l'unità elementare, ma è utile evidenziare una differenza. *L'item-catalografico* è un oggetto reale, *l'istanza* concreta, ben diverso dall'*item-entità* inserito in un modello concettuale: il primo può essere un *esemplare* di cinquecentina, una *copia* di un libro moderno, *l'originale* di una scultura o di un quadro in un museo, un *documento* in un archivio. L'altro è *l'entità* di uno dei tanti modelli possibili identificabile attraverso le proprietà definite per quella specifica *entità*, cioè i suoi attributi e le relazioni create all'interno del modello stesso.

## Conclusioni

Il nuovo universo bibliografico, molto più vasto ed eterogeneo, spinge verso una inevitabile ridefinizione della terminologia; la necessità di un nuovo vocabolario risulta direttamente proporzionale alle novità rilevabili nel contesto operativo e nei concetti a esso sottesi. Inoltre, per offrire una base solida al dialogo interdisciplinare è indispensabile un lessico professionale comune e condiviso, che consenta di definire univocamente i problemi. In una prospettiva di integrazione tra domini disciplinari diversi, la consapevolezza crescente dell'evoluzione terminologica apre scenari più ampi di collaborazione tra gli attori che ne fanno parte. La comunità professionale si interroga sul proprio ruolo nel nuovo ecosistema informativo rispetto alla mission consolidata della biblioteca: organizzare, conservare e rendere accessibile la conoscenza in ogni sua forma; *l'IFLA Code of Ethics for Librarians and other Information Workers* dichiara nel preambolo che "The need to share ideas and information has grown more important with the increasing complexity of society in recent centuries and this provides a rationale for libraries and the practice of librarianship. The role of information institutions and professionals, including libraries and librarians, in modern society is to support the optimisation of the recording and representation of information and to provide access to it".<sup>17</sup>

---

<sup>17</sup> <<https://www.ifla.org/publications/node/11092#responsibilities>>.

## Riferimenti bibliografici

Biagiotti, Denise. 2020. *Continuità e innovazioni terminologiche nella catalogazione descrittiva: Luigi Crocetti e la promozione di ISBD in Italia*, Tesi, Università di Firenze.

Biagiotti, Denise. 2021. «Che vi sia ciascun lo dice, dove sia nessun lo sa». Discussions on Italian cataloging terminology at the turn of the 20th and 21 st centuries.” *JLIS.it*, 12, 3:135-148.

Bianchini, Carlo, e Mauro Guerrini. 2018. “New terms for New Concepts: Reflections about the Italian Translation of RDA.” *JLIS.it*, 9, 1:1-5.

Bianchini, Carlo, e Mauro Guerrini. 2019. *Guida alla biblioteca per gli studenti universitari*. Con la collaborazione di Andrea Capaccioni; postfazione di Alberto Salarelli. Milano: Editrice Bibliografica.

Buckland, Michael K. 1997. *What is a “Document”?*, «*Journal of the American Society for Information Science*», vol. 48, no. 9, p. 804-809.

Guarasci, Roberto, e Mauro Guerrini. 2022. *Cos'è l'indicizzazione*. Milano: Editrice Bibliografica. 2022.

Guerrini, Mauro, e Cristina Cavallaro. 2011. *Per un tesoro delle scienze del documento*, in «*Books seem to me to be pestilent things*». *Studi in onore di Piero Innocenti per i suoi 65 anni*, Manziana: Vecchiarelli Editore: 1121-1126.

Guerrini, Mauro, e Giuliano Genetasio. 2012. *I principi internazionali di catalogazione (ICP): universo bibliografico e teoria catalografica all'inizio del XXI secolo*. Milano: Editrice bibliografica.

Guerrini, Mauro. 2015. “Old Wine, New Bottle? Principi e metodi per una reale innovazione nelle prospettive LIS: il parere di Marshall Breeding, Mauro Guerrini, David Weinberger, Paul Gabriele Weston, Maja Žumer”. *AIB Studi*, 55, n. 3:385-403.

*IFLA Library Reference Model. A conceptual model for bibliographic information*. 2017. Pat Riva, Patrick Le Boeuf, and Maja Žumer Consolidation Editorial Group of the IFLA FRBR Review Group, <<https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017.pdf>>.

IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional requirements for bibliographic records. Final report*. München: Saur, 1998. Edizione italiana: *Requisiti funzionali per record bibliografici*. Roma, ICCU, 2000.

International Federation of Library Associations and Institutions. 1977. *ISBD(G): General International Standard Bibliographic Description: Annotated Text*, prepared by the Working Group on the General International Standard Description set up by the IFLA Committee on Cataloguing, London: IFLA International Office for UBC.

International Federation of Library Associations and Institutions. 2007. *International Standard Bibliographic Description (ISBD). Preliminary consolidated edition*, recommended by the ISBD Review Committee, approved by the Standing Committees of the IFLA Cataloguing Section. München: Saur.

International Federation of Library Associations and Institutions. 2009. *Statement of International Cataloguing Principles 2009*, <[https://www.ifla.org/wpcontent/uploads/2019/05/assets/cataloguing/icp/icp\\_2009-en.pdf](https://www.ifla.org/wpcontent/uploads/2019/05/assets/cataloguing/icp/icp_2009-en.pdf)>.

International Federation of Library Associations and Institutions. 2010. *International Standard Bibliographic Description (ISBD)*. Edizione consolidata preliminare, raccomandata dall'ISBD Review Group, approvata dallo Standard Committee dell'IFLA Cataloguing Section; Edizione italiana a cura dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche. Roma: ICCU.

International Federation of Library Associations and Institutions. 2012. *ISBD: International Standard Bibliographic Description*. Edizione consolidata, raccomandata dall'ISBD Review Group, approvata dallo Standing Committee dell'IFLA Cataloguing Section; Edizione italiana a cura dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche. Roma: ICCU.

International Federation of Library Associations and Institutions. 2016. *Statement of International Cataloguing Principles 2016*, <[https://www.ifla.org/files/assets/cataloguing/icp/icp\\_2016-en.pdf](https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf)>.

ISO <International Organization for Standardization>, *ISO 5127:2017, Information and documentation: foundation and vocabulary*, ISO, 2017, <<https://www.iso.org/standard/59743.html>>.

RDA. *Resource Description and Access*. 2010. Joint Steering Committee for Development of RDA. Chicago: American Library Association, <<https://original.rdatoolkit.org>>.

RDA. *Resource Description and Access*. 2015. Developed in a collaborative process led by the Joint Steering Committee for Development of RDA (JSC), representing the American Library Association, the Australian Committee on Cataloguing, the British Library, the Canadian Committee on Cataloguing, CILIP: Chartered Institute of Library and Information Professionals, Deutsche Nationalbibliothek, and the Library of Congress, Edizione italiana a cura dell'ICCU, Roma, <[http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/RDA\\_Traduzione\\_ICCU\\_5\\_Novembre\\_REV.pdf](http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2015/RDA_Traduzione_ICCU_5_Novembre_REV.pdf)>.

*Seminario FRBR: Functional requirements for bibliographic records = Requisiti funzionali per record bibliografici*. 2000. Atti a cura di Mauro Guerrini. Roma: Associazione italiana biblioteche.

Svenonius, Elaine. 2000. *The intellectual foundation of information organization*, Cambridge, Mass.: MIT Press; traduzione italiana: Svenonius, Elaine. 2008. *Il fondamento intellettuale dell'organizzazione dell'informazione*, a cura di Maria Letizia Fabbrini, Firenze: Le Lettere.

Tillett, Barbara. 2008. *Bibliographic universe and the new IFLA cataloging principles: lectio magistralis in library science = L'universo bibliografico e i nuovi principi di catalogazione dell'IFLA: lectio magistralis di biblioteconomia*. Fiesole (Firenze): Casalini libri, <<http://digital.casalini.it/10.1400/257027>>.

## Names, things, places: towards a semantic, sustainable, usable integration?

Pierluigi Feliciati<sup>(a)</sup>

a) University of Macerata, <https://orcid.org/0000-0002-2499-8528>

Contact: Pierluigi Feliciati, [pierluigi.feliciati@unimc.it](mailto:pierluigi.feliciati@unimc.it)

Received: 23 May 2022; Accepted: 1 June 2022; First Published: 15 September 2022

### ABSTRACT

In the nowadays world characterized by complexity, the digital systems for archival and bibliographic description are, especially in Italy, a faithful mirror of too many horizontal (among bodies with national functions) and vertical (among central and peripheral levels) complications. The integration of data sets conceived according to the up-to-date conceptual domain models would simplify the dynamics of representation, improve user experiences, and optimize the environmental costs of computing infrastructures. The concepts to be shared should relate to agents (individual or collective), places, and chronological data, deepening on how to address the more complex issues opened by *things*. This paper introduces three possible scenarios of semantic cooperation between archivists and librarians, not necessarily alternative: activating a conceptual matching through common entities, relying on a neutral semantic data infrastructure such as Wikidata, or, finally, developing a transversal core ontology. Hoping that disciplinary boundaries will not impede cooperation, two types of impediments must be considered: the organizational one (bottom-up, top-down, or a virtuous synergy between the two organizational models?) and the crucial issue of offering easy-to-use interfaces to end users, not only constituted by software agents.

### KEYWORDS

Conceptual models; Archival description; Bibliographic description; Semantic integration; Wikidata.

## Nomi, cose, città: verso un'integrazione semantica sostenibile e usabile?

### ABSTRACT

In un mondo caratterizzato dalla complessità, i sistemi digitali di restituzione delle descrizioni archivistiche e bibliografiche sono, specie in Italia, specchio fedele delle troppe complicazioni orizzontali (tra enti con funzioni nazionali) e verticali (tra livelli centrali e periferici). L'integrazione tra *set* di dati basati sui più aggiornati modelli concettuali di dominio avrebbe come vantaggi la semplificazione delle dinamiche di rappresentazione, il miglioramento delle esperienze d'uso degli utenti e l'ottimizzazione dei costi ambientali delle infrastrutture di calcolo. I concetti condivisibili possono essere prima di tutto quelli relativi agli agenti (individuali o collettivi), ai luoghi e ai dati cronologici, approfondendo poi le più complesse problematiche relative alle "cose". Il contributo introduce tre possibili scenari di cooperazione semantica tra archivisti e bibliotecari, non necessariamente alternativi: far dialogare i modelli concettuali attraverso entità comuni, appoggiarsi ad un'infrastruttura dati semantica neutrale come Wikidata oppure, infine, elaborare una *core ontology* trasversale. Auspicando che i confini disciplinari non siano un impedimento alla cooperazione, si devono considerare però almeno due ordini di problemi: quello organizzativo (bottom-up, top-down o una virtuosa sinergia tra i due approcci?) e quello cruciale relativo all'offerta di interfacce usabili per gli utenti finali, non solo costituiti da agenti software.

### PAROLE CHIAVE

Modelli concettuali; Descrizione archivistica; Descrizione bibliografica; Integrazione semantica; Wikidata.

*Un comune destino  
ci tiene qui. Lo sapevamo.  
Ma non troppo bene.  
Mariangela Gualtieri, Novemarzoduemilaventi*

## Modellare la complessità, non complicarla

Il perno epistemologico intorno a cui ruota la società contemporanea, dalla fine del secolo scorso, è la complessità. Calvino scriveva nella sua ultima opera che “Carlo Emilio Gadda cercò per tutta la sua vita di rappresentare il mondo come un garbuglio, o groviglio, o gomitolo, di rappresentarlo senza attenuarne affatto l’inestricabile complessità, o per meglio dire la presenza simultanea degli elementi più eterogenei che concorrono a determinare ogni evento.” (Calvino 1985: 78) E ancor prima, nel 1962, Umberto Eco, parlando di poetica aperta, aveva osservato che l’autore dovrà “narrare il suo personaggio così come egli si manifesta nella situazione, narrarlo nei modi proposti dalla situazione, descrivere la complessità e la imprecisione dei suoi rapporti, l’inesistenza dei suoi parametri di comportamento, attraverso la messa in crisi dei parametri narrativi.” (Eco 2017: 678). La complessità non deve spaventarci, però, come sostiene Mauro Ceruti in alcune pagine illuminanti di un suo recente volume: “per tre secoli almeno, abbiamo pensato di trovarci sempre di fronte a realtà o sistemi, più o meno complicati, da dover semplificare per scoprirne l’intelligibilità”. Dai primi decenni del ‘900 infatti “si è preso coscienza del fatto che, semplificando un sistema complesso, si finisce per mutilarlo e per inficiarne a priori l’intelligibilità” (Ceruti e Bellusci 2020: 10). Oggi si pone “la sfida di un nuovo umanesimo planetario, generato da una cultura della complessità, che deve diventare cultura dell’educazione libera e aperta, della condivisione delle responsabilità e delle conoscenze, della democrazia intesa come progettazione solidale globale.” (Ibidem: 141). Dobbiamo avere paura piuttosto dei *semplificismi*, delle soluzioni pronte, “di un pensiero disgiuntivo e riduttivo, cui fa da *pendant* la frammentazione ancora accentuata dei saperi e delle istituzioni formative, inadeguate a innalzarsi alla nuova scala globale dei problemi” (Ibidem: 97).

In quali modi e sulla base di quali modelli concettuali i professionisti della documentazione possono, nel proprio lavoro di mediazione esperta delle registrazioni frutto dell’agire umano (Bruni et al. 2016; Guerrini 2020), integrare le entità conoscitive relative agli agenti, ai tempi, agli spazi, agli oggetti documentari?

Il contesto politico del nostro paese di certo non ci aiuta. Assistiamo attoniti alla moltiplicazione dei centri decisionali sia in senso verticale, applicando confusamente le novità in tema di sussidiarietà della Carta costituzionale occorse dopo il 2001, sia in senso orizzontale, aumentando gli istituti centrali e le direzioni generali del Ministero della Cultura con l’effetto di amplificare le sovrapposizioni di funzioni e di progetti. Nel nostro paese la complessità, al di là degli slogan che gridano all’innovazione per l’innovazione, è resa sempre più complicata. Questo panorama confonde nei cittadini la percezione delle responsabilità dei decisori, impegnati sempre in nuovi progetti *in progress*, e minano la credibilità delle istituzioni e l’efficacia dei servizi, paradossale (o forse paradigmatica) in tempi di rinnovata disponibilità di risorse finanziarie (Piazzini 2022).

Impegnarsi per l’integrazione dei dati, dei sistemi e dei servizi, dunque, deve trovare le sue motivazioni più profonde nella nobile idea di *servizio*, ispirandosi alle più avanzate esperienze di ri-

pensamento dei dogmi relativi ai modelli di conoscenza e attivando così una spinta virtuosa verso una facilitazione delle condizioni di accesso e uso degli oggetti di conoscenza da parte degli utenti finali. Utenti e operatori sono costretti oggi a barcamenarsi tra molti, troppi punti di accesso, non comprendendo con facilità quali contenuti possano essere trovati, e dove, lasciando la copertura informativa di ciascun servizio, indipendentemente dalle dichiarazioni implicite dei loro titoli, notevoli spazi di vuoto, le cosiddette *hidden collections* (Yakel 2005), pur a fronte dei molti, troppi spazi di sovrapposizione. Coloro che usano i sistemi non possono non scegliere di rivolgersi ai motori di ricerca generalisti, che illudono di poter rispondere a tutte le possibili richieste e curiosità (Valacchi 2016; Bianchini 2017).

L'obiettivo dell'integrazione dei significati, infine, può portare anche a un'ottimizzazione a lungo termine della sostenibilità dei servizi digitali. Ridurre al minimo le ridondanze e le sovrapposizioni riduce i costi di produzione e di aggiornamento delle risorse conoscitive e semplifica la gestione e il mantenimento delle risorse tecnologiche, hardware e software, in linea con diversi tra gli obiettivi dell'Agenda ONU 2030 per lo sviluppo sostenibile (Di Domenico, Bilotta e Damiani 2021).

## **L'integrazione delle risorse informative: cosa e come? Tre scenari per l'integrazione**

Gli snodi critici sopra accennati, certo, non sono di agevole soluzione: richiedono riflessioni profonde sulla formazione, sulle competenze, sulle politiche di reclutamento da parte degli istituti della memoria, intervenendo anche sui contesti organizzativi così come le normative in vigore nel nostro paese oggi, spesso infelicemente, li configura. Convergere su una nozione estesa di catalogo, ad esempio, intendendolo come collezione sistematica e ordinata di registrazioni, un sistema di metadati (Guerrini 2020), potrebbe essere un punto di partenza per facilitare il dialogo – non solo teorico – tra gli archivisti e i bibliotecari impegnati nella costruzione di servizi informativi digitali. Venendo ai contenuti, sono convinto che sia giunto il tempo di ammettere che esistono spazi di integrazione tra risorse bibliografiche e risorse archivistiche. In un determinato contesto storico, dinamico, mutevole e *imperfetto* per definizione, le stesse persone, negli stessi luoghi e negli stessi momenti, da sole o riunite in famiglie od organizzazioni più o meno formali, hanno redatto e ricevuto documenti, letto e scritto opere letterarie o scientifiche, prodotto o acquistato oggetti artistici, di cui forse sono state anche soggetto, hanno costruito e abitato mura, case e palazzi e contribuito e respirato la propria cultura immateriale. Certo, ogni risorsa documentaria che oggettivizza i fatti che costruiscono tali dinamiche non è certo da considerare isolatamente, immersa come è nell'intreccio di contesti concorrenti, con/rispondendo alle motivazioni, alle forme, alle attività (quanto mai diverse), di cui costituisce la manifestazione, materiale o immateriale. Motivazioni necessarie, intenzionali, o anche casuali.

Rappresentare le dinamiche di concezione, formalizzazione, sedimentazione e conservazione insieme alle caratteristiche dei sedimenti stessi richiede, è ovvio, competenze specialistiche, perché i contesti e i contenuti possano essere correttamente compresi, ricostruiti e rappresentati. Inevitabili allora i lessici disciplinari, gli standard descrittivi di settore, i cataloghi e gli inventari. Da evitare, invece, che i sistemi di rappresentazione dei sedimenti siano parcellizzati tipologicamente talvolta fino al paradosso, concentrati in modo ossessivo sulle distinzioni piuttosto che sulle proprietà in comune.

Ancora, è evitabile che il trattamento informativo delle stesse tipologie di risorse documentarie sia effettuato in modo del tutto difforme a seconda del contesto (talvolta fortuito) di conservazione e di descrizione. I contesti, certo, sempre sono significativi e non devono essere sacrificati sull'altare della digitalizzazione, ma sarebbe virtuoso condividere almeno gli stessi approcci e attributi descrittivi. Penso agli archivi nei musei e nelle biblioteche, alle fotografie ovunque siano conservate, alle biblioteche dei musei e degli archivi, alle collezioni artistiche nelle biblioteche e negli archivi, ai giornali negli archivi o nelle biblioteche, alle registrazioni audio e video negli archivi, nei musei e nelle biblioteche, agli oggetti fisici negli archivi, nelle biblioteche e nei musei.

Le rappresentazioni descrittive di queste risorse e dei loro contesti sono prodotte, oltre che per fondamentali intenti gestionali e conservativi, con l'obiettivo di fungere da strumenti essenziali per offrire un efficace processo di *reference*, di mediazione, di cui sono inevitabilmente parte i contenuti, i sistemi/canali e gli utenti.

Un'azione virtuosa e praticabile nella quale si potrebbero impegnare - insieme - le nostre comunità disciplinari sarebbe individuare le possibili integrazioni semantiche tra le entità previste dai modelli concettuali che le diverse comunità disciplinari si sono date. Il primo livello di integrazione è quello tra registrazioni, connettendo ad esempio le risorse bibliografiche con le risorse relative ad archivi, documenti o soggetti produttori di archivi, referenziandole a vicenda e delegandone la qualità al controllo d'autorità svolto al di fuori dello specifico servizio informativo. Anzi, al di fuori del contesto locale, accettando finalmente che viviamo all'interno una rete di conoscenza globale e le mura degli istituti di memoria non proteggeranno ancora a lungo dall'isolamento. E riconoscere pienamente che gli utenti sono soggetti centrali nella progettazione dei servizi digitali. Entrando nel merito delle entità previste dalle ontologie informatiche, intese come rappresentazioni formali, condivise ed esplicite di una tra le possibili concettualizzazioni di un dominio di conoscenza, emerge con chiarezza come risulti più agevole l'interoperabilità semantica prima di tutto tra i concetti chiave relativi alle persone, ai luoghi e ai contesti cronologici. Meno semplice – per ovvi motivi – connettere le entità riferite alle “cose”, a causa delle differenze strutturali, di origine (non solo estrinseche) che esistono tra le risorse documentarie frutto di un'azione creativa e quelle prodotte come conseguenza diretta – anche involontaria - di attività individuali o collettive. Per collegare tali entità, simili dal punto di vista tecnico ma molto dissimili in quanto a contesto di produzione e conservazione, ci si dovrebbe basare forse su un livello ontologico più generale, una macro-classe come la *Thing* di RiC-CM e RiC-O 0.2 (ICA-EGAD 2021a; ICA-EGAD 2021b) e la *Res* di IFLA-LRM (Riva et al. 2017). Ma su questo fronte sono di certo necessari ulteriori approfondimenti e a conti fatti non è necessario che oggetti tanto diversi come un articolo di rivista e un *instrumentum* del XIII secolo debbano somigliarsi nelle loro rappresentazioni descrittive: sarebbe già ottimo che la registrazione bibliografica dell'articolo relativo a quella pergamena si connettesse correttamente al contesto nel quale essa è correttamente rappresentata, così come la descrizione dell'atto notarile rinviasse direttamente all'articolo che lo ha studiato o trascritto.

Non è opportuno entrare qui nel merito specifico delle potenzialità di connessione tra le ontologie di dominio, anche perché si tradirebbe il requisito fondamentale da cui qui si sono prese le mosse, ovvero che sia necessaria la cooperazione tra comunità disciplinari. È però utile ricordare che talvolta i modelli concettuali sembrano essere stati concepiti escludendo ogni opportunità di interoperabilità semantica *ab origine* (Felicati 2021). Voglio piuttosto riprendere alcune considerazioni già presentate alla conferenza internazionale di Firenze *Bibliographic Control in the Digital*



*Ecosystem* (Bergamin e Guerrini 2022) in merito alle possibili strategie attuabili per affrontare tali obiettivi (Felicati 2022). Sembra a chi scrive che gli scenari di cooperazione possano essere almeno tre, non necessariamente alternativi tra loro.

La prima strategia, apparentemente la più semplice da affrontare, sarebbe far dialogare i modelli concettuali, ovvero verificare la potenziale convergenza sugli stessi concetti (entità) e definirne le possibili relazioni. Tentando un abbinamento di massima delle entità principali di RiC-CM 0.2 e di IFLA-LRM, la *Thing* dell'entità RIC-E01 non sembra essere così lontana dall'entità *Res* di IFLA-LRM, come si accennava sopra, considerando i loro rapporti da un lato con la *Record Resource* (considerando la sostanziale unicità dei record), con *Agent*, *Event* e *Date* e dall'altro con *Work/Item* e con *Time-span*, *Place* e *Agent (Person o Collective Agent)*. La concezione di *Nomen* nella LRM come condizione reificante della *Res* rappresenta a mio parere una questione interessante su cui dovrebbe riflettere RiC fin dalla sua prima versione stabile, in via di presentazione alla IFLA Conference 2022, considerando la complessità delle denominazioni sia nella catalogazione bibliografica che nella descrizione archivistica: originale, derivata, normalizzata, sintetizzata.

IFLA-LRM	RiC-O 0.2
Res + Nomen	Thing
Time-span	Date (Single Date, Date Range, Date Set)
Place	Place
Agent (Person, Collective Agent)	Agent (Person, Group, Position, Mechanism)

Tavola 1 – Corrispondenza generale tra le *core entities* di IFLA-LRM e RiC-O (adattato da Felicati 2021)

Il secondo percorso possibile sarebbe concentrare attivamente la cooperazione su una meta piattaforma neutrale, un ambiente di dati condivisi, come Wikidata (Vrandecic 2013; Tharani 2021; Bianchini e Sardo 2022), che utilizza i Linked Open Data per memorizzare asserzioni su elementi che fungono da nodi, collegati dalle proprietà come vertici, senza preventive delimitazioni di dominio. Il progetto è un repository di dati aperto e collegato, disponibile con una licenza aperta CC0. Tim Berners-Lee aveva affermato che la criticità principale per realizzare il Web semantico era che le ontologie devono essere sviluppate, gestite e approvate da specifiche comunità di pratica, evidenziando gli effetti critici e di rallentamento che questo comporta. Con Wikidata come sistema aperto, riusabile e trasversale, la visione del Semantic Web sembra più agevolmente affrontabile. Molti gruppi di lavoro, anche in Italia, sono attivi nella gestione e arricchimento di Wikidata, definendo una struttura di metadati per le risorse bibliografiche e le biblioteche e caricando e condividendo metadati locali a livello globale (Bergamin e Bacchi 2018, Bianchini e Sardo 2022). Alcuni archivisti, specialmente francesi, hanno nondimeno lanciato di recente la comunità di pratica *Wikidata:WikiProject Archival Description*, con l'obiettivo «di creare il database di alta qualità più completo al mondo di fondi archivistici e raccolte di patrimoni, di rappresentare le strutture archivistiche all'interno di Wikidata dove ciò è ritenuto utile e di garantire l'interconnessione tra le risorse archivistiche su Wikidata». Il progetto, collegato al *Wikidata:WikiProject Archives Linked Data Interest Group* sta lavorando all'elaborazione semantica a partire dagli standard descrittivi

dell'International Council of Archives prima di RiC, ISAD(G) e ISAAR, soprattutto (ICA-CDS 2000; ICA-CDS 2003).

In Italia, nel 2014, si è effettuata invece l'elaborazione dell'ontologia del Sistema Archivistico Nazionale<sup>1</sup>. In questo caso, a partire dal modello relazionale direttamente ispirato agli standard descrittivi di cui sopra, si è definito “un modello concettuale del SAN formalizzato in una ontologia espressa in linguaggio OWL pienamente corrispondente alla struttura informativa veicolata dagli schemi XML proposti ai sistemi aderenti come tracciati di scambio per il conferimento al Catalogo delle risorse archivistiche – CAT dei propri dati relativi a complessi archivistici, soggetti produttori, strumenti di ricerca e soggetti conservatori”. Inoltre, si è attivata la “produzione e la pubblicazione dei dati SAN in formato LOD, resi disponibili principalmente attraverso due canali: un *endpoint* SPARQL dedicato; un'area di download organizzata in specifici datasets realizzati in considerazione delle principali classi di oggetti definite dal modello e della loro organizzazione concettuale”.<sup>2</sup> Pur apprezzando il progetto, va rilevato come tali modellazioni semantiche siano realizzate *ex post*, rappresentando entità e relazioni di uno specifico modello relazionale, locale, non potendo così fungere da modello concettuale di riferimento.

Infine, ancora in Italia, è attivo dal 2020 il *Wikidata:Gruppo Wikidata per Musei, Archivi e Biblioteche* (GWDMA), ispirato al *Wikidata Affinity Group*. Lanciata principalmente da bibliotecari, questa estesa comunità di pratica è interessata alle potenzialità di Wikidata come spazio di cooperazione per Musei, Archivi e Biblioteche e potrebbe costituire il substrato ideale nel quale innestare la riflessione e l'attività sull'integrazione semantica tra risorse archivistiche e bibliografiche. La terza possibile strategia che si intravede per costruire un sistema di integrazione semantica tra le risorse informative prodotte nei contesti archivistici e bibliotecari sarebbe convergere verso un nuovo, condiviso, modello concettuale *core*. Questa linea di lavoro dovrebbe prevedere la selezione di classi, entità e proprietà condivise, sempre a partire dagli *Agents* (persone, enti, famiglie), dai loro *Roles/functions* in diversi contesti, dai *Place*, i nomi dei luoghi (anche storici), dai dati cronologici (date esatte o intervallo di date) e dalle azioni/eventi all'origine delle  *cose*, qualificando le loro possibili, molteplici relazioni. Lo sviluppo di una meta-ontologia potrebbe consentire l'integrazione degli *authority record* e il loro controllo ed estrazione sotto forma di asserzioni RDF. La raccolta, la connessione, l'arricchimento e il controllo di informazioni semantiche di alta qualità fornite da diverse fonti di dati aumenterebbe il potenziale semantico dei servizi online, con i vantaggi citati in apertura. Inoltre sarebbe un'occasione unica per ragionare sulla fattibilità di inserire nei modelli concettuali gli utenti, *Agent* con specifici *Roles* (identificazione, diritti, accesso, annotazioni, etc.), così come già si auspicava nella fase di discussione della prima bozza di RiC (Vitali 2017) e come previsto d'altra parte in OAIS (CSDS 2012).

## Conclusioni

Se i vantaggi epistemologici e in termini di qualità e sostenibilità dei servizi di una meta-modellazione della conoscenza sembrano difficili da contestare, sarebbe ingenuo negare che esistano

<sup>1</sup> Vedi <http://dati.san.beniculturali.it/lode/> e <http://www.san.beniculturali.it/SAN/san-lod.owl>.

<sup>2</sup> Vedi <http://www.san.beniculturali.it/web/san/ontologia-san-lod> e <http://dati.san.beniculturali.it/dataset/>.

alcuni problemi per la sua realizzazione. Al di là dei disciplinarismi che non facilitano il dialogo tra le diverse comunità, dobbiamo prima di tutto chiederci quale approccio organizzativo sarebbe più efficace per l'obiettivo di un'integrazione applicativa. Proseguire con le iniziative *bottom-up*, sostenute da una parte delle comunità disciplinari su base volontaria, oppure promuovere progetti *top-down*, strutturati e sufficientemente finanziati, nell'ambito dei quali ci si possa concentrare sul perseguimento di obiettivi sostenibili all'interno di un contesto robusto e durevole? O ancora, come oggi spesso si auspica e come sembrerebbe più che sensato, creare le condizioni per l'incrocio virtuoso e trasparente tra i due approcci organizzativi?

Un secondo aspetto critico riguarda la effettiva *discoverability* (Manoni 2022) e l'usabilità degli insiemi di dati semantici aperti in rete da parte degli utenti finali. Possiamo auspicare che si svilupperanno in tempi brevi soluzioni automatiche di aggregazione di metadati relativi alle risorse documentarie che offrano interfacce di selezione, ricerca, navigazione e accesso alle rappresentazioni? Oppure crediamo che la prospettiva debba essere che gli utenti – anche quelli professionali – debbano imparare ad esprimere interrogazioni in SPARQL, come capita di sentire da alcuni addetti ai lavori? Nuovamente, non va dimenticato che la qualità dei sistemi risiede nella loro capacità di offrire risposte efficaci, efficienti e soddisfacenti ai bisogni degli utenti finali, non nella raffinatezza a monte dei modelli concettuali o delle soluzioni tecnologiche. Gli utenti digitali sono entità difficili da *afferrare*, per la variabilità di bisogni, comportamenti e competenze di cui sono/ siamo portatori (Feliciati 2016; Galeffi 2017; Feliciati 2018; Galeffi e Sardo 2021) e semplificarne i profili sulla base di nostre proiezioni o pregiudizi non può che produrre sistemi inefficaci. La motivazione entro la quale ogni iniziativa di mediazione documentaria deve essere concepita non può essere altro che l'offerta di un servizio soddisfacente.

## Riferimenti bibliografici

- Bergamin, Giovanni e Bacchi, Cristian. 2018. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data." *JLIS.it* 9, 3: 35-74. <http://dx.doi.org/10.4403/jlis.it-12458>. Accessed April 24, 2022.
- Bianchini, Carlo. 2017. "Funziona come Google, vero?. Prima indagine sull'interazione utente-catalogo nella biblioteca del Dipartimento di musicologia e beni culturali (Cremona) dell'Università di Pavia." *AIB studi* 57, 1: 23-49. <https://doi.org/10.2426/aibstudi-11557>. Accessed April 24, 2022.
- Bruni, Silvia, Capetta, Francesca, Lucarelli, Anna, Pepe, Maria Grazia, Peruginelli, Susanna e Rulent, Marco. 2016. "Towards the integration of archives, libraries and museums." *JLIS.it* 7, 1: 225-244. <https://jlis.it/index.php/jlis/article/view/183>. Accessed April 24, 2022.
- Calvino, Italo. 1985. *Lezioni americane. Sei proposte per il prossimo millennio*. Milano: Garzanti.
- Ceruti, Mauro e Bellusci, Francesco. 2020. *Abitare la complessità. La sfida di un destino comune*. Milano: Mimesis.
- CSDS - Consultative Committee for Space Data Systems. 2012. *Reference Model for an Open Archival Information System (OAIS). Recommended Practice CCSDS 650.0-M-2*. Washington, DC, USA: Magenta book. <https://public.ccsds.org/Pubs/650x0m2.pdf>. Accessed April 24, 2022.
- Di Domenico, Giovanni, Bilotta, Anna e Damiani, Concetta. 2021. *Il paradigma della biblioteca sostenibile*. Milano: Ledizioni.
- Eco, Umberto. 1962. *Opera aperta. Forma e indeterminazione nelle poetiche contemporanee*, Milano, Bompiani.
- Feliciati, Pierluigi. 2016. "L'usabilità degli ambienti bibliotecari e archivistici digitali come requisito di qualità: contesto, modelli e strumenti di valutazione." *JLIS.it* 7, 1: 113-130. <https://jlis.fupress.net/index.php/jlis/article/view/179/178>. Accessed April 24, 2022.
- Feliciati, Pierluigi. 2018. "La qualità dell'universo documentario digitale: Dai contenuti al servizio." *AIB Studi* 58, 1: 53-63. <https://doi.org/10.2426/aibstudi-11740>. Accessed April 24, 2022.
- Feliciati, Pierluigi. 2021. "Archives in a Graph. The Records in Contexts Ontology Within the Framework of Standards and Practices of Archival Description." *JLIS.it* 12, 1: 92-101. <https://doi.org/10.4403/jlis.it-12675>. Accessed April 24, 2022.
- Feliciati, Pierluigi. 2022. "Call me by your name: towards an authority data control shared between archives and libraries." *JLIS.it* 13,1: 203-214. <https://doi.org/10.4403/jlis.it-12733>. Accessed April 24, 2022.
- Galeffi, Agnese. 2017. "Se Il Catalogo Parlasse, Lo Capiremmo? Cinque Assiomi Della Comunicazione Catalografica." *AIB Studi* 57, 2: 239-252. <https://doi.org/10.2426/aibstudi-11648>. Accessed April 24, 2022.
- Galeffi, Agnese e Sardo, Lucia. 2021. "Comunicare la catalogazione: un'indagine sulle aspettative degli studenti e sulle esigenze professionali." *AIB Studi*, 61, 1: 31-53. <https://doi.org/10.2426/aibstudi-13009>. Accessed April 24, 2022.

Guerrini, Mauro. 2020. *Dalla catalogazione alla metadattazione. Tracce di un percorso*. Roma: Associazione Italiana Biblioteche.

ICA (International Council of Archives) – EGAD (Experts Group on Archival Description). 2021a, *Records in Contexts Ontology (ICA RiC-O) version 0.2*, 2021-02-12. [https://www.ica.org/standards/RiC/RiC-O\\_v0-2.html](https://www.ica.org/standards/RiC/RiC-O_v0-2.html). Accessed April 24, 2022.

ICA (International Council of Archives) – EGAD (Experts Group on Archival Description). 2021b, *Records in Contexts. A conceptual model for archival description. Consultation Draft v0.2*, July 2021. [https://www.ica.org/sites/default/files/ric-cm-02\\_july2021\\_0.pdf](https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf). Accessed April 24, 2022.

ICA (International Council of Archives) – CDS (Committee on Descriptive Standards). 2000. *ISAD(G): General International Standard for Archival Description, Second Edition*. Ottawa, [https://www.ica.org/sites/default/files/CBPS\\_2000\\_Guidelines\\_ISAD%28G%29\\_Second-edition\\_EN.pdf](https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf). Accessed April 24, 2022.

ICA (International Council of Archives) - CDS (Committee on Descriptive Standards). 2003. *ISAAR (CPF): International Standard Archival Authority Record For Corporate Bodies, Persons and Families. Second Edition*, <https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd>. Accessed April 24, 2022.

IFLA (International Federation of Library Associations and Institutions). 2017. *IFLA Library Reference Model. A Conceptual Model for Bibliographic Information*. Edited by Pat Riva, Patrick Le Boeuf, and Maja Zumer. Den Haag: IFLA. [https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla\\_lrm\\_2017-03.pdf](https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla_lrm_2017-03.pdf). Accessed April 24, 2022.

Manoni, Paola. 2022. “Discoverability’ in the IIF Digital Ecosystem.” *JLIS.it* 13, 1: 312-320. <https://doi.org/10.4403/jlis.it-12770>. Accessed April 24, 2022.

Piazzini, Tessa. 2022. “Bibliographic Control and Institutional Repositories: Welcome to the Jungle.” *JLIS.it* 13, 1: 132-142. <https://doi.org/10.4403/jlis.it-12717>. Accessed April 24, 2022.

Tharani, Karim. 2021. “Much more than a mere technology: A systematic review of Wikidata in libraries.” *The Journal of Academic Librarianship* 47, 2. <https://doi.org/10.1016/j.acalib.2021.102326>. Accessed April 24, 2022.

Valacchi, Federico. 2016. “Pezzi di cose di cose nel mondo. Il processo di integrazione delle descrizioni archivistiche nei sistemi interculturali.” *JLIS.it* 7.2: 333-369, <http://dx.doi.org/10.4403/jlis.it-11529>. Accessed April 24, 2022.

Vitali, Stefano. 2017. “Records in Contexts: A conceptual model for archival description: Il contributo italiano.” *Quaderni del Mondo degli archivi* 2: 3–7, <http://www.ilmondodegliarchivi.org/rubriche/i-quaderni/530-i-quaderni-del-mondo-degliarchivi-n-2-records-in-contexts-a-conceptual-model-for-archival-description-il-contributo-italiano/>. Accessed April 24, 2022.

Vrandečić, Denny. 2013. “The rise of Wikidata.” *IEEE Intelligent Systems*, 28(4): 90–95. <https://dl.acm.org/doi/abs/10.1109/MIS.2013.119>. Accessed April 24, 2022.

Yakel, Elizabeth. 2005. “Hidden collections in archives and libraries.” *OCLC Systems & Services: International digital library perspectives* 21, 2: 95-99. <https://doi.org/10.1108/10650750510598675>.

## Archival description and conceptual transversality

Concetta Damiani<sup>(a)</sup>

a) University of Campania “Luigi Vanvitelli”, <https://orcid.org/0000-0002-3301-0075>

**Contact:** Concetta Damiani, [concetta.damiani@unicampania.it](mailto:concetta.damiani@unicampania.it)

**Received:** 31 May 2022; **Accepted:** 4 July 2022; **First Published:** 15 September 2022

### ABSTRACT

Archival description has become more and more clearly characterized in terms of transversality, as the process of assembling the representations of different information entities; these different components contribute to defining the return of an objective.

To this “internal” transversality is then added the question posed by the growing need for multidimensionality functional to the technological system, in an increase which feeds reflections on normalization and the need for contextual plurality

### KEYWORDS

Archives; Archival description; Context; Intercultural system.

## Descrizione archivistica e trasversalità concettuale

### ABSTRACT

La descrizione archivistica si è andata sempre più nettamente caratterizzando in termini di trasversalità, in quanto processo di montaggio delle rappresentazioni di diverse entità informative; tali diverse componenti concorrono a definire la restituzione di un obiettivo.

A tale trasversalità “interna” si aggiunge poi la questione posta dal crescente bisogno di multidimensionalità funzionale agli apparati tecnologici, in un crescendo che alimenta le riflessioni sulla normalizzazione e sul bisogno di pluralità contestuali.

### PAROLE CHIAVE

Archivi; Descrizione archivistica; Contesto; Sistema interculturale.

## Appunti di viaggio

La riflessione sulle strategie, sulle modalità e sugli sviluppi delle attività di descrizione e comunicazione delle risorse documentali è parte integrante del nostro rapporto con gli archivi.

Se guardiamo allo stato dell'arte, la descrizione archivistica si è andata sempre più nettamente caratterizzando in termini di trasversalità, in quanto processo di montaggio delle rappresentazioni di diverse entità informative che concorrono a definire la restituzione di un obiettivo.

C'è stata una fase in cui è stata avvertita come primaria l'urgenza di restituire un ordine ai documenti e di fornire strumenti di interpretazione istituzionalmente e strutturalmente definiti; tale processo pur nella sua legittimità e, tutto sommato, necessità, ha rischiato di rendere autoisolanti le descrizioni dei complessi documentali; si è pertanto fatto strada il bisogno di riportare le descrizioni archivistiche a una pluralità di contesti, che esulano dall'ambito tradizionalmente delineato e tendono al superamento di una percezione della descrizione finalistica e compiuta in sé stessa (Valacchi 2022). La ricostruzione di un ordine, codificato e proposto negli strumenti di ricerca, ha rappresentato un punto di partenza irrinunciabile oltretutto l'imprescindibile premessa ad una disponibilità dell'archivio a letture, rappresentazioni e riagggregazioni aperte e trasversali. Ma quella stessa ricostruzione, se perseguita a qualunque costo in un tentativo di identificazione perfetta tra il soggetto produttore e il complesso documentale di riferimento, rischia di essere forzata e di portarci lontano dalla realtà delle cose. La declinazione sostanzialmente acritica del principio di provenienza in quanto garanzia qualitativa delle informazioni ha irrigidito in quel senso il processo descrittivo, limitandone in qualche modo la portata.

Le criticità lucidamente segnalate da Claudio Pavone nella difficile operazione di bilanciamento tra la puntuale emersione della storia dell'istituzione che ha prodotto l'archivio e «l'applicazione rigorosa di questo criterio all'opera di riordinamento degli archivi e di stesura degli inventari ha incontrato e incontra molte e gravi difficoltà» (Pavone 1970, 145) perdurano; sono anzi aumentate anche in relazione al mutare delle dinamiche storiche, sociali, economiche e istituzionali e, naturalmente, dei meccanismi che sottendono alla produzione e alla diffusione documentaria contemporanea (Michetti 2018). Naturalmente val la pena di rammentare il ruolo della descrizione archivistica come attività di lunga durata che si sviluppa in ogni fase del ciclo vitale, come esplicitato nell'introduzione di ISAD(G).

Quando descriviamo ci sforziamo di rappresentare i fondi archivistici quali strutture documentarie complesse e articolate, composte di entità informative autoesplicative all'interno di archival relationship, secondo il dettato di ISAD(G). La teoria dei livelli rimane centrale per dar conto di questo tipo di descrizione e tali entità, collegate tra loro, prodotte e gestite da uno o più soggetti, rispondono a finalità di esplicitazione della struttura, del contesto – più opportunamente, dei contesti – di produzione e del contenuto degli archivi.

L'elemento di continuità è stato rappresentato dallo sforzo di normalizzare la descrizione, a favore di una impostazione condivisa e di una maggior qualità della comunicazione. Per quanto riguarda le "buone pratiche di descrizione archivistica", da circa un ventennio la prassi scientificamente validata ci ha allenati a ragionare e lavorare prevalentemente in termini di costruzione di strutture gerarchicamente modellate ad albero logico rovesciato, capaci di restituire il fondo documentario nei termini di un'interpretazione gerarchica e multilivellare, organizzata per separazione tra i diversi tipi di informazione. Questo ha generato iniziali vantaggi nella condivisione dei metodi e

nella resa dei risultati; tuttavia alla concretizzazione di relazioni verticali soddisfacenti, ma a volte limitanti, si è andata affiancando una sempre crescente attenzione alla possibilità di legami orizzontali capaci di restituire più efficacemente - in termini di multidimensionalità e di molteplicità di relazioni - i complessi documentali, ma anche di provare a rappresentare in maniera idonea, correlandoli adeguatamente, i sempre più ricorrenti archivi multimaterici. Si avverte inoltre il bisogno di moltiplicazione dei contesti, cioè dell'introduzione di possibili chiavi interpretative trasversali che consentano una lettura più articolata e circostanziata.

Il percorso compiuto nella progettazione dei software di descrizione archivistica consente di recuperare le evoluzioni della resa descrittiva: per un lungo periodo il modello basato su relazioni gerarchiche ha permeato di sé i prodotti informatici e ha contribuito all'elaborazione di applicativi che rispondessero a questa logica; poi anche per tali strumenti, in funzione dell'evoluzione teorica, si è data la necessità di un ripensamento e di una rivisitazione, in qualche caso già in corso (Arianna4 2020; Sinapsi 2019). Significativa in questo senso anche la norma *Elementi per la valutazione di applicazioni per la descrizione e l'ordinamento d'archivio* (UNI 2020). La norma è uno strumento di orientamento per l'analisi e la comprensione delle caratteristiche degli applicativi dedicati all'ordinamento e alla descrizione dei fondi documentari, anche in rapporto alla sempre più frequente gestione integrata di patrimoni archivistici, librari e museali. Il documento tecnico pone inoltre particolare accento sulla possibilità di stabilire relazioni, gerarchiche e non, all'interno della stessa tipologia di record e tra differenti tipologie e affronta i temi dell'indicizzazione e dell'interoperabilità con altri sistemi e applicazioni.

Non è una novità che molta della partita, oggi, si gioca intorno a *Record in context. A conceptual model for archival description*, sia pure con i limiti che il modello concettuale palesa quando lo si voglia calare nel quadro concreto della descrizione archivistica (Di Marcantonio e Valacchi a c. di 2018; Feliciati 2021; ICA – EGAD 2021a). Con tale modello concettuale gli standard di descrizione archivistica arrivano a un inequivocabile punto di svolta: il documento mostra infatti come sollecitazioni di diversa natura abbiano indotto la comunità archivistica a riflettere sull'esigenza di un modello descrittivo basato sulle relazioni tra risorse informative. Tutto ciò avviene a partire dalle soluzioni tecnologiche disponibili, perché RiC dichiara esplicitamente il suo debito al riguardo, differenziandosi da ISAD(G) che altrettanto esplicitamente si presentava come “atecnologico”, per quanto poi abbia avuto un ruolo determinante collocandosi alla base dei sistemi informativi e non solo. Il nuovo modello prova ad armonizzare i singoli standard sviluppati nel corso degli ultimi vent'anni e, nel contempo, a modificarne la prospettiva; la proposta offre una modalità innovativa per affrontare la descrizione archivistica, affiancando all'ormai tradizionale impianto basato sulla descrizione multilivello un approccio multidimensionale costruito sulle relazioni, aprendo inoltre alla concreta possibilità di un'ontologia del dominio archivistico (ICA – EGAD 2021b).

L'archivio – e di conseguenza la sua descrizione – supera i confini di un mondo chiuso e autoesplicativo. Al contrario, è il serbatoio informativo per l'immissione in un ambito di relazioni e conoscenza dagli innumerevoli sviluppi possibili. Diventa indispensabile quindi rapportarsi al contesto, o più opportunamente ai contesti, travalicando i confini di tracciati precostituiti e provando a impostare percorsi descrittivi aperti (Tomasi 2022).

L'ordine e la struttura restano gli elementi fondamentali da cui prendono origine e forma le rappresentazioni dell'archivio. La restituzione del patrimonio archivistico, destinata a diventare parte della rappresentazione di una realtà più ampia, deve dotarsi di un adeguato equipaggiamento.



Le descrizioni archivistiche vengono ripensate anche in virtù di esigenze di rappresentazione e integrazione in vista di aggregazioni con dati provenienti da ambiti disciplinari affini e da domini diversi dagli abituali ambiti di azione<sup>1</sup>.

I processi mirano all'inclusione delle tecnologie del web semantico e guardano agli sviluppi di una più ampia spendibilità delle risorse documentarie. Gli apparati descrittivi si vanno strutturando in termini di entità aperte all'interoperabilità; questo, oltre ad auspicabili soluzioni di aggregazione, integrazione e contaminazione in domini nuovi e diversi, oltre al debutto in un mondo aperto, deve tener conto anche di cambiamenti generati da particolari ambiti applicativi della descrizione stessa. È il caso dei cosiddetti *invented archives*. Tali "prodotti documentali" non possono essere considerati associazioni decontestualizzate e pertanto derubricati in termini archivistici. Si tratta di aggregazioni di documenti digitali, generati da uno più soggetti produttori, selezionati e proposti intorno a uno specifico tema, ma anche di riproduzioni digitali di complessi documentali proposte integralmente o parzialmente in una nuova dimensione di contesto, che determina quindi il crearsi e il definirsi di nuovi legami e relazioni che in qualche modo alterano o addirittura esulano da quelli con i soggetti produttori e con i conservatori istituzionali. Il prodotto è un fenomeno documentario nuovo e diverso dall'originale analogico e dalla sua riproduzione digitale, rispetto ai quali assume proprie caratteristiche di autonomia e originalità (Pezzica 2020, 17; Valacchi 2019, 173).

Condividere, integrare, riutilizzare sono azioni da preparare adeguatamente. Le descrizioni vanno opportunamente profilate per rispondere a precisi obiettivi informativi e comunicativi e per sostenere adeguatamente l'impatto con il sistema di relazioni che si avviano a costruire.

La tendenza alla redazione di descrizioni esaurienti, strutturate, rafforzate dal dettato degli standard e ricche di potenziale informativo è volta, oltre che a una corretta comunicazione del patrimonio culturale, a raggiungere livelli di interoperabilità e integrazione della risorsa con il più ampio contesto culturale (Tomasì e Daquino 2015). Alla dimensione dei sistemi descrittivi "chiusi" e a loro volta autoesplicativi si sostituisce una concreta interazione con i sistemi interculturali (Valacchi 2016). In vista della costruzione e dell'ingresso in portali interculturali la descrizione standardizzata supera i confini di una normalizzazione strutturale e si apre a contemplare altre dimensioni, prima tra tutte quella tecnologica. In questo contesto la disponibilità all'integrazione prevede anche l'apertura all'utilizzazione delle Intelligenze Artificiali, che nei fatti non devono rappresentare una dittatura tecnologica ma esprimere il valore funzionale del digitale (Floridi 2022).

## Dalla parte dell'archivio: il frattale immagine del fondo documentale

Descrizione e rappresentazione degli archivi sono dunque un problema aperto; se i fondi storici delle organizzazioni istituzionali rimandano a un'apparente mansuetudine, pur nella problematicità e nella complessità che li riguarda (Valacchi 2019), gli archivi "del presente" e in particolare gli archivi di persona reclamano più di un'attenzione (Di Domenico 2020).

---

<sup>1</sup> Il ripensamento è da un lato logico e dall'altro fisico. È logico perché scardina il principio di provenienza e i sistemi di relazione che ne derivano in quanto garanzia di controllo e qualità; fisico nel momento in cui si pone invece come mashup di linguaggi e quindi di tecniche descrittive. Il punto sta nell'individuazione di una crasi descrittiva che coniughi "le schede archivistiche" con le triple, ovvero con le formule e le azioni che sostengono il web semantico e il learning machine.

La resa dell'innumerabile potenziale di relazioni che accompagnano e legano le diverse tipologie documentali, ravvisabili ben oltre l'oggettiva (lo è mai?) sedimentazione archivistica, le volontà più o meno esplicite del soggetto produttore nonché i contesti di produzione, di riferimento e di trasmissione rappresentano un banco di prova non facile per l'impostazione delle attività di descrizione e restituzione. Bisogna poi considerare che quanto più il processo di descrizione si arricchisce e si allontana dalla sua definizione di scuola, tanto più diventa indispensabile un confronto serrato con la concretezza degli archivi e con le effettive peculiarità della loro sedimentazione.

In questa prospettiva l'archivio di Tomás Maldonado (1922–2018), offerto alla consultazione in occasione del centenario della nascita dalla Fondazione Giangiacomo Feltrinelli, offre motivi di riflessione. L'intenzione non è quella di presentare una sperimentazione o un caso di studio, visto che si tratta di un intervento a cui non ho partecipato, ma di ragionare sulle modalità di restituzione descrittiva.

Intellettuale, “giramondo” e “giramestieri”, come viene definito, Maldonado rilascia nelle componenti documentali dell'archivio la ricchezza intellettuale del designer, del filosofo, dell'artista militante, del pensatore, dell'accademico. Lettere, manoscritti, progetti, fotografie, opere d'arte, libri, riviste - che consentono di descrivere le vicende accademiche, culturali, artistiche, politiche di cui è stato protagonista - non sono che il punto di partenza per una più ampia fruizione multidimensionale e multicontestuale.

Siamo di fronte a un archivio analogico (1943-2018) che accoglie anche opere a stampa, frutto dell'attività scientifica e artistica dell'autore e che, a ragione, sono state ritenute parte integrante della struttura del fondo.

Le attività di riordinamento hanno prodotto un'articolazione in cinque serie documentali (*Corrispondenza*, *Attività scientifica*, *Attività artistica*, *Attività professionale* e *Documentazione personale*) e relative sottoserie che vengono proposte nella tradizionale struttura gerarchica, ma si aprono poi anche a percorsi e relazioni trasversali. I curatori, infatti, si sono da subito resi conto del fatto che la produzione documentale, che rispecchia fortemente l'evoluzione e la complessità del pensiero autoriale, sarebbe stata penalizzata e impoverita da scelte di rappresentazione improntate esclusivamente a strutture archivistiche dalle maglie rigide e precostituite. È stato pertanto preferito un modello descrittivo che valorizzasse i legami e le possibilità di rimando sia tra le diverse componenti del fondo, sia tra l'archivio e la personalità del soggetto produttore, sia tra il soggetto produttore e i diversi mondi con i quali è stato in costante dialogo e relazione.

Una sollecitazione, in termini di lettura complementare, capace di generare suggestioni sulla ricchezza e sulla complessità culturale rispecchiata dal fondo Maldonado, è offerta inoltre da alcuni percorsi tematici resi disponibili nell'area Archivi digitali della Fondazione Feltrinelli. In questa sede, con la restituzione di immagini che rimandano al patrimonio documentale, viene proposto una sorta di itinerario maldonadiano. Senza alcuna pretesa di esaustività la pratica artistica, la riflessione sulla progettazione ambientale, il lavoro sperimentale per Olivetti e La Rinascenza - Upim, l'impegno nel rinnovamento dei modelli formativi e dei programmi di studio delle discipline di suo interesse, la partecipazione al dibattito culturale e politico nella compagine del Partito comunista italiano vengono illustrate attraverso una rassegna di documenti a cui è affidato il delicato compito di lasciare intravedere le potenzialità dell'archivio.

Nel bel saggio che ripercorre le tappe del lavoro di riordinamento e di descrizione archivistica gli autori, nel presentare il fondo di un “creator” dalla personalità complessa e dalle molteplici

attività, pur dichiarandosi debitori alla «nota metafora dell'archivio come albero» (Califano e Rubinelli 2022, 10), che ha rappresentato una bussola nei momenti più critici, sostengono che forse l'immagine più appropriata per rendere conto di un fondo archivistico “eterogeneo” è quella del frattale, peraltro cara a Maldonado.

Il frattale è un oggetto in cui le parti sono simili, ma non identiche al tutto; offrono quindi dettagli rivelatori per ottenere una visione più articolata e complessa del tutto. L'archivio Maldonado può essere rappresentato attraverso l'immagine del frattale «perché ogni sua parte riflette in dettaglio la ricerca di trasversalità disciplinare e la volontà progettuale che il soggetto produttore ha posto come principi del suo lungo itinerario intellettuale» (Califano e Rubinelli 2022, 10).

Si tratta di un'interpretazione interessante che a mio avviso, al di là del caso specifico, rilancia ulteriormente la riflessione in merito alla trasversalità – questa volta concettuale – indispensabile per descrivere al meglio un archivio.

Un fondo archivistico è costituito da componenti che, nelle rispettive specificità, consentono di offrire una restituzione stratificata del complesso di azioni che sottendono alla produzione documentale, esplicitando la rete relazionale che congiunge le singole parti tra loro e il fondo con gli innumerevoli contesti di riferimento.

Saremo capaci di muoverci in tal senso? Saremo capaci di far emergere e di rendere disponibile il potenziale relazionale spesso soffocato dalla rigidità delle partizioni? Gli strumenti di mediazione descrittiva riusciranno, senza essere prescrittivi, a restituire all'utenza quella stessa vitalità e quel sistema di rapporti istituzionali, sociali, intellettuali, scientifici che nella compagine del fondo sono ben saldi e presenti? Credo che ci sia molto da costruire, in questo senso, anche per dotare le descrizioni di un potere narrativo ed evocativo.

## Conclusioni

Le attività di descrizione e mediazione archivistica necessitano di un'impostazione e di uno sguardo che siano al contempo trasversali e sistemici. Nel momento in cui consolidati modelli descrittivi sembrano entrare in crisi o, quanto meno, risultare affaticati, si assiste a un paradossale rilancio dell'anima del policentrismo, declinato non più in termini di stretta provenienza ma come spunto di più ampia contestualizzazione del fattore archivio in quanto componente di un sistema che ha la sua ragion d'essere soprattutto fuori dall'archivio stesso. Ciò orienta la descrizione archivistica alla confluenza verso sistemi informativi più ampi e articolati e ne fa l'innescò di processi multiculturali. Questa è una delle possibili e più credibili prospettive per creare e restituire all'utenza la più ampia possibilità di individuazione e tessitura di una rete di relazioni tra soggetti produttori, fondi archivistici e contesti. Soltanto con un'effettiva integrazione descrittiva, declinata innanzitutto nei termini di condivisione delle informazioni tra i diversi domini, si intravede la possibilità di descrizioni multidimensionali e multicontestuali, oltre alla possibilità di sperimentare un'ecologia della memoria, recuperando il pensiero maldonadiano (Maldonado 2005).

L'integrazione, però, non può essere un auspicio teorico né una manifestazione di buona volontà, ma deve prendere corpo in tecniche descrittive capaci di sprigionarsi nella loro dimensione applicativa: questo è forse il principale obiettivo da perseguire in una fase in cui molti, ormai, conven-gono sul bisogno di tali integrazioni.

## Riferimenti bibliografici

Califano, Pierfrancesco e Serena Rubinelli. 2022. *La stanza delle meraviglie: Un'introduzione all'Archivio Maldonado*. In *Tomás Maldonado e la sfida della trasversalità*. 2022. Con i testi di Pierfrancesco Califano, Davide Fornari, Francesco Giasi, Simona Morini, Raimonda Riccini, Serena Rubinelli, Marco Santambrogio. Milano: Fondazione Giangiacomo Feltrinelli. <https://fondazionefeltrinelli.it/app/uploads/2022/04/TomasMaldonadoelasfidadellatrasversalit%C3%A0.pdf>.

Di Domenico, Giovanni. 2020. *Le ragioni di un nuovo convegno su archivi e biblioteche personali*. In *Il privilegio della parola scritta. Gestione conservazione e valorizzazione di carte e libri di persona. Atti del convegno internazionale, Fisciano 10-12 aprile 2019*, a c. di Giovanni Di Domenico e Fiammetta Sabba. Roma: Associazione italiana biblioteche: 9-25.

Di Marcantonio, Giorgia e Federico Valacchi a c. di. 2018. *Descrivere gli archivi al tempo di RIC. Atti del convegno, Ancona 17-18 ottobre 2017*. Macerata: EUM.

Feliciati, Pierluigi. 2021. "Archives in a Graph. The Records in Contexts Ontology within the framework of standards and practices of Archival Description". *JLIS.it* 12(1): 92-101. <https://doi.org/10.4403/jlis.it-12675>.

Floridi, Luciano. 2022. *Etica dell'intelligenza artificiale: Sviluppi, opportunità sfide*. Milano: Raffaello Cortina.

Fondazione Feltrinelli. 2022. <http://fondazionefeltrinelli.it/archivi-digitali/tomasaldonado>.

Hyperborea, Arianna4. 2020. <https://www.hyperborea.com/prodotti/arianna4view-valorizzazione-beni-culturali/>

ICA – CDS (Committee on Descriptive Standards). 2000. *ISAD(G): General International Standard for Archival Description, Second Edition*. Ottawa, <https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>

ICA – EGAD (International Council on Archives – Expert Group on Archival Description). 2021a. *Records in Contexts. A conceptual model for archival description. Consultation Draft version 0.2*. <https://www.ica.org/en/records-in-contexts-conceptual-model>.

ICA – EGAD (International Council on Archives – Expert Group on Archival Description). 2021b. *Records in Contexts Ontology version 0.2*. <https://www.ica.org/en/records-in-contexts-ontology>

Maldonado, Tomás. 2005. *Memoria e conoscenza: Sulle sorti del sapere nella prospettiva digitale*. Milano: Feltrinelli.

Michetti, Giovanni. 2018. "Se un leone potesse parlare, noi non potremmo capirlo: La comunicazione del patrimonio culturale in ambiente digitale". *AIB Studi*, (58)2. <https://doi.org/10.2426/aibstudi-11820>.

Pavone, Claudio. 1970. "Ma è poi tanto pacifico che l'archivio rispecchi l'istituto?". *Rassegna degli Archivi di Stato* (XXX) 1: 145-149.

Pezzica, Lorenzo. 2020. *L'archivio liberato. Guida teorico-pratica ai fondi storici del Novecento*. Milano: Editrice Bibliografica.

Progetto Sinapsi. 2019. <https://www.progettosinapsi.it/>.

Tomasi, Francesca. 2022. *Organizzare la conoscenza: digital humanities e web semantico*. Milano: Editrice Bibliografica.

Tomasi, Francesca e Marilena Daquino. 2015. “Modellare ontologicamente il dominio archivistico in una prospettiva di integrazione disciplinare”, *JLIS.it*, 6 (3): 13-38. <https://doi.org/10.4403/jlis.it-11133>.

UNI (Ente nazionale italiano di unificazione). 2020. *Norma 11769. Applicazioni per i beni culturali (ABC) - Elementi per la valutazione di applicazioni per la descrizione e l'ordinamento d'archivio (VADO)*.

Valacchi, Federico. 2016. “Pezzi di cose di cose nel mondo. Il processo di integrazione delle descrizioni archivistiche nei sistemi interculturali”. *JLIS.it* 7(2): 333-369. <https://doi.org/10.4403/jlis.it-11529>.

Valacchi, Federico. 2019. “Ripartire da Pavone: Spunti di archivistica”. *Parolechiave* 27(1-2): 161-175.

Valacchi, Federico. 2022. *Diventare archivisti: Competenze tecniche di un mestiere sul confine*. Milano: Editrice Bibliografica.

## Agrifood data processing in Italy: a conceptual model\*

Annantonia Martorano<sup>(a)</sup>, Elena Gonnelli<sup>(b)</sup>

a) University of Florence, <https://orcid.org/0000-0001-8795-447X>

b) University of Florence, <https://orcid.org/0000-0002-6908-8106>

**Contact:** Annantonia Martorano, [annantonia.martorano@unifi.it](mailto:annantonia.martorano@unifi.it); Elena Gonnelli, [elena.gonnelli@unifi.it](mailto:elena.gonnelli@unifi.it)

**Received:** 16 June 2022; **Accepted:** 9 July 2022; **First Published:** 15 September 2022

### ABSTRACT

Through the study and analysis of a precise economic sector, that of agribusiness, this contribution attempts to elaborate a replicable conceptual model that, in compliance with scientific methodologies, can contribute to the development of current archives, which will later be historical, capable of managing traceable and secure digital flows. In that absolute semantic individuality that characterizes each documentary complex, subject to “rules” of treatment only by productive affinities and common provenance characteristics, the application of standards has been a matter of discussion for the scientific literature. The example from which we wish to move concerns a specific category of Software Packages (LIMS) that are currently used in Italy by analytical laboratories and that implement integrated management of multiple data and processes.

### KEYWORDS

Conceptual models; LIMS; Agrifood; Integrated systems; Agritech.

## Il trattamento dei dati agroalimentari in Italia: un modello concettuale

### ABSTRACT

Attraverso lo studio e l'analisi di un preciso settore economico, quello dell'agroalimentare, il presente contributo tenta di elaborare un modello concettuale replicabile che, nel rispetto delle metodologie scientifiche, possa contribuire allo sviluppo di archivi correnti, che poi saranno storici, capaci di gestire flussi digitali tracciabili e sicuri. In quell'assoluta individualità semantica che caratterizza ogni complesso documentario, assoggettabile a “regole” di trattamento solo per affinità produttive e per caratteristiche di provenienza comune, l'applicazione di standard ha costituito materia di discussione per la letteratura scientifica. L'esempio da cui vogliamo muovere riguarda una specifica categoria di Software Packages (LIMS) che vengono usati attualmente in Italia da laboratori di analisi e che attuano una gestione integrata di molteplici dati e processi.

### PAROLE CHIAVE

Modelli concettuali; LIMS; Agroalimentare; Sistemi integrati; Agritech.

\* Le autrici, pur avendo lavorato collegialmente alla redazione del testo, hanno specificatamente redatto: Elena Gonnelli, paragrafo 3, Annantonia Martorano, paragrafo 1 e 2. Le conclusioni sono da imputarsi ad entrambe le autrici.

## 1. Agritech e PNRR

Il settore dell'agroalimentare rappresenta un asset strategico per l'economia del nostro Paese sia in termini numerici (addetti coinvolti, valore complessivo della produzione), sia in un'ottica di valorizzazione del brand Made in Italy e costituisce un indotto fondamentale per il turismo enogastronomico. L'agrifood, del resto, rappresenta una componente imprescindibile del nostro sistema Paese così come evidenziato dall'ultimo rapporto CREA (Consiglio per la ricerca dell'economia agraria) (CREA 2022), che ha confermato il comparto dell'agroalimentare italiano uno dei cardini sui cui poggia la nostra economia. Solo l'agricoltura e l'industria alimentare pesano per oltre il 4% sul PIL nazionale. Se poi vi si aggiungono tutti i settori collegati (commercio, ristorazione e servizi legati al cibo) il sistema si attesta su un peso pari al 15% con un valore complessivo di circa 522 miliardi di euro. L'agricoltura italiana è dunque la prima in Europa per valore aggiunto e la terza per produzione lorda vendibile.

La pandemia COVID19 ha però inciso fortemente comportando una riduzione del valore aggiunto – dati Istat 2020 – (CREA 2022) dell'8,6% rispetto al 10,2% delle attività produttive nel loro complesso. Nonostante questo, l'intero comparto si è dimostrato capace di assicurare le esigenze dei consumatori grazie ad un modello produttivo e gestionale ben rodato. Nell'ultimo periodo inoltre una maggiore consapevolezza degli strumenti digitali ha permesso di affrontare le grandi sfide globali investendo sempre di più nelle nuove tecnologie, al fine di sviluppare un modello di agricoltura 4.0. Questo nuovo paradigma dovrebbe coniugare i fabbisogni crescenti di cibo a livello mondiale, ridurre l'impatto ambientale delle produzioni ma, soprattutto, preservare e tracciare la qualità dei prodotti e delle filiere che, insieme alla sicurezza e alla sostenibilità, si inseriscono nell'ambito delle strategie richieste dal Green Deal europeo.

Qualità dei prodotti, sicurezza alimentare, tracciabilità e sostenibilità devono però essere misurabili e per questo è necessario disporre di dati certi e oggettivi su tutto il ciclo di vita di un prodotto e sulle procedure di analisi e certificazione. Parametrizzare è la parola chiave che dovrà sostenere lo sviluppo di regole uguali e misurabili e consentire di apportare al flusso dei dati una serie di migliorie che avranno come primo beneficio quello di certificare e tracciare la qualità introdotta nella supply chain.

Gli elementi di questo nuovo paradigma sono principalmente: dati verificabili, contratti intelligenti con le parti terze, accesso alle informazioni in tempo reale, registro elettronico decentralizzato, monitoraggio del luogo di provenienza semplificato, migliore trasparenza della supply chain, sostenibilità del prodotto e del processo di produzione e attività di contraffazione.

Il Piano Nazionale di Ripresa e Resilienza (PNRR) ha previsto, infatti, un pacchetto di investimenti e riforme articolato in sei missioni in piena coerenza con i pilastri del Next Generation EU. La missione 1 “Digitalizzazione, innovazione, competitività e cultura” e la missione 2 “Rivoluzione verde e transizione ecologica” coinvolgono in pieno il settore agroalimentare grazie anche alla creazione di un Centro Nazionale per le Tecnologie dell'Agricoltura (Agritech), che avrà la sua sede presso l'Università degli Studi di Napoli Federico II.

La nascente istituzione dovrà svolgere ricerca e promuovere lo sviluppo di tecnologie innovative nel settore agricolo per migliorare quantità e qualità delle produzioni, garantendo l'adattamento sostenibile ai cambiamenti climatici anche attraverso la prevenzione, la resistenza e la resilienza rispetto ai rischi.

Tra le aree maggiormente interessate dal PNRR vi è certamente quella della tracciabilità alimentare in cui si stanno diffondendo tecnologie che puntano su Big Data, Intelligenza Artificiale e Machine learning, Internet of Things, Cloud, Blockchain, 5G e mobile network per ottimizzare il settore agroalimentare. In particolare, la Blockchain viene utilizzata come “diario di bordo” del processo produttivo e dei relativi dati che si creano, in quanto l’adozione di questa tecnologia permette a tutti i protagonisti della filiera di garantire trasparenza, conoscibilità e attendibilità del flusso documentale e dunque del prodotto agroalimentare dalla fase di produzione sino a quella di trasformazione e distribuzione.

Il digitale dunque sta entrando nelle imprese agroalimentari del nostro Paese con un trend positivo così come afferma Chiara Corbo, Direttrice dell’Osservatorio Smart Agrifood Politecnico di Milano, che nel marzo di quest’anno ha rilevato come “il mercato dell’agricoltura 4.0 abbia visto una crescita progressiva, che nel 2020 è arrivata a circa il 20% con investimenti in innovazione dell’ordine di 540 milioni di euro” (Costa 2022).

Va inoltre segnalato che nel febbraio di quest’anno il gruppo di lavoro Filiere produttive 4.0 di Anitec – Assinform in stretta collaborazione con il gruppo di lavoro Agroalimentare dello Steering Committee Innovazione Digitale nelle Filiere di Confindustria Digitale ha realizzato il White Paper *Il Digitale e l’innovazione tecnologica a supporto del settore agrifood italiano* (Confindustria Digitale e Anitec-Assinform 2022). Muovendo dalla consapevolezza che l’innovazione e la trasformazione del sistema agroalimentare siano una parte importantissima della competitività del nostro Paese, il documento mette bene in evidenza come il ruolo dell’innovazione digitale sia fondamentale sia per la tutela e la valorizzazione che per lo sviluppo e il raggiungimento degli obiettivi di sostenibilità e tracciabilità dell’intero comparto agroalimentare.

Oggi più che mai il digitale è un paradigma necessario per garantire la sostenibilità dell’agricoltura 4.0, sia da un punto di vista ambientale che economico. È una sfida importante che bisogna cogliere, ora in questo momento, sia per il sistema Paese che per chi si occupa di innovazione e gestione dei dati.

L’attuale congiuntura politica ed economica legata al PNRR rappresenta un momento cruciale se si pensa al ruolo che può e che dovrà svolgere l’innovazione digitale anche sui temi della tracciabilità, della gestione della qualità, della sicurezza alimentare, della conservazione dei dati e conseguentemente degli archivi che si creeranno e sedimenteranno.

Noi archivisti siamo dunque chiamati a confrontarci con una realtà, quella dell’Agritech, che seppure sembri distante è invece parte integrante della nostra società, del suo sviluppo e di conseguenza parte attiva e politica attraverso i suoi archivi del nostro sistema Paese.

Obiettivo di questo contributo e della relativa ricerca che stiamo sviluppando è quello di realizzare un modello concettuale replicabile che, nel rispetto delle metodologie scientifiche, possa contribuire allo sviluppo di archivi correnti, poi storici, capaci di gestire flussi digitali tracciabili e sicuri sia in ambito archivistico che in ambito agroalimentare.

## 2. Modelli concettuali

I cosiddetti Archivi dell’Agricoltura sono stati indagati da una bibliografia più o meno recente, nazionale e internazionale, anche se l’attenzione è tendenzialmente ricaduta sulle modalità descrittive



tive, sugli aspetti storici, sulle modalità di conservazione (Soprintendenza Archivistica del Lazio 2009; LaMotta 1996). Più raramente si è cercato di riflettere sulle modalità di sedimentazione, sull'archivio corrente o sul workflow di dati, soprattutto se digitali.

Questo approccio invece appare quanto mai impellente, soprattutto nel ramo d'impresa privato che, come è noto, muove da istanze lontane dal fronte culturale e legate a esigenze di mercato, a logiche imprenditoriali che garantiscano la conservazione del patrimonio e soprattutto il lucro tramite meccanismi di efficienza e di produttività.

Come è noto gli attuali sistemi agroalimentari sono permeati da un numero crescente di standard privati relativi a qualità e sicurezza alimentare, operanti accanto alla normativa pubblica, che permettono di soddisfare la crescente domanda espressa dal consumatore, costituendo, al tempo stesso, la base della differenziazione di prodotto nei mercati agroalimentari. Le iniziative private di standardizzazione sono oggi a carico dei grandi *stakeholders* e ne riflettono l'elevato potere di mercato e le strategie competitive. Nell'ambito delle iniziative individuali di differenziazione si sono, quindi, sviluppati standard comunicati al consumatore in maniera diretta, tra i quali si distinguono, in particolare quelli definiti nel contesto di una *relazione verticale cliente-fornitore*. Partendo dagli elementi minimi di qualità richiesti dagli enti certificatori, volti a normalizzare le condizioni di produzione, la decisione strategica di implementare uno standard di matrice privata dovrebbe tenere conto di due aspetti: dello sforzo da intraprendere per consentire un eventuale adeguamento delle caratteristiche della base produttiva e delle opportunità di mercato. Pertanto, l'individuazione dei fattori che motivano le imprese all'implementazione di standard privati, dalle condizioni di efficacia degli stessi e dei relativi effetti sugli altri attori della filiera (produttori a monte e consumatori), richiede un'attenta analisi delle interazioni strategiche tra l'impresa di trasformazione/distribuzione e i produttori a monte, e del comportamento del consumatore (Giraud-Héraud, Grazia, e Hammoudi 2009).

Alla luce di quanto detto è necessaria la creazione di un'infrastruttura per il sistema agroalimentare che deve consentire da un lato l'integrazione tra piattaforme e dall'altro la capacità di superare logiche settoriali. I dati raccolti sul campo devono essere trasformati in informazioni utili, disponibili e verificabili e allo stesso tempo archivisticamente corretti. In quest'ottica è innegabile che il flusso di gestione documentale agroalimentare presenti due distinti, ma complementari, ambiti di modellizzazione a cui guardare, ossia il settore commerciale e di assistenza agronomica e gli asset dei laboratori di prova. Per realizzare tutto ciò le aziende sono state in grado di implementare il loro livello di innovazione digitale applicando processi e logiche 4.0 come ad esempio alcuni software packages denominati LIMS.

### 3. I sistemi informatici LIMS nei laboratori di analisi

Il *Laboratory Information Management System* (LIMS) è un acronimo che identifica una categoria di *software packages* usati nei laboratori di analisi per la gestione integrata di molteplici dati e processi. Il suo utilizzo spazia da laboratori pubblici e privati in tutti i settori merceologici, ma con una particolare attenzione al supporto per le attività di controllo qualità e di ricerca e sviluppo.

Per poter essere utilizzato, e quindi conforme alle procedure sviluppate in un laboratorio accre-

ditato dall'ente certificatore nazionale<sup>1</sup>, il software deve possedere delle funzionalità specifiche, tra le quali un controllo sul personale (dal mansionario, alle abilitazioni, alla formazione), sulle dotazioni del laboratorio (apparecchiature, materiali di consumo e di riferimento), sui prodotti e i servizi forniti dall'esterno.

Una particolare attenzione, però, va anche alla gestione delle offerte e dei contratti con le loro relative specifiche, quali archiviazione, controllo, spedizione e monitoraggio; così come la selezione, la verifica e la validazione dei metodi. Come per qualsiasi sistema dedicato al trattamento e alla gestione dei dati, la garanzia della tracciabilità delle informazioni diventa punto nodale e cardine: la normativa italiana ha, ormai da tempo, riservato una specifica attenzione agli standard internazionali e alle linee guida nazionali relative all'individuazione di quei criteri fondamentali che un sistema informativo deve rispettare per assicurare a enti sia pubblici che privati qualità e affidabilità nell'esercizio della funzione archivistica.

«L'interesse specifico degli archivisti (ma anche dei responsabili di un'amministrazione pubblica o d'impresa) per la costruzione di programmi di gestione documentaria e, soprattutto, l'attenzione della dottrina per lo studio degli archivi in formazione costituiscono un fatto recente, direttamente legato al ruolo crescente dell'informazione, in particolare di quelle informazioni di qualità, significative ed essenziali, affidabili, controllate e non ridondanti, di cui qualunque struttura organizzativa ha ormai bisogno per rispondere ad esigenze di trasparenza ed efficienza e sfruttare adeguatamente le possibilità di comunicazione e conoscenza consentite da un utilizzo avanzato degli strumenti tecnologici e che si traducono nella maggior parte dei casi in produzione di documenti giuridicamente rilevanti.» (Guercio, 2019: 15-16).

I primi sistemi LIMS fanno la loro comparsa sul mercato già dagli anni Ottanta, come strumento di supporto per l'inserimento di dati analitici, ma anche come primordiale automazione di *workflow management*. Oggi si parla di un sistema in continua evoluzione, implementabile in base alle diverse esigenze, ma capace di supportare alcuni dei processi fondamentali propri dell'attività di laboratorio: accettazione dei campioni, assegnazione e *scheduling* dell'attività analitica, attività analitica vera e propria, monitoraggio delle attività, verifica e approvazione dei risultati, *reporting* (rapporti di prova, statistiche, export).

La logica integrata, propria del sistema, permette moltissime attività complementari, quali: la gestione delle specifiche di riferimento (leggi, norme, prodotti, clienti), l'acquisizione dati da strumenti e la loro conferma metrologica (manutenzione e taratura), la gestione dei materiali di magazzino, l'elaborazione e l'analisi delle statistiche, la gestione di preventivi e contratti, la predisposizione di fatture, prove e servizi altri. Non ultimo la creazione, il mantenimento e il continuo aggiornamento di un portale web dedicato ai servizi di laboratorio.

Un sistema LIMS si muove sulla gestione di due elementi: gli oggetti (possono essere campioni in analisi, la strumentazione del laboratorio, i materiali di supporto, oppure le controparti principali del business, come clienti, fornitori, collaboratori...) e gli eventi (il cosiddetto "ciclo di vita" degli

---

<sup>1</sup> Esiste una Banca Dati dei laboratori di prova accreditati, gestita da Accredia e aggiornata dopo ogni seduta del Comitato Settoriale di Accreditamento Laboratori di prova. Accredia è l'Ente Unico nazionale di accreditamento designato dal governo italiano, posto sotto la vigilanza del Ministero dello Sviluppo Economico, in applicazione del Regolamento europeo 765/2008. Esso ha lo scopo di attestare la competenza, l'indipendenza e l'imparzialità degli organismi di certificazione, ispezione e verifica, e dei laboratori di prova e taratura. <https://www.accredia.it> (ultima consultazione maggio 2022).

oggetti). In altre parole l'obiettivo è quello di pianificare, rilevare e monitorare tutte le fasi a cui è sottoposto un campione di analisi permettendo di gestire anche le offerte commerciali, gli strumenti, il campionamento, la fatturazione e così via. La struttura dati si articola quindi attraverso:

1. Matrici, ossia ambiti di prova che possono anche prevedere più livelli ma che condividono, di solito, riferimenti e informazioni comuni (Olio di Oliva, Alimenti e Mangimi per animali, Bevande spiritose, Birra, Vino, Mosti, Aceto, Superfici di lavoro e utensili, Acque e Terreno, Materie prime come i tappi di sughero...);
2. Metodi<sup>2</sup>, visualizzabili con un codice dedicato. Nella compilazione dell'anagrafica denominata "prove del laboratorio", l'utente potrà specificare il/i metodo/i di prova per ogni combinazione e tipo campione (o matrice, prodotto, materiale). Nella compilazione del metodo l'utente potrà, anziché utilizzare la casella descrittiva in libera digitazione, richiamare e codificare direttamente i metodi utilizzati (da normativa o interni) e associarli alla relativa composizione prova/matrice. Questa seconda metodologia, grazie alla presenza di una "libreria" di metodi codificati e utilizzati dal laboratorio di prova, garantisce l'associazione tra il documento ufficiale descrittivo del metodo (ad esempio in formato pdf) e il campo informativo. Ciò consente l'irradiamento delle procedure di dettaglio utili a definire le condizioni strumentali ottimali per le proprie apparecchiature, in linea con la corretta applicazione della norma (le così dette "procedure di prova").
3. Prove, la risultante dell'applicazione di un metodo a una determinata matrice.

[...] il modello assume quindi un ruolo fondamentale come strumento di rappresentazione e di conoscenza della realtà di interesse: è il risultato di un'attività di astrazione con cui si associa all'oggetto o al processo reale una sua rappresentazione mediante costrutti simbolici che abbiano un preciso significato all'interno di un linguaggio formale [...] (Michetti 2019, 342).

La modellazione può, dunque, aiutare a comprendere meglio un sistema complesso di gestione informatica, rappresentando un utile strumento per condividere obiettivi e metodologie, nonché implementarne le potenzialità.

Se partiamo dall'assunto che un modello concettuale sia una rappresentazione astratta di entità, attributi, ma soprattutto relazioni, utile a definire un terreno comune che da una parte riduca il più possibile le ambiguità interpretative e dall'altra realizzi sistemi quanto più omogenei, ci rendiamo conto delle necessità di analizzare i casi specifici per verificare le effettive praticabilità archivistiche. In quell'assoluta individualità semantica che caratterizza ogni complesso documentario, assoggettabile a "regole" di trattamento solo per affinità produttive e per caratteristiche di provenienza comune, l'applicazione di standard ha costituito materia di discussione per la lettera-

---

<sup>2</sup> I metodi per i quali viene richiesto/ottenuto l'accreditamento sono scelti tra quanto normato, in ultima edizione, da organizzazioni nazionali ed internazionali e dalle Gazzette Ufficiali dell'Unione Europea e della Repubblica Italiana. La scelta del metodo è operata dopo un esame di fattibilità relazionato alle risorse disponibili. L'esigenza di adozione ed accreditamento di metodi non ufficiali o normalizzati, è comunque possibile attraverso una selezione operata fra i potenziali metodi noti e pubblicati da Organizzazioni tecniche, Comunità e/o Riviste scientifiche specializzate, Costruttori delle apparecchiature di prova. È possibile accreditare i metodi, mantenendo i riferimenti originali, se le modifiche apportate dal laboratorio ad un metodo ufficiale/normalizzato/non normalizzato sono tali da migliorarne le prestazioni senza snaturarne il principio/tecnica (es. impiego di apparecchiatura con migliore risoluzione, materiali di riferimento e reagenti di più elevata purezza).

tura scientifica. Guardando all'archivistica come alla rappresentazione e organizzazione dell'informazione dentro a uno scenario polifunzionale marcato da una forte eterogenesi dei fini, dovremo allora accettare che la descrizione, per esempio, possa essere vista quale astrazione rappresentativa, capace di contemperare le diverse esigenze: dalla ricostruzione dei contesti alla reperibilità dei contenuti. Sarà allora necessaria una concezione più dinamica e aperta dell'archivio che tenga conto dei recenti dibattiti così come delle problematizzazioni teoriche e, ancora, della polimorfia documentale e archivistica che contraddistingue il contesto, perlomeno, italiano.

Un articolo pubblicato su «Archival Science» nel 2002 presenta due diversi punti di vista opposti sulla descrizione (uno basato sul *respect des fonds* e l'altro incentrato sulle serie) per mettere in evidenza e spiegare le diverse prospettive che sottendono e influenzano gran parte del dibattito di cui si faceva riferimento poc'anzi (Duff e Harris 2002).

Se, come sappiamo, la descrizione archivistica "tradizionale" è tipicamente composta da dati strutturati, da testi narrativi/descrittivi che portano con sé forti componenti interpretative e soggettive, ragionando in prima battuta in termini di struttura; l'evoluzione recente, figlia invece dell'interoperabilità, cerca ibridazione di contenuti nella diversificazione degli oggetti finali della descrizione, impone una maggiore parcellizzazione dei dati, nonché tende al riuso e alla condivisione.

Il punto critico, allora, sta nell'equilibrio: sul confine tra gerarchie contestualizzanti e contenuti, distribuiti alla base e non lungo i lati della piramide. Un equilibrio da ricercare nell'affidabilità contestuale così come nei contenuti informativi non necessariamente cristallizzati in una gerarchia. Il rischio è inevitabile e va dalla decontestualizzazione del dato, per esempio, a temi quali affidabilità e autorevolezza (Duranti e Rogers 2012; 2014).

In un certo senso i documenti sono sempre in fase di creazione, poiché le "loro" storie non finiscono mai, così come le storie di coloro che sono convenzionalmente chiamati soggetti produttori, record managers, archivisti, utenti e così via, sono parti (mutevoli e mescolate) di storie più grandi comprensibili solo nei contesti più ampi e, a loro volta, mutevoli della società. In senso ampio, l'esigenza odierna, enfatizzata dal digitale, è quella di documentare la documentazione (Alfieri 2020): ci serve descrivere anche i processi (dall'audit ai paradata), oltre alle informazioni e alle loro relazioni.

Alla luce di quanto detto, l'idea di modello concettuale punta alla ricerca di equilibrio tra le diverse componenti funzionali, nel tentativo di spostare l'attenzione verso un *record management* intelligente e consapevole dei problemi di accesso e conservazione (Bilotto e Perondi 2008).

Nell'ingegneria del software, più precisamente nell'ingegneria dei requisiti (disciplina dedicata alla identificazione, comprensione e documentazione dei problemi, necessità e requisiti di un software), si definisce come requisito ogni richiesta, necessità, processo o funzionalità di un sistema: essi sono distinti tra funzionali (tutti i bisogni, necessità, pure le caratteristiche o le funzionalità previste in un processo che può essere soddisfatto dal software) e non funzionali (proprietà del sistema che devono essere soddisfatte). Tra questi si segnalano i requisiti necessari per una corretta impostazione:

- garanzia dell'imparzialità rispetto al processo di auto-documentazione del soggetto produttore in rapporto alle sue attività svolte e, allo stesso tempo, confermi anche l'autenticità connessa al reale bisogno di documentazione e che il sistema documentario deve conservare sia nel momento della gestione attiva che nelle fasi successive di trattamento e tenuta.

Come d'altra parte chiarisce anche MoReq2 (Sez. 2.2.)<sup>3</sup> il fatto che un sistema di gestione assicuri la conservazione di *authoritative records* o di documenti di riferimento principale (ISO 15489-I) è un concetto cruciale perché il «loro contenuto è fidato in quanto ciascuno di essi costituisce completa e accurata rappresentazione delle fasi dei procedimenti, delle attività, degli atti di cui è la testimonianza» (Guercio 2019, 109).

- utilizzo di vocabolari controllati conformi agli standard ISO 2788 e 5694 (mono e multilingue);
- predisposizione di tutte le funzionalità che supportino il flusso di dati del laboratorio. Oggi un LIMS deve essere un *Enterprise resource planning* (letteralmente “pianificazione delle risorse d’impresa”);
- facilità di uso per l’operatore. Sistemi eccessivamente complessi possono essere accettati con più difficoltà dagli operatori, seppur formati;
- Integrità con l’infrastruttura gestionale pre esistente;
- scalabilità sia in termini verticali (specializzazioni) che in termini orizzontali (estensioni), ma anche flessibile in modo da potersi adattare con facilità alle esigenze evolutive di ogni singolo laboratorio;
- possibilità di avere una configurazione personalizzabile direttamente dall’utente, nonché implementabile a costi contenuti. Il tentativo è quello di evitare l’obsolescenza;

Ne consegue che all’interno di un laboratorio di analisi, il sistema LIMS permetterà la gestione dei dati scandendo cinque fasi fondamentali: 1. accettazione campioni, 2. assegnazione e scheduling dell’attività analitica, 3. attività analitica vera e propria, 4. reporting o rapporti di prova, 5. archiviazione.

La fase 1 prevede che il cliente mandi il campione all’accettazione dove un addetto specializzato aprirà la scheda cliente, vedrà il contratto e assocerà i dati delle analisi richieste con i dati commerciali. Si tratta della prima registrazione, una fase molto importante e delicata che sarà totalmente a carico del responsabile della gestione in ingresso così come sarà di sua competenza l’identificazione univoca dei campioni da sottoporre a prova.

L’offerta ingressata all’interno del LIMS dovrà contenere, per la corretta gestione digitale del flusso, un set di dati minimi, quali: il nome dell’azienda alla quale si deve addebitare il servizio, l’identificazione univoca del campione; l’identificazione propria del cliente; le analisi richieste.

Nel caso in cui un cliente voglia effettuare una pre-accettazione on line, inserendo i dati in maniera autonoma, potrà farlo entrando nella sua area riservata con password e username. In tempo reale il sistema genererà un *alert* per il personale dedicato all’interno dell’azienda, che dovrà – in fase di acquisizione reale del campione – verificarne solo la corrispondenza e stampare il codice identificativo generato dal sistema. In questo modo, gli elementi costitutivi del documento entrano

---

<sup>3</sup> *Model Requirements for the Management of Electronic Record* (d’ora in avanti MoREQ e la successiva revisione MoREQ2) fornisce le linee guida per l’utilizzo di tecnologie per la gestione informatica dei documenti mediante la definizione dei requisiti funzionali di un *Electronic Records Management System* sia in ambito pubblico che privato. <https://moreq.info>.

a far parte del “profilo”<sup>4</sup> che accompagna ogni unità documentaria e che deve includere tutti i metadati utili alla sua identificazione e gestione dall’inizio del processo fino alle fasi finali.

È al termine dell’accettazione che si apre la seconda fase, quella di assegnazione e *scheduling* dell’attività, generando un foglio di accettazione commessa recante tutti i dati e recapitato in formato digitale al cliente.

Il sistema dovrà essere predisposto per assegnare automaticamente a ogni singolo campione la seguente codifica AANNNNNN (dove AA identifica l’anno di riferimento e NNNNNN il numero progressivo di identificazione del campione). Tale numero di riferimento sarà da stampare su un’etichetta adesiva, applicata sul campione, che seguirà tutte le fasi successive.

Questa prima codifica è da considerarsi come un numero di protocollo, composto da data e numero progressivo che si andrà integrando con il codice del metodo di prova, la data di stampa della effettuazione della prova, il codice campione e il nome analista.

L’assegnazione delle autorizzazioni e delle relative responsabilità avverrà in automatico a seguito dei privilegi impostati nel LIMS, così da permettere anche la visualizzazione delle relative prove e informazioni a seconda delle fasi di lavorazione del campione.

Sarà infine utile che il software permetta di esportare una serie di campioni aventi lo stesso metodo su un foglio elettronico recante, oltre al metodo stesso, il parametro, il numero del campione, nome e cognome del tecnico nonché la sua firma digitale.

All’avvio della terza fase, ossia quella dell’attività analitica vera e propria, è necessario tenere in considerazione i seguenti aspetti diversi:

- l’input dei valori. Quando il tecnico di laboratorio procede ad effettuare le analisi secondo la metodica di prova, leggerà i dati ottenuti dall’analisi, eseguirà i calcoli necessari per elaborarne i risultati e trascriverli sul foglio di lavoro in corrispondenza di ciascun campione. Successivamente gli stessi saranno da inserire nel computer manualmente o da importare per via informatica. Durante la fase di importazione dei risultati dal foglio elettronico è importante che il LIMS provveda ad allegare il link al file di importazione e che tale documento non possa più essere modificato. Solo a questo punto il tecnico potrà firmare il modulo ad evidenza del completamento e a verifica della prova.
- il monitoraggio delle attività. Il responsabile di laboratorio deve controllare giornalmente, o più volte durante la giornata, che le prove siano eseguite secondo quanto programmato, verificando che i moduli di lavoro corrispondano al numero di determinazioni previste. Un “Cruscotto di controllo” permetterà la supervisione dello stato di avanzamento dei campioni, tenendo sotto controllo le performance analitiche e la loro coerenza.
- la verifica/approvazione dei risultati. A fine giornata e comunque prima della stesura del rapporto di prova, il responsabile deve verificare che vi sia corrispondenza tra i dati riportati sul foglio di lavoro e quelli presenti nel computer. Sarà in questa sede che dovrà provvedere alla vidimazione dei risultati accedendo nel software di laboratorio al livello “vidimatore”, tramite un’apposita password. Il rapporto di prova a quel punto

---

<sup>4</sup> Il termine profilo fa ampiamente parte del vocabolario informatico, mentre entra nel lessico archivistico negli anni Novanta, poi approfondito dallo studio europeo MoReq2: «Il profilo ha, in sostanza, la forma di un modulo elettronico collegato a ogni documento d’archivio e mantenuto in tutte le fasi di gestione» (Guercio 2019, 35). Per più ampie e dettagliate indicazioni cfr. (Guercio 2007).

sarà emesso soltanto quando tutti i valori delle analisi associate al campione saranno stati validati.

Ogni laboratorio di prova possiede un sistema di registrazione automatica dei dati delle prove eseguite: per questa quarta fase dovrà essere utilizzato uno specifico software che archivi in un'apposita cartella sita all'interno del server e che, più in generale, si integri a tutti i livelli con il sistema LIMS. La gestione documentaria all'interno dei laboratori di analisi potrebbe essere definita come un sistema ibrido, in quanto oltre ai supporti magnetici sono utilizzati anche quelli cartacei, anche se tutti i dati storici e di funzionamento sono oggetto di back-up periodici e i rapporti di prova vengono compilati elettronicamente, vidimati e firmati digitalmente dal responsabile del laboratorio stesso. Le correzioni a mano sono permesse solo sui documenti di registrazione cartacei e devono essere contrassegnate da data e firma di chi esegue le correzioni senza rendere illeggibile la parte annullata, mantenendo sempre il motivo della correzione; negli altri casi deve essere emessa una revisione del documento.

L'ultima fase, quella dell'archiviazione di tutti i documenti è da effettuarsi in modo informatico nel server aziendale ogni qualvolta ne sussista la possibilità e, possibilmente demandata a un Responsabile del servizio di conservazione (art.6, comma 5, DPCM 3/12/2013).

Normalmente i documenti del sistema di qualità sono contenuti in una specifica cartella, nella quale vi è una suddivisione tra documenti prescrittivi, documenti di registrazione, documenti esterni e documenti superati, a loro volta divisi secondo un criterio funzionale. Nel caso di registrazioni in cui non è possibile un'archiviazione elettronica, sono conservate in originale dal Responsabile di Settore, in apposito archivio.

Un sistema LIMS, per adempiere a questo scopo, dovrà permettere una distribuzione interna, elettronica, dei documenti, gestita attraverso una base dati XML che contenga l'elenco di tutta la documentazione con i links ai files, che sia accessibile da tutto il personale e che faccia riferimento sempre alle ultime revisioni. L'accesso a documenti elencati dovrà essere tuttavia limitato e garantito tramite password o protezione dell'accesso delle cartelle del server.

I documenti del sistema di gestione, invece, devono essere univocamente identificati, in quanto riportano un codice identificativo, l'indice di revisione, la data di revisione, un titolo, una numerazione su ogni pagina.

La codifica prevede un format tipo: XX 1.2.3-n, in cui XX corrisponde alla sigla della tipologia di documento (per esempio, PG procedura; MD modulo di registrazione o di pianificazione; DT documento tecnico o di controllo; PdC piano di controllo) e 1.2.3 alla numerazione del documento contenente il riferimento allo specifico paragrafo della norma UNI CEI EN ISO/IEC 17025 (capitolo-paragrafo-sottoparagrafo); n, infine, identifica il numero progressivo o la lettera.

Ovviamente alcuni documenti tecnici potranno avere una nomenclatura particolare: per gli ordini a fornitori viene attribuito un numero progressivo per anno; per le offerte e per le accettazioni d'ordine si identifica univocamente il documento con una sigla numerica che rappresenti sia il numero di protocollo che il numero di revisione dell'offerta, insieme all'anno di emissione della stessa. L'esigenza fondamentale di controllare gli accessi e gli interventi a e sul sistema documentario deve essere assicurata da specifiche *policy* sulla sicurezza: tutti i dati relativi ai clienti sono trasferiti all'interno di uno specifico software, accessibile solamente da personale autorizzato attraverso l'uso di apposite password.

I privilegi di accesso allo stesso software sono tuttavia attribuiti in base ai ruoli e alle responsabilità degli utenti all'interno della struttura. È noto come in questo ambito il modello concettuale di riferimento debba essere l'*Open Archival Information System* (OAIS) - standard ISO 14721:2003: la sua flessibilità per una *comunità designata* fa sì che si adatti alle esigenze binarie di utente e di soggetto produttore (Pigliapoco 2005).

I documenti generati dal LIMS (per esempio i rapporti di prova, i fogli di lavoro, i preventivi e i contratti, le fatture, ecc.) così come tutti gli allegati (verbali, row-data, cromatogrammi, foto...), dovranno essere automaticamente archiviati, completi dei metadati associati a ogni categoria documentale (ISO23081).

La normativa italiana prevede alcuni metadati minimi che devono essere associati al documento informatico: identificativo, data di chiusura, oggetto, soggetto produttore, destinatario. Se tali elementi servono per attribuire al documento un'identità ben precisa, è facile intuire perché la fase di formazione dell'archivio sia ritenuta cruciale per la corretta tenuta e conservazione della documentazione digitale (Pigliapoco 2010).

Tuttavia lo stesso documento, una volta formato, avrà un suo percorso di esistenza che lo porterà ad essere gestito da diversi sistemi e applicazioni con la contestuale produzione e associazione di ulteriori informazioni: i metadati tendono allora ad accumularsi nel corso della vita degli stessi per tracciarne l'utilizzo, ad esempio gli accessi, le modifiche, i trasferimenti, le copie, nonché le modalità della sua conservazione (*Preservation description information*).

## Conclusioni

Già nel 160 A.C., Marco Porcio Catone, detto il Censore, nella sua opera "Liber de Agri cultura" rimarcava il ruolo fondamentale dell'agricoltura, sul piano sociale, morale ed educativo, ma anche su quello del profitto economico. A distanza di più di due millenni, l'affermazione è ancora attuale: l'agricoltura rimane uno degli asset strategici di un Paese come l'Italia, caratterizzato da una straordinaria diversità biologica ed una forte cultura agroalimentare 4.0.

Come abbiamo visto in apertura, per questo settore investire in innovazione è certamente una delle leve strategiche sui cui puntare per ottenere processi produttivi più efficienti. In questo senso il ruolo delle soluzioni digitali archivistiche si presenta come fondamentale in particolare all'interno di attività tecniche produttive che prevedono l'analisi dei dati agroalimentari e la loro corretta sedimentazione e gestione.

In questo lavoro abbiamo immaginato un modello concettuale, che grazie all'uso delle tecnologie digitali orientate all'utilizzo e alla gestione dei flussi archivistici di dati - abilitatori di un nuovo modello di business - potesse essere capace di creare una sinergia costruttiva tra Archivistica e Agritech, al fine di realizzare archivi correnti capaci di permanere nel tempo ed essere testimoni di un momento storico senza eguali, ma soprattutto archivi capaci di rendere credibile e sostenibile il grande patrimonio informativo in esso contenuto.



## Riferimenti bibliografici

- Alfieri, Alessandro. 2020. *Il sistema di documentazione digitale*. Inarchivio 2. Milano: Editrice bibliografica.
- Bilotto, Antonella, e Gianluca Perondi. 2008. «Gli archivi informatici nelle imprese». *Digitalia* 3 (1): 90-97.
- Confindustria Digitale, e Anitec-Assinform. 2022. «Il Digitale e l'innovazione tecnologica a supporto del settore agrifood italiano. White Paper». Milano: Anitec-Assinform. [https://www.innovation-nation.it/wp-content/uploads/2022/03/wp\\_agrifood\\_def.pdf](https://www.innovation-nation.it/wp-content/uploads/2022/03/wp_agrifood_def.pdf).
- Costa, Claudia. 2022. «Innovare il settore agrifood per sostenibilità, tracciabilità e sicurezza». *Agrifood.Tech*. 7 marzo 2022. <https://www.agrifood.tech/digital-farming/innovare-il-settore-agrifood-con-il-digitale-per-indirizzare-le-sfide-globali-di-sostenibilita-tracciabilita-e-sicurezza/>.
- CREA. 2022. «Report 2020». Ministero Politiche Agricole alimentari e forestali. <https://www.crea.gov.it/media-kit>.
- Duff, Wendy M., e Verne Harris. 2002. «Stories and Names: Archival Description as Narrating Records and Constructing Meanings». *Archival Science* 2 (3): 263–85. <https://doi.org/10.1007/BF02435625>.
- Duranti, Luciana, e Corinne Rogers. 2012. «Trust in Digital Records: An Increasingly Cloudy Legal Area». *Computer Law & Security Review* 28 (5): 522–31. <https://doi.org/10.1016/j.clsr.2012.07.009>.
- . 2014. «Trust in Records and Data Online». In *Integrity in government through records management: essays in honour of Anne Thurston*, 203–16. London: Routledge.
- Giraud-Héraud, Eric, Cristina Grazia, e Abdelhakim Hammoudi. 2009. «Comportamento strategico e organizzazione della qualità nelle filiere agroalimentari». *Agriregionieuropa*, 2009. <https://agrireregionieuropa.univpm.it/en/content/article/31/17/comportamento-strategico-e-organizzazione-della-qualita-nelle-filiere?qt-eventi=1>.
- Guercio, Maria. 2007. «MoReq2 Specifications: una sintesi commentata dei nuovi requisiti modello per la gestione di documenti elettronici». *Archivi&Computer* 2–3: 64–86.
- . 2019. *Archivistica informatica: i documenti in ambiente digitale*. Aulamagna 74. Roma: Carocci editore.
- LaMotta, Greg. 1996. «Electronic Records in the National Archives Relating to the History of Agriculture and Rural America». *Agricultural History* 70 (2): 439–48.
- Michetti, Giovanni. 2019. «Metodologie di analisi per l'automazione dei sistemi documentari». In *Archivistica informatica: i documenti in ambiente digitale*, 337–74. Roma: Carocci.
- Pigliapoco, Stefano. 2005. *La memoria digitale delle amministrazioni pubbliche: requisiti, metodi e sistemi per la produzione, archiviazione e conservazione dei documenti informatici*. Santarcangelo di Romagna (RN): Maggioli.
- . a c. di. 2010. *Conservare il digitale: Convegno Conservare il Digitale. Riflessioni su Modelli Archivistici, Figure Professionali e Soluzioni Applicative*, svoltosi a Macerata il 7 e 8 maggio 2009. 1. ed. EUM x Archivistica informatica. Macerata: EUM.

Soprintendenza Archivistica del Lazio. 2009. *Gli archivi dell'agricoltura del territorio di Roma e del Lazio*. A cura di Stefano Lepre. Vol. 96. Pubblicazioni degli Archivi di Stato. Saggi. Roma: Ministero per i beni e le attività culturali, direzione generale per gli archivi. [http://archivi.beniculturali.it/dga/uploads/documents/Saggi/Saggi\\_96.pdf](http://archivi.beniculturali.it/dga/uploads/documents/Saggi/Saggi_96.pdf).