

Wikibase, or The search for the unicorn

Giovanni Bergamin^(a)

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

Contact: Giovanni Bergamin, 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 Wikidata¹. 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.² 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).

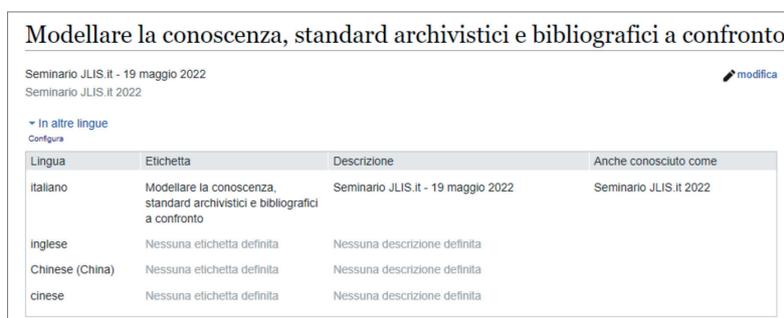
In questo intervento l’interesse per Wikibase sarà centrato su un particolare aspetto della piattaforma tecnologica, ovvero sul modello dei dati: il *Wikibase data model*.³ Si tratta in realtà di un metamodello che offre la struttura di base per costruire ontologie relative a particolari domini (ad esempio 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)*.



Modellare la conoscenza, standard archivistici e bibliografici a confronto

Seminario JLIS.it - 19 maggio 2022 modifica

Seminario JLIS.it 2022

▼ In altre lingue Configura

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)

¹ <https://wikidata.org>.

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

³ 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 Turtle⁴ – 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 LOD⁵ 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”;⁶
- 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;⁷

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

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

⁶ <https://www.w3.org/TR/rdf-schema/>

⁷ <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*).⁸

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).⁹

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.¹⁰ 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 esempio¹¹ 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 livelli¹⁴ *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*.¹⁵ Viene citato anche il

⁸ <https://schema.org>.

⁹ Il corsivo è stato aggiunto per migliorare la leggibilità.

¹⁰ 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).

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

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

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

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

¹⁵ «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*¹⁶ 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*.¹⁷

FRBR/RDA/LRM	BIBFRAME	Wikidata
<i>frbr:work</i>	<i>bf:work</i> ¹⁸	<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 LRM¹⁹ 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. La versione consultata è del 29 marzo 2022, ore 21:49 disponibile su <https://tinyurl.com/2p87mbwb>.

¹⁶ «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.

¹⁷ 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.

¹⁸ «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).

¹⁹ 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*²⁰ 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 *nomen*» (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.²¹

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*.²² 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*.²³

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

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

²² L'elenco completo in (IFLA 2020: 53-58).

²³ 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». ²⁴

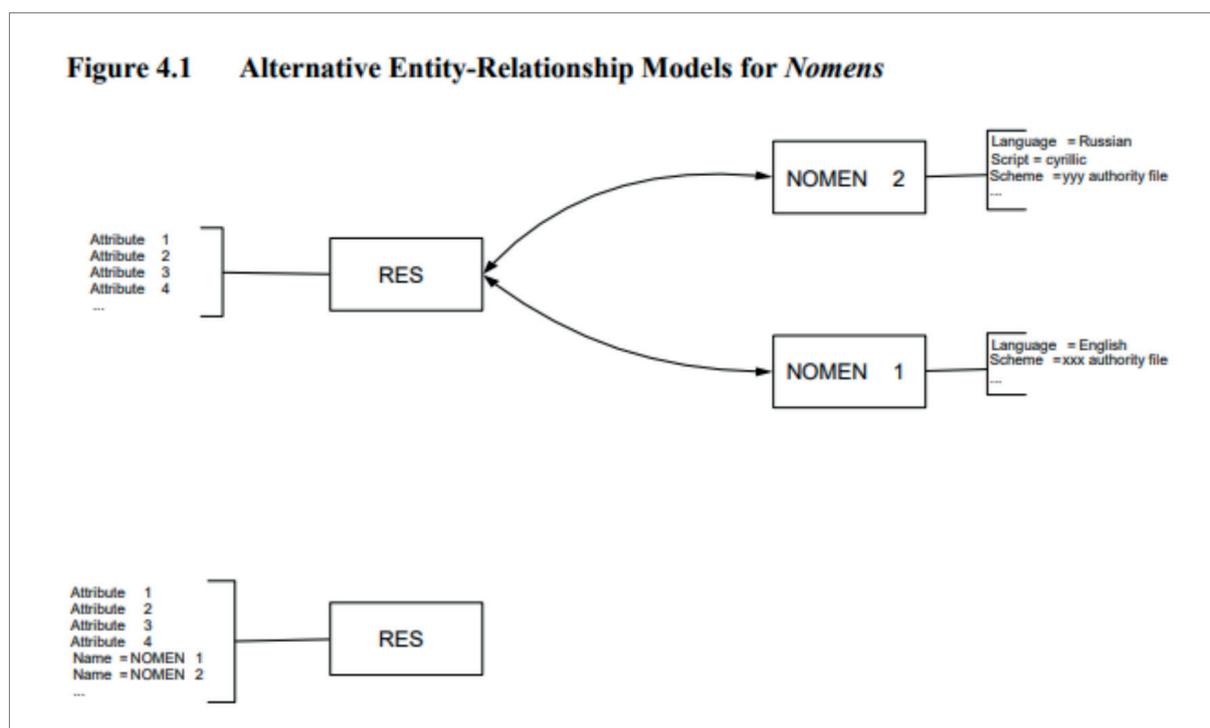


Figura 5: Modelli alternativi per “dare un nome” alle entità

²⁴ 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.²⁵ 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.²⁶ 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).²⁷ 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.²⁸

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.²⁹ 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

²⁵ (Tennant 2002).

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

²⁷ 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).

²⁸ «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.

²⁹ «[...] 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.³⁰

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*.³¹ 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).

³⁰ «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).

³¹ 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]).³² 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).

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

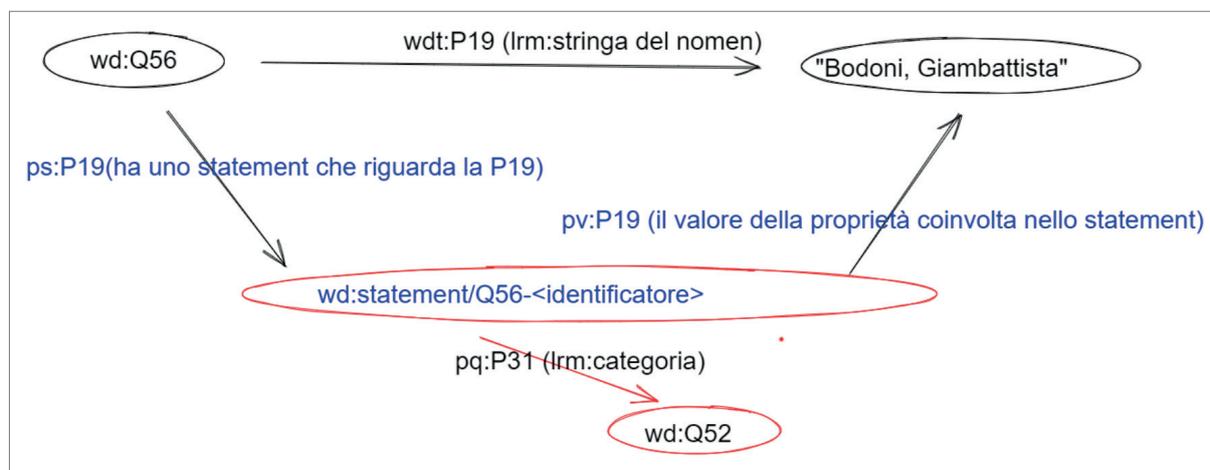


Figura 7: Sperimentazione IFLA LRM *nomen*: in blu l'entità *statement* e il collegamento alla tripla di partenza; in rosso esempio di tripla che supporta gli attributi che LRM prevede per l'entità *nomen* (nel *Wikibase data model* equivalente a quel determinato *statement*)

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.³³ 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

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

ben noto acronimo GAFAM³⁴. 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*.³⁵

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.³⁶

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.

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

³⁵ 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).

³⁶ 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.
- 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.
- 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.
- 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.
- 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>.