Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Nuovi modi di produrre e condividere l'informazione bibliografica: una sperimentazione dell'uso del Wikibase Data Model per i dati UNIMARC

Giovanni Bergamin, Cristian Bacchi

Contatti: Giovanni Bergamin, giovanni.bergamin@gmail.com; Cristian Bacchi, cristian.bacchi@gmail.com

ABSTRACT

A partire dalla considerazione che l'UNIMARC (e in generale il MARC) è di fatto una ontologia, questo contributo si propone di esplicitare e convertire - solo a livello sintattico - l'ontologia del MARC in strutture Linked Data/RDF mediante l'uso del Wikibase data model. Il punto di arrivo potrebbe essere non solo la pubblicazione dei dati come LOD, ma anche un ambiente per la produzione dei dati bibliografici che permette la coesistenza di differenti visioni ontologiche.

Si illustra la possibilità di realizzare una ristrutturazione del record UNIMARC in Item distinti per tipologie di dato (potenzialmente riferibili anche alle diverse entità FRBR), mantenendo la possibilità di recuperare tutte le informazioni del formato originale.

Quindi si evidenziano le soluzioni di Wikibase che diventano sfruttabili per il MARC: "usable version" del record, con esplicitazione dei valori codificati, e definizioni collegate al dato nel medesimo sistema; individuazione dei dati univoci mediante URI, così come richiesto nel contesto del Semantic Web; puntatore alla fonte del dato, al livello di ogni singolo campo; statistiche sulla presenza di campi e sottocampi; formato di archiviazione nativamente ideato per la condivisione e la modifica partecipata; export di tutti gli elementi descrittivi in RDF standard; supporto di modifica dall'esterno attraverso API aperte.

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Introduzione

L'obiettivo di questo lavoro è dare un contributo alle iniziative in corso sui nuovi modi di produrre e di condividere i dati bibliografici. L'iniziativa più nota, oggi in fase di avanzata sperimentazione, è sicuramente BIBFRAME (Bibliographic Framework Initiative) che, nei suoi obiettivi, definisce anche una possibile direzione di un percorso di transizione: non si tratta di perseguire l'obiettivo della *fine del MARC* (Tennant 2017), ma di valorizzare nel W*eb dei dati* la solidità di un formato di scambio che da oltre mezzo secolo ci ha assicurato (e continua ad assicurarci) la condivisione (sia a livello di produzione che di fruizione) delle risorse bibliografiche.[1]

Il nuovo contesto delle iniziative di transizione sono le tecnologie note con l'espressione *linked data* o *linked open data* che si affermano a partire dal 2006 con l'obiettivo di rendere operative nella vita quotidiana le visioni del Web semantico, proposte agli inizi di questo secolo: affiancare al *Web dei documenti* un *Web dei dati*.[2]

La tesi di questo contributo è che il MARC o "catalogazione leggibile da macchine"[3] ha tutti i titoli per affermare la sua presenza nel Web semantico che - in questo contesto - potrebbe essere definito come "web leggibile e comprensibile da macchine" dove ovviamente le macchine non leggono e non capiscono, ma sono solo in grado di elaborare informazioni per obiettivi pratici che noi decidiamo.

È opportuno brevemente ricordare qui tre espressioni che vengono usate nel contesto del trattamento informatizzato dell'informazione bibliografica: *formato*, *schema di metadati* e *ontologia*. Si tratta di espressioni che si riferiscono da diversi punti di vista allo stesso contenuto informativo: l'informazione bibliografica può essere strutturata solo grazie a un determinato *formato*[4]; il formato è definito grazie a un determinato *schema di metadati* (o *insieme definito* di metadati); con l'affermarsi del Web semantico e dei *Linked data* l'espressione *ontologia* ha sostituito l'espressione *schema di metadati* (Coyle 2012, p. 15). Come è noto il termine *ontologia* (nel contesto informatico) fa riferimento a modelli di rappresentazione formale di un determinato dominio di interesse basati principalmente su RDF (Resource Description Framework)[5] che - vale la pena ricordare - è un modello (o grammatica[6]) standard per lo scambio di dati sul web. In questo contesto si può dire che ad esempio lo standard MARC21 è *di fatto* una ontologia (oltre che un *insieme definito* o *schema* di metadati rappresentabile con sintassi ISO 2709, XML o JSON).

I percorsi di transizione - dal MARC al *Web dei dati* - fino a oggi hanno seguito sostanzialmente due strade: 1) *ristrutturazione semantica dell'informazione* (senza garantire un ritorno al formato di origine); 2) *mappatura sintattica di tutti gli elementi* del MARC (con garanzia di ritorno). La prima

---

[1] "A major focus of the initiative will be to determine a transition path for the MARC 21 formats while preserving a robust data exchange that has supported resource sharing and cataloging cost savings in recent decades" http://loc.gov/bibframe.

[2] Il riferimento di base per i Linked data è: Berners-Lee 2006; per il Web semantico e il Web dei dati: Berners-Lee, Hendler, Lassila 2001.

[3] https://www.loc.gov/marc/faq.html#definition: "MARC is the acronym for MAchine-Readable Cataloging".

[4] Formato come "structured information about an information resource" dove "structured information" significa che l'informazione "must be recorded in accordance with some documented metadata scheme" (Caplan 2003. p. 3).

[5] Proposto dal W3C nel 1997: https://www.w3.org/TR/WD-rdf-syntax-971002/; qui la versione corrente https://www.w3.org/RDF.

[6] "La lingua del web semantico ha una grammatica. Questa grammatica, chiamata Resource Description Framework (RDF), definisce una struttura per esprimere frasi – nel linguaggio tecnico: "statements"– secondo un semplice modello tripartito [...] composto da soggetto [...], un predicato [...] e un oggetto" (Baker 2013).

strada è stata oggetto di molte iniziative documentate in letteratura (Hallo, Luján-Mora, Maté, Trujillo 2016) che si sono occupate principalmente del modo di pubblicare i dati bibliografici su web, con l'eccezione di BIBFRAME che sta affrontando con sistematicità anche il modo di produrre i dati. La seconda strada è stata percorsa con la mappatura completa in RDF del formato MARC21[7], ma non sembra essere stata sperimentata in nessun progetto. Il primo percorso parte dal MARC, ma persegue programmaticamente un cambio di ontologia di riferimento, il secondo è invece interessato alla "traduzione" del MARC nel linguaggio del *Web dei dati* e a mantenere l'ontologia di partenza.
Il contributo che qui viene proposto vuole perseguire il secondo obiettivo (mappatura completa e garanzia di ritorno) prendendo come punto di partenza un *modello di successo* non proveniente dal dominio bibliografico, ma che ha un forte impatto nel modo di produrre, di condividere e di riusare l'informazione nel web di oggi. Il riferimento è all'iniziativa Wikidata e a tutta l'infrastruttura (modelli di dati e soluzioni tecnologiche riusabili) che la accompagna.

Wikidata, Wikibase e modello di dati

Wikidata è una "Knowledge-Base collaborativa" (*collaboratively edited knowledge base* o *free linked database*) mantenuta dalla Wikimedia Foundation e aperta in consultazione e modifica a utenti e sistemi automatizzati (i *bot*).[8] Wikidata svolge una funzione di archiviazione centrale dei "dati strutturati" contenuti nei diversi progetti Wikimedia, quali Wikipedia, Wikiquote o Wikisource: vi si archiviano, cioè, i dati rappresentabili mediante la tripla del web semantico (soggetto, proprietà, oggetto), come la data di nascita di un personaggio.
La centralizzazione dei dati comuni è mirata a facilitare la loro manutenzione in modo indipendente dalle specificità dei diversi progetti Wikimedia (per esempio le diverse lingue) e ad agevolare il loro riutilizzo in nuovi contesti, con un'impostazione Linked Open Data (Vrandečić, Krötzsch 2014).
L'iniziativa è stata avviata da un gruppo di ricercatori di Wikimedia Deutschland, con il supporto della Paul Allen Institute for Artificial Intelligence, della Gordon and Betty Moore Foundation e di Google.
Dal punto di vista dell'utenza, la centralizzazione dei "dati strutturati" può essere verificata considerando un esempio come la voce "Dante Alighieri" in Wikipedia: le pagine nelle diverse lingue di tale voce Wikipedia, contengono sia descrizioni discorsive sia "dati strutturati";

Figura 1: voce "Dante Alighieri" in Wikipedia, in 2 diverse lingue

[7] http://www.marc21rdf.info.
[8] Sito di riferimento dell'iniziativa: https://www.wikidata.org/wiki/Wikidata:Main_Page.

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

tali pagine Wikipedia rimandano tutte alla medesima "pagina" (*Item*) Wikidata della voce "Dante Alighieri"[9] in cui sono raccolti i "dati strutturati".

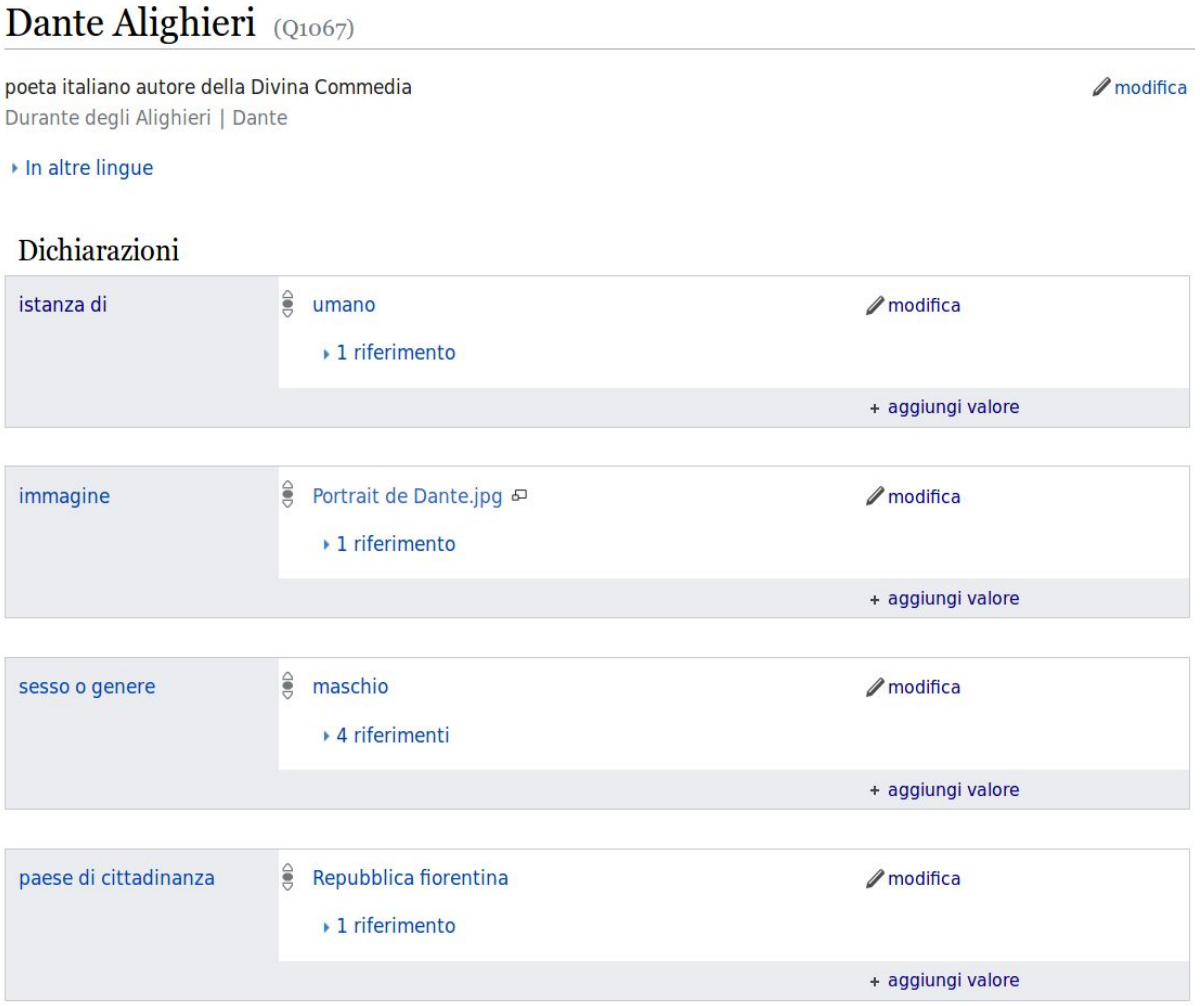

Figura 2: *Item* "Dante Alighieri" in Wikidata

[9] *Item* Wikidata su Dante Alighieri: https://www.wikidata.org/wiki/Q1067.

I puntatori alle diverse pagine dei progetti Wikimedia sono gestiti nella sezione *Link* dell'*Item*; questa sezione rende l'idea della funzione svolta da Wikidata a supporto della manutenzione dei diversi progetti Wikimedia: per ciascun progetto Wikimedia (Wikipedia, Wikibooks, Wikinews, Wikiquote) vengono raccolti i link alle pagine nelle diverse lingue.

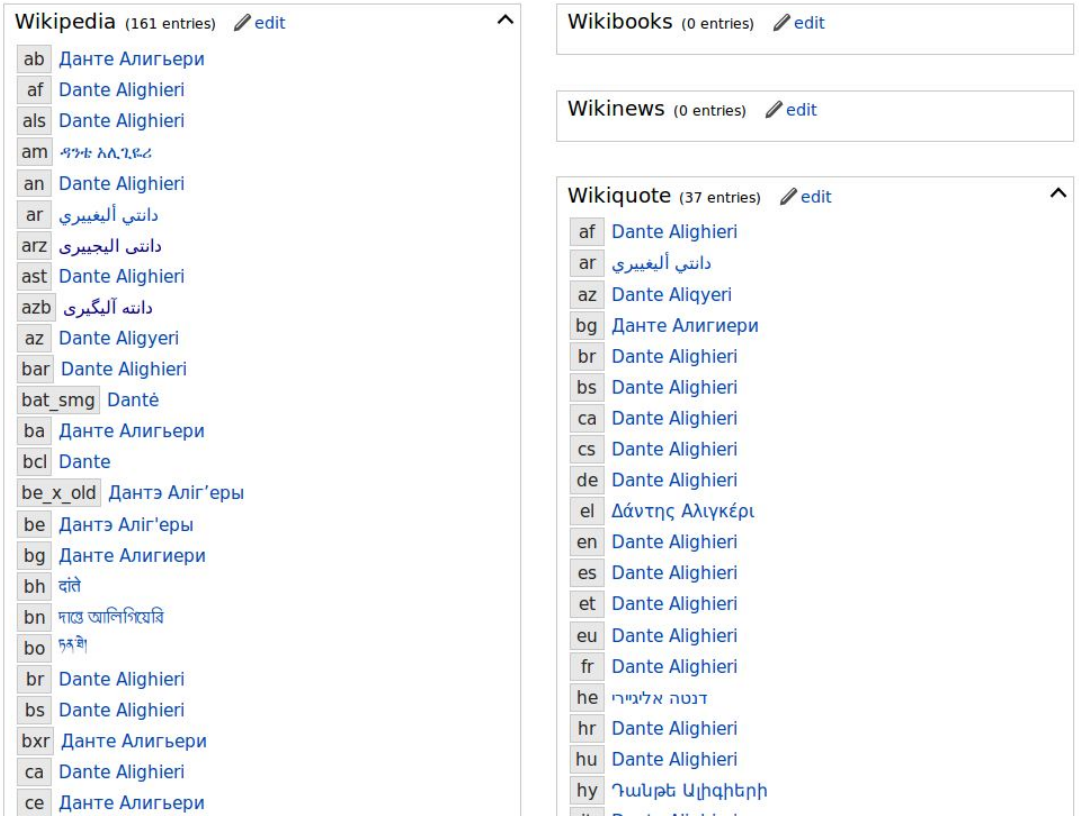

Figura 3: sezione *Link* dell'*Item*

Wikibase Data Model

Il servizio di Wikidata viene gestito con il software libero Wikibase[10] e si basa su di un modello di dati, il *Wikibase Data Model*,[11] che archivia in maniera uniforme tutte le unità informative (le *Entity*) distinguendo le specifiche caratteristiche (*Property*) dai soggetti descritti o messi in relazione (gli *Item*).

[10] Sito di riferimento del software: http://wikiba.se.

[11] Per lo studio del *Wikibase Data Model* si è fatto riferimento alle due fonti di documentazione proposte da Wikimedia come "living document": https://www.mediawiki.org/wiki/Wikibase/DataModel e https://www.mediawiki.org/wiki/Wikibase/DataModel/Primer.

Item e *Property* vengono individuati dall'identificativo assegnato da Wikibase (un numero progressivo preceduto da una "Q" nel caso degli *Item*, e da una "P" nel caso delle *Property*) e da un *Fingerprint* diverso per ognuna delle lingue supportate per quella *Entity*: il *Fingerprint* è costituito dall'insieme di etichetta (*Label*), *Description* e le versioni alternative della *Label* (gli *Alias*, solitamente visualizzati con la definizione "Also known as").

Dante Alighieri (Q1067)

Italian poet — edit

Dante | Durante degli Alighieri

In more languages Configure

Language	Label	Description	Also known as
English	Dante Alighieri	Italian poet	Dante Durante degli Alighieri
Italian	Dante Alighieri	poeta italiano autore della Divina Commedia	Durante degli Alighieri Dante
French	Dante Alighieri	poète, homme politique et écrivain italien	Dante
Sardinian	Dante Alighieri	No description defined	
Abkhazian	Данте Алигьери	No description defined	
Afrikaans	Dante Alighieri	No description defined	
Amharic	ዳንቴ አሊጊዬሪ	No description defined	
Aragonese	Dante Alighieri	No description defined	
Arabic	دانتي أليغييري	No description defined	
Egyptian Arabic	دانتى اليجييرى	No description defined	
Asturian	Dante Alighieri	poeta italianu	Dante

Figura 4: *Fingerprint* dell'*Item* "Dante Alighieri" in Wikidata

Le *Entity* contengono, poi, gli *Statement* che ospitano le triple semantiche con cui si registrano i "dati strutturati" mediante *Property* e rispettivi *Target*. Il *Target* della tripla (cioè il suo oggetto) può contenere un puntatore a un'altra *Entity* oppure un valore (testuale, numerico, un link, o altro tipo di dato codificato).
Nell'esempio dell'*Item* di Dante, si può notare che è presente il riferimento al luogo in cui è morto: questo è gestito come un "dato strutturato" (una tripla semantica) che mette in relazione due entità, Dante Alighieri come soggetto e città di Ravenna come oggetto, mediante la *Property* "place of death".

place of death	Ravenna — edit 2 references
	+ add value

Figura 5: *Statement* con il riferimento al luogo della morte nell'*Item* "Dante Alighieri" in Wikidata

Gli *Statement* si basano sulle *Property*, come avviene di consueto nelle triple del web-semantico. Le *Property*, come gli *Item*, sono identificate con *Fingerprint* e "descritte" con *Statement*

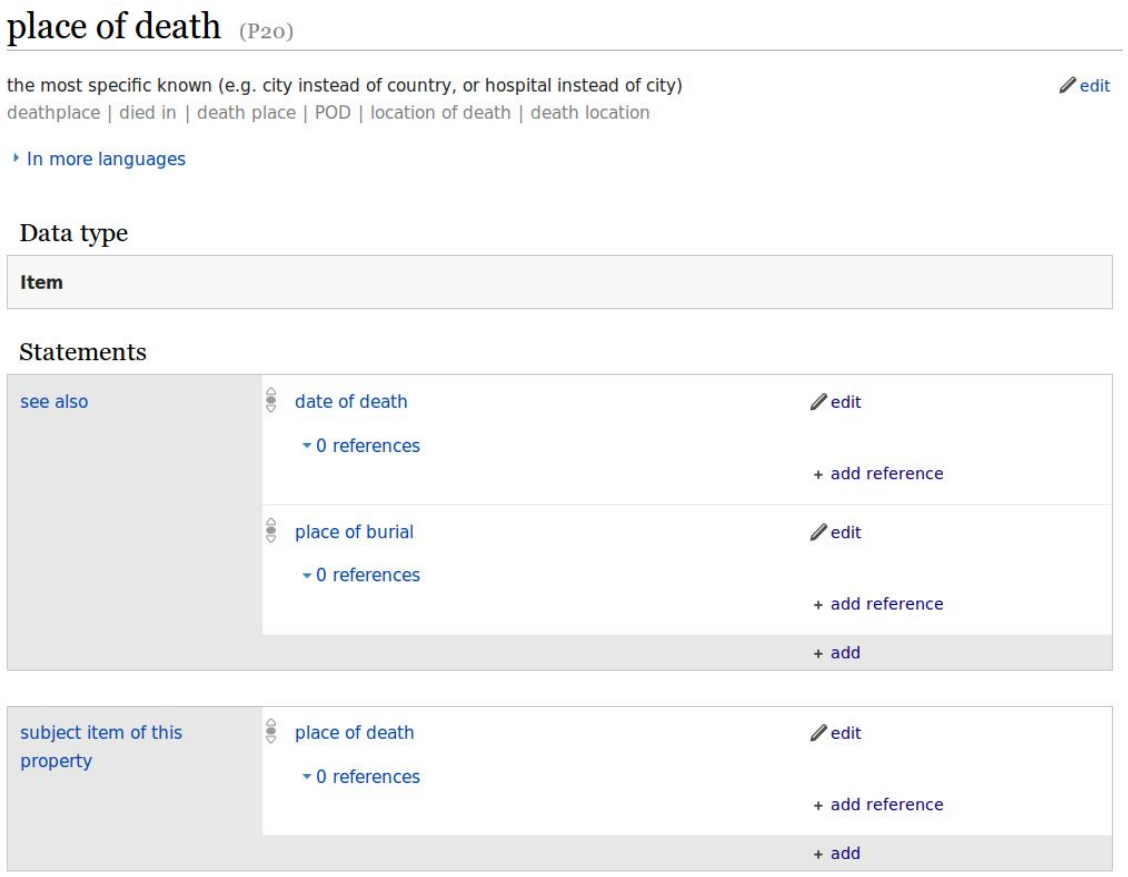

Figura 6: *Property* "Place of death" in Wikidata

La caratteristica del *Wikibase Data Model* da evidenziare ai fini dell'idea qui esposta è che, all'interno degli *Statement*, la tripla semantica di partenza (*Claim*) può essere integrata con due tipologie di informazioni: il *Qualifier* che raccoglie dettagli aggiuntivi sullo *Statement*, e la *Source* (o *Reference*) in cui si registra la fonte della singola informazione.

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

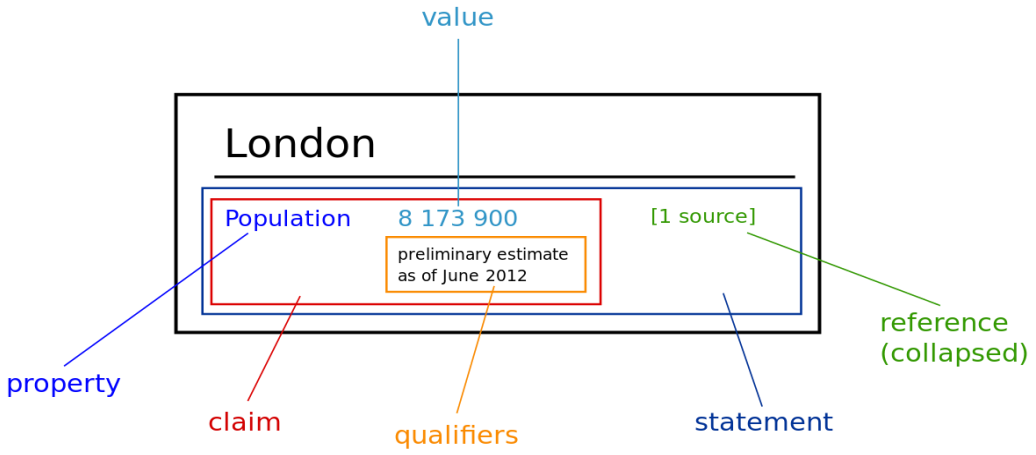

Figura 7: schema grafico del *Wikibase Data Model*

In questo senso, ancora nell'esempio dell'*Item* di Dante, si consideri che lo *Statement* della *Property* "data di nascita" contiene delle specificazioni aggiuntive - i *Qualifier* - anch'esse descritte con la tripla semantica, usando specifiche *Property*: "earliest date", limite inferiore possibile per la data (terminus post quem) e "latest date", il limite massimo possibile per la data (terminus ante quem).

Figura 8: *Statement* con il riferimento alla data di nascita nell'*Item* "Dante Alighieri" in Wikidata

E si noti pure che questo *Statement* è anche corredato con l'indicazione delle fonti - *Source* - dalle quali il dato viene acquisito (nella Figura 8 sono riportate le prime due); ciascuna fonte è a sua volta descritta con una serie di triple semantiche: la prima presenta il riferimento alla fonte (referenziata con lo specifico *Item*) e la data in cui l'informazione è stata presa, la seconda presenta anche l'URL del sito web nel quale l'informazione può essere verificata.

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Sperimentazione dell'uso del Wikibase Data Model per i dati UNIMARC

Sfruttando "l'arricchimento" della tripla semantica implementato nel *Wikibase Data Model*, si è ipotizzata la possibilità di replicare in tale modello di dati tutto il dettaglio informativo del record bibliografico UNIMARC, così come per esempio viene esposto dal Catalogo del Servizio Bibliotecario Nazionale (OPAC SBN)[12].
Si è dunque selezionato in SBN un record di media complessità (identificativo BID[13] IT\ICCU\CFI\0893220) che presenta dei rimandi ad altri record bibliografici e a punti di accesso (intestazioni) a nome, soggetto e classificazione, oltre che a voci descrittive di vocabolari controllati;

```
LEADER 05771nam0 2201813 i 4500
001 IT\ICCU\CFI\0893220
003 http://id.sbn.it/bid/CFI0893220
005 20150525130956.6
010 $a978-88-581-1170-3
020 $aIT$b2014-2498
100 $a20140514d2014 ||||0itac50 ba
101 $aita$cita
102 $aIT
181 1$6z01$ai $bxxxe
182 1$6z01$an
183 1$6z01$cnc$2rdacontent
200 1 $a<<Lo >>Stato innovatore$esfatare il mito del pubblico contro il privato$fMariana Mazzucato$gtraduzione di Fabio Galimber
210 $aRoma$aBari$cGLF editori Laterza$d2014
215 $aXXVI, 351 p.$d21 cm
410 0$1001IT\ICCU\RAV\1797816$12001 $aAnticorpi$v43
500 10$a<<The >>entrepreneurial State$3IT\ICCU\CFI\0893222$9Mazzucato, Mariana
606 $aInnovazione tecnologica$xInterventi statali$2FN $3IT\ICCU\CFIC\227454
676 $a338.45$cINDUSTRIE SECONDARIE E SERVIZI. EFFICIENZA PRODUTTIVA$v23
700 1$aMazzucato$b, Mariana$3IT\ICCU\UFIV\143190
702 1$aGalimberti$b, Fabio$f <1972- >$3IT\ICCU\CFIV\224780
801 3$aIT$bICCU$c20160705
```

Figura 9: record UNIMARC BID IT\ICCU\CFI\0893220

e si è condotta una sperimentazione di caricamento in Wikibase e quindi di trasformazione verso il *Wikibase Data Model* con procedure semiautomatiche - input e correzioni manuali coadiuvate da caricamenti con procedure batch - al fine di verificare i seguenti assunti di base dell'idea proposta:

- Le *Property* di Wikibase possono essere utilizzate per mappare tutti gli elementi strutturali del MARC: codici di campo, rispettivi indicatori e codici di sottocampo.

[12] http://opac.sbn.it.
[13] L'identificativo del record UNIMARC (etichetta 001) verrà indicato nel corso del testo con il nome "BID" adottato nel contesto SBN, mentre il sottocampo 3 del blocco delle etichette 7xx (Numero di registrazione di autorità) verrà indicato come VID.

- Gli *Item* possono accogliere i record bibliografici o di authority e le voci di liste controllate (quali le lingue o i paesi di pubblicazione), usando
    - il *Target* dello *Statement* per registrare la stringa (eventualmente completa di punteggiatura convenzionale) così come compare nei campi MARC,
    - i *Qualifier* per registrare i dettagli del MARC (come i sottocampi e gli indicatori)
    - e la *Source* per registrare, in ciascun campo, la fonte del record di origine.
- Il sistema di relazioni fra gli *Item* di Wikibase può essere utilizzato per supportare le relazioni dei record bibliografici MARC con gli altri elementi descrittivi:
    - altri record bibliografici MARC (come nel caso delle descrizioni a più livelli),
    - voci di liste controllate,
    - record di authority (come i nomi, i soggetti, le classificazioni, ecc).

UNIMARC nel Wikibase Data Model

La prima operazione necessaria nella prospettiva proposta è la "mappatura" della semantica dell'UNIMARC, che si traduce sostanzialmente nell'implementazione dell'ontologia dello standard UNIMARC all'interno del *Wikibase Data Model*: gli elementi sintattici dell'UNIMARC (label del record, indicatori, campi, sottocampi) e i dati codificati (codici di lingue, paesi, …) devono essere opportunamente distribuiti fra *Property* e *Item* di Wikibase con un dettaglio e una logica di relazioni che tengano conto da una parte dell'esigenza di preservare integralmente l'informazione originale e dall'altra delle prospettive di riuso del dato nel nuovo sistema.

Property: struttura dell'UNIMARC

Nella sperimentazione sono state create le *Property* per gli elementi strutturali dell'UNIMARC (campi, sottocampi e indicatori) presenti nei record bibliografici e di autorità selezionati; i *Fingerprint* delle *Property* create sono stati impostati in base ai seguenti criteri:

- il nome per esteso dell'elemento UNIMARC è stato inserito nella *Label* della *Property*, traducendolo in più lingue
- i codici degli elementi UNIMARC sono stati introdotti negli *Alias*, secondo uno schema che raccoglie le sigle identificative, dalla più generale a quella più specifica, separate da underscore.

Per esempio: "unimarc_bib_200_a" per il sottocampo "a" (Titolo proprio) del campo UNIMARC 200 (Titolo e formulazione responsabilità)

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Figura 10: *Property* corrispondente al sottocampo "Titolo proprio" del campo UNIMARC 200

Per mantenere l'univocità delle *Label*, è stato introdotto il suffisso [bib] nelle *Label* delle *Property* per il record UNIMARC bibliografico, e il suffisso [auth] (o [aut] per l'italiano) in quelle per il record UNIMARC delle voci di autorità.

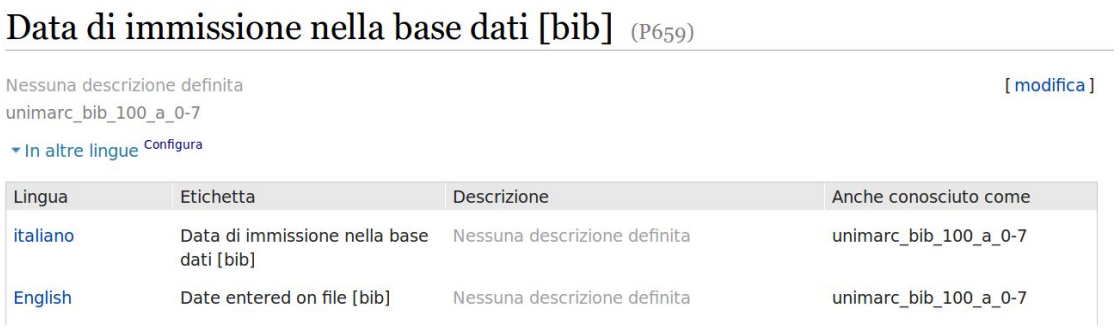

Figura 11: *Property* corrispondente alle posizioni dalla 0 alla 7 del campo UNIMARC 100, sottocampo "a"

Natura flessibile del Wikibase Data Model

La natura particolarmente flessibile del Wikibase data-model rende possibile (ma anche necessario) evitare una traduzione della sintassi UNIMARC in un elenco di tutte le combinazioni possibili delle sue parti componenti (campo, indicatori, sottocampo):[14] è stata, dunque, sfruttata la possibilità offerta da Wikibase di qualificare gli elementi con un sistema di relazioni fra le *Entity*.
Per esempio, il campo UNIMARC 200 (l'area del titolo) non è stato registrato con più *Property* per

[14] Si fa qui riferimento al lavoro di mappatura "marc21rdf" realizzato da Gordon Dunsire (http://www.marc21rdf.info/) e richiamato nell'introduzione dove, ad esempio per i campi 2xx dello standard, vengono create "*Properties representing every combination of MARC 21 tags 210 to 264, first and second indicators, and their subfield codes.*"

rendere conto delle diverse casistiche di "titolo significativo" o "titolo non significativo" date dai valori presenti nel suo primo indicatore ("Indicatore di titolo significativo"): il primo indicatore viene, invece, individuato autonomamente mediante una specifica *Property*

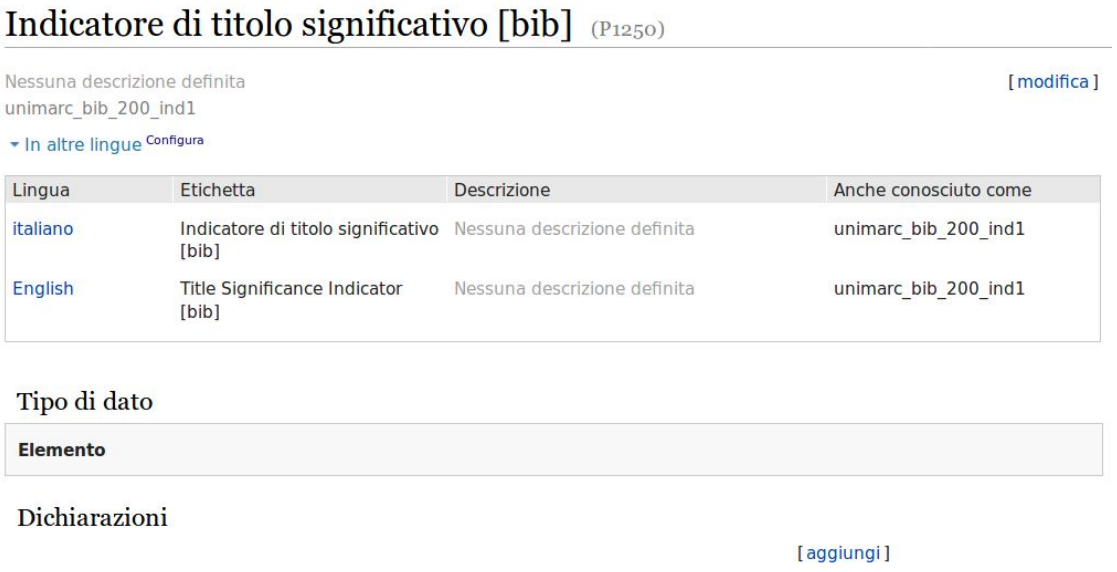

Figura 12: *Property* corrispondente al primo indicatore del campo UNIMARC 200

la *Property* del primo indicatore viene utilizzata in un *Qualifier* dello *Statement* che ospita il campo UNIMARC 200, e la sua compilazione permette di ricostruire se si tratti o meno di un titolo significativo.

Item: vocabolari controllati

Per completare la mappatura della semantica UNIMARC nel *Wikibase Data Model* si devono, dunque, importare in appositi *Item* le voci dei vocabolari controllati dello standard.
Nella sperimentazione sono stati caricati i coded-data utilizzati nei record selezionati; continuando l'esempio del campo UNIMARC 200, sono stati caricati i possibili valori utilizzabili nel primo indicatore: uno di questi è "Titolo significativo".

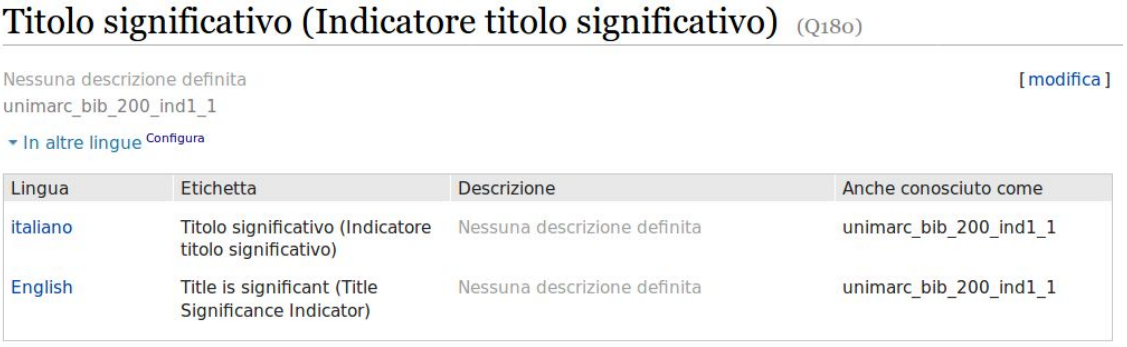

Figura 13: *Item* corrispondente al valore "Titolo significativo" per il primo indicatore del campo UNIMARC 200

Item: record dell'UNIMARC

Una volta caricate in Wikibase le *Entity* necessarie a rappresentare la struttura UNIMARC dei dati presi come esempio, si è sperimentato il caricamento dei record bibliografici e di autorità negli *Item*, utilizzando le diverse triple semantiche di cui si compongono gli *Statement* del *Wikibase Data Model*:

- i campi UNIMARC per intero (eventualmente con la relativa punteggiatura convenzionale) sono stati registrati nel *Target* dell'intero *Statement*,
- mentre gli indicatori e i sottocampi sono stati registrati nei *Qualifier*.
- Le *Source* accolgono, per ciascuna occorrenza di campo UNIMARC, il puntatore alla fonte del record originale.

Figura 14: *Statement* corrispondente al campo UNIMARC 200 del record BID IT\ICCU\CFI\0893220

Negli *Item* corrispondenti a record UNIMARC bibliografici, l'identificativo assegnato nel sistema di origine (il BID SBN) è stato registrato nell'*Alias* dell'*Item*; in questo modo, anche dopo il caricamento del record in Wikibase, può essere garantita una continuità nel modo di identificare e reperire il dato. La distribuzione di queste informazioni fra la *Label* e gli *Alias* (sia per i record bibliografici, sia per le voci di vocabolari controllati o di liste di autorità) segue le consuetudini del *Wikibase Data Model*, in cui la "main label" viene utilizzata per descrivere la *Entity* in varie lingue, mentre l'*Alias* viene principalmente utilizzato per le ricerche e per fornire punti di accesso alternativi: in questo senso, l'*Alias* risulta anche più adatto a ospitare molteplici identificativi alfanumerici nel caso sia necessario.

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Item Discussione Leggi Cronologia

Lo Stato innovatore (Q156)

Nessuna descrizione definita [modifica]
CFI0893220 | IT\ICCU\CFI\0893220

In altre lingue Configura

Lingua	Etichetta	Descrizione	Anche conosciuto come
italiano	Lo Stato innovatore	Nessuna descrizione definita	CFI0893220 IT\ICCU\CFI\0893220
English	Lo Stato innovatore	Nessuna descrizione definita	CFI0893220 IT\ICCU\CFI\0893220
français	Nessuna etichetta definita	Nessuna descrizione definita	
sardu	Nessuna etichetta definita	Nessuna descrizione definita	

Figura 15: *Fingerprint* dell'*Item* corrispondente all'intero record UNIMARC BID IT\ICCU\CFI\0893220. Nell'*Alias* è registrato il BID SBN.

Ciascuno *Statement* corrispondente a un campo UNIMARC, contiene la fonte (*Source* o *Reference*) da cui proviene il dato. Sono stati registrati, con apposite *Property*:

- l'istituzione da cui proviene il record, con un puntatore al relativo *Item* caricato in Wikibase,
- e il puntatore al record originale, con URL alla fonte

Area della pubblicazione [bib]	Roma ; Bari : GLF editori Laterza, 2014	[modifica]
Luogo di pubblicazione [bib]	Roma Bari	
Nome dell'editore [bib]	GLF editori Laterza	
Data di edizione [bib]	2014	
MARC ordine sottocampi	a[1] a[2] c[1] d[1]	
1 riferimento		
		[modifica]
Importato da	National Central Library of Florence English	
Sito web ufficiale	http://purl.org/bncf/marc/CFI0893220	
		[aggiungi riferimento]
		[aggiungi]

Figura 16: *Statement* corrispondente al campo UNIMARC 210 del record BID IT\ICCU\CFI\0893220

L'ordine originale dei sottocampi è registrato in un apposito Qualifier

Lo standard UNIMARC prevede, fra le sue caratteristiche fondamentali, che i campi siano eventualmente ripetibili e che anche i sottocampi possano essere eventualmente ripetuti all'interno di ciascuna occorrenza di un campo. Meno scontata è la situazione in cui i sottocampi vengano ripetuti frammisti ad altri sottocampi e meno scontata ancora è la situazione (peraltro mai esplicitamente prevista o vietata dallo standard) in cui un sottocampo sia ripetuto contenendo il medesimo valore.
Un esempio[15] che presenti entrambe le problematiche può essere individuato nell'area del titolo (campo UNIMARC 200) del seguente record (BID IT\ICCU\URB\0620565), dove si può notare la presenza del sottocampo $f ripetuto due volte con lo stesso valore e in diverse posizioni:

```
200 1 $aPour les valeurs bourgeoises$fpar Georges Hourdin$cContre les
valeurs bourgeoises$fpar Georges Hourdin
```

Entrambe le necessità - quella di ripetere i sottocampi (*Qualifier*) in diverse posizioni e quella di registrare sottocampi con contenuti uguali (quindi *Qualifier* con stessa *Property* e stesso *Target* nel medesimo *Statement*) - non sarebbero nativamente supportate nel *Wikibase Data Model*; il software Wikibase non accetta due *Claim* con identici *Property* e *Target* e, infatti, risponde con un errore al tentativo di caricamento sia attraverso l'interfaccia di input manuale del software, sia se si prova a inviare i dati attraverso API utilizzando una procedura automatica (si veda di seguito):

```
pywikibot.data.api.APIError: modification-failed: Claim has already a
qualifier with hash d4d0bad3dc0bbd58a8c3c218fd135e73265a03e9
```

Mentre, riguardo la ripetizione dei sottocampi in diverse posizioni, l'ordine verrebbe perso nella loro rappresentazione mediante *Qualifier*, poiché i *Claim* dei *Qualifier* (all'interno dello stesso *Statement*) sono raggruppati in base alle *Property* e dunque più sottocampi UNIMARC $f presenti in diverse posizioni del campo (come nel caso descritto) verrebbero raccolti in un'unica lista riferita alla *Property* con cui si è mappato il sottocampo $f.

Nella sperimentazione si è data soluzione a queste due esigenze introducendo un *Qualifier* aggiuntivo - il *Qualifier* con *Property* "MARC subfields order" - per la registrazione dell'esatto ordine dei sottocampi (di ciascuna occorrenza dei campi UNIMARC e quindi degli *Statement*).

Nel caso del record (BID IT\ICCU\URB\0620565) portato sopra ad esempio, lo *Statement* corrispondente all'area del titolo (campo UNIMARC 200) conterrà un *Qualifier* con *Property* "MARC subfields order" e *Target* con stringa "a[1] f[1] c[1] f[1]": questi sono i codici di sottocampo, nella loro esatta sequenza (registrata a prescindere da come poi Wikibase archivia il contenuto di ciascun sottocampo); il numero fra parentesi quadre di seguito ai codici è il contatore della ripetizione di ciascuno di essi.

[15] Con tutta probabilità (ma non abbiamo visto il libro) si tratta di un errore di trascrizione. Una versione corretta del record bibliografico si può trovare qui: http://catalogue.bnf.fr/ark:/12148/cb33045411k.

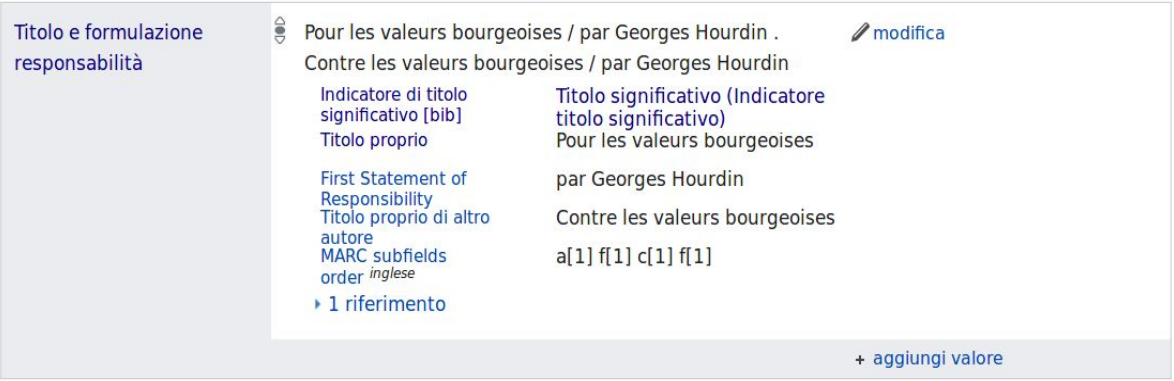

Figura 17: *Statement* corrispondente al campo UNIMARC 200 del record BID IT\ICCU\URB\0620565

In questo modo, l'ordine dei sottocampi viene correttamente registrato, mentre la seconda occorrenza di sottocampo (ripetuta identica) non viene archiviata, viene solo referenziata.
Va da sé che con questa soluzione tutti i meccanismi di indicizzazione e ricerca di Wikibase vengono conservati: la stringa "par Georges Hourdin" viene correttamente collegata alla specifica occorrenza dello *Statement* con cui si registra l'area del titolo (UNIMARC 200) nell'*Item* creato per il record BID IT\ICCU\URB\0620565; mentre in fase di estrazione dati, sarà sempre possibile ricostruire l'esatta sequenza dei sottocampi, comprese le loro ripetizioni.

Impostazione relazionale

Nella sperimentazione si è necessariamente affrontata l'impostazione "relazionale" del *Wikibase Data Model* verso il quale è stata trasformata la descrizione UNIMARC: il record UNIMARC è notoriamente un "record piatto" nel quale tutte le informazioni comuni vengono ripetute; ma è altrettanto assodato che, con il suo supporto di identificativi univoci e puntatori fra notizie messe in relazione (come nel caso dei punti di accesso o delle descrizioni a più livelli gestite nel Linking Entry Block dello standard), l'UNIMARC veicola con coerenza l'impostazione relazionale degli elementi che "organizzano" l'universo bibliografico.[16]
Il processo di trasformazione dall'UNIMARC al *Wikibase Data Model*, pertanto, ha previsto la scomposizione di tutti gli elementi UNIMARC riconducibili ad *Entity* Wikibase univoche, registrando e qualificando tutte le necessarie relazioni fra di esse.

In un *Item* corrispondente a un record UNIMARC bibliografico, dunque, i campi UNIMARC contenenti voci di autorità sono gestiti con puntatori ad altri *Item*: il punto di accesso al nome, ad esempio, viene registrato in uno *Statement* con la *Property* che mappa lo specifico tipo di responsabilità, e il contenuto - il *Target* - viene sostituito da un puntatore all'*Item* della voce di authority del nome (si noti che Wikibase espone automaticamente la *Label* dell'*Item* collegato)

[16] "Data elements on bibliographic records might be classified into two categories: those that describe the entity in hand and those that relate the entities to other entities. Thus, in considering data elements to be included on bibliographic records, account needs to be taken not only of those that represent the format attributes of the entities described (descriptive elements) but also those whose purpose is to organize catalogues and by so doing to structure the bibliographic universe (organizing elements)" (Svenonius 1992).

Figura 18: 2 *Statement* corrispondenti ai punti di accesso al nome del record BID IT\ICCU\CFI\0893220

Aggiuntivamente, nei *Qualifier* degli *Statement* utilizzati per le "relazioni" fra *Item*, sono stati registrati dei dati specifici (quando presenti) come il Relator-code del punto di accesso al nome oppure l'indicazione del volume nel puntatore alla collezione (dal Linking Entry Block dell'UNIMARC).

Figura 19: *Statement* corrispondente al puntatore alla Serie, campo UNIMARC 410 del record BID IT\ICCU\CFI\0893220

Nella sperimentazione, è stata ipotizzata una distinzione fra (a) puntatori a record di liste di autorità (che possono avere nel formato UNIMARC un record autonomo), e (b) puntatori a voci di liste chiuse di coded-data:

a) Gli *Statement* che puntano a voci di autorità provenienti da apposito record UNIMARC (nomi, soggetti, classificazioni, titoli assegnati), contengono il puntatore nel *Target* dello *Statement*.
Nel record bibliografico, vengono scartati i dati ridondanti con la voce di autorità mentre vengono conservati eventuali "attributi di relazione" (si vedano le Figure 18 e 19).

b) Gli *Statement* che contengono coded-data riferiti a "liste chiuse" (vocabolari controllati), presentano il puntatore nei *Target* dei *Qualifier*

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Dati per la elaborazione [bib]	20140514d2014 ||||0itac50 ba	[modifica]
Dati generali (informazioni codificate) [bib]	20140514d2014 ||||0itac50 ba	
MARC ordine sottocampi	a[1]	
Data di immissione nella base dati [bib]	20140514	
Tipo data di pubblicazione [bib]	Monografia completa quando pubblicata o pubblicata in parti ma nello stesso anno solare	
Data pubblicazione 1 [bib]	2014	
Codice record modificato [bib]	record non modificato	
Lingua della catalogazione [bib]	Lingua italiana	
Codice di traslitterazione [bib]	traslitterazioni multiple: usati schemi ISO o altri	
Character set [bib]	Set di caratteri: ISO 10646 Level 3 (Unicode, UTF-8)	
Alfabeto del titolo [bib]	Alfabeto: latino	

Figura 20: *Statement* corrispondente al campo UNIMARC 100 del record BID IT\ICCU\CFI\0893220

Nei casi che lo richiedono, come i campi "Language of the Item" (UNIMARC 101) e "Country of Publication" (UNIMARC 102), i puntatori a livello di sottocampo permettono di gestire anche i valori multipli (cioè diverse lingue) nella medesima occorrenza.

La voce della lista di autorità, ad esempio quella di un autore, diventa un *Item* autonomo, completo di *Statement* e specifiche *Property* tanto quanto l'*Item* del record bibliografico, e con l'identificativo della voce di autorità (in questo caso il VID SBN) registrato nel segmento *Alias* del *Fingerprint* (sia per la versione italiana sia per quella inglese).

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

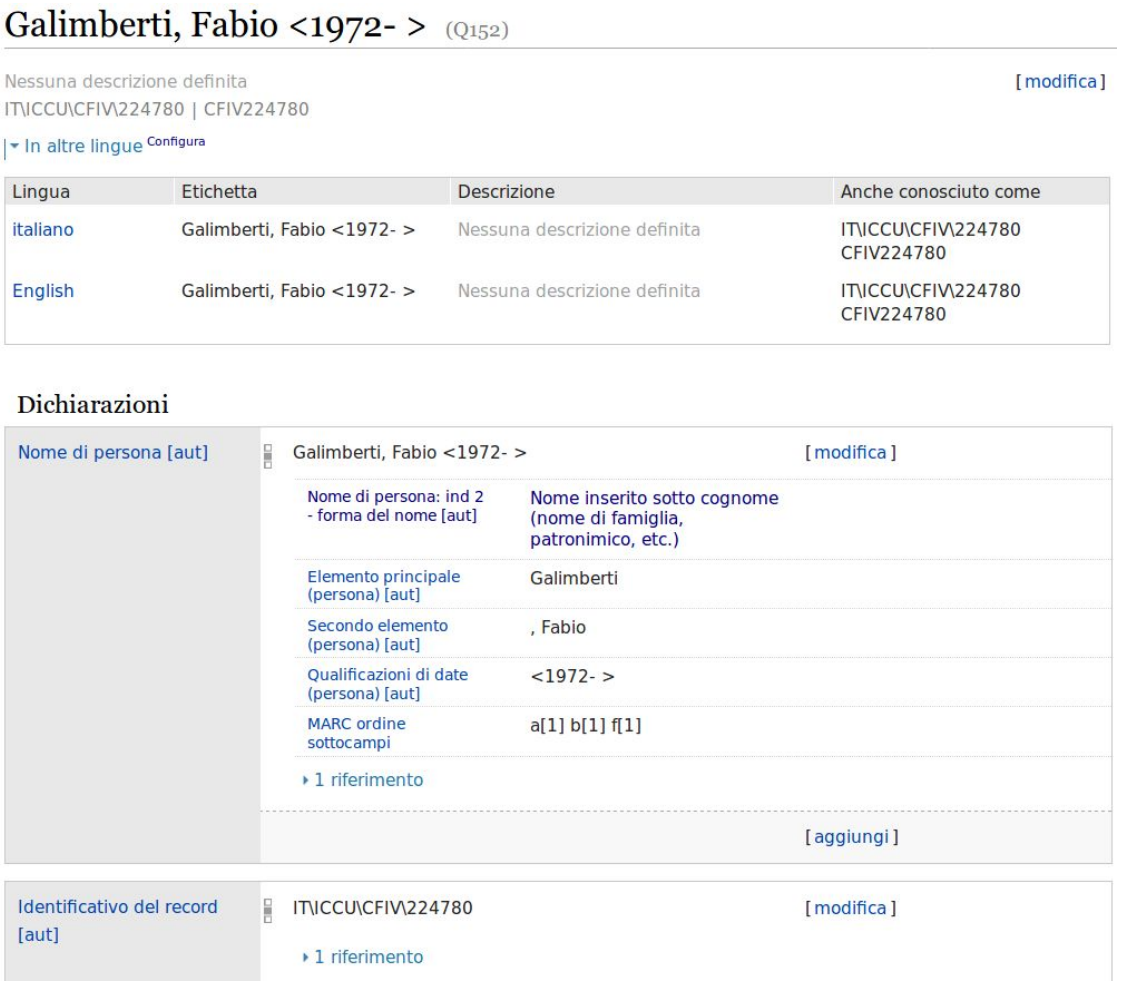

Figura 21: *Item* corrispondente al record di autorità per l'autore Galimberti, Fabio

Ristrutturazione del record UNIMARC in Item distinti per tipologia di dato

In questo modo con la sperimentazione si è potuta realizzare una ristrutturazione del record UNIMARC in *Item* distinti per tipologie di dato (potenzialmente riferibili anche alle diverse entità FRBR) e al tempo stesso si è testata la possibilità di realizzare una mappatura di tutti gli elementi dell'UNIMARC, mantenendo la possibilità di recuperare tutte le informazioni del formato originale.

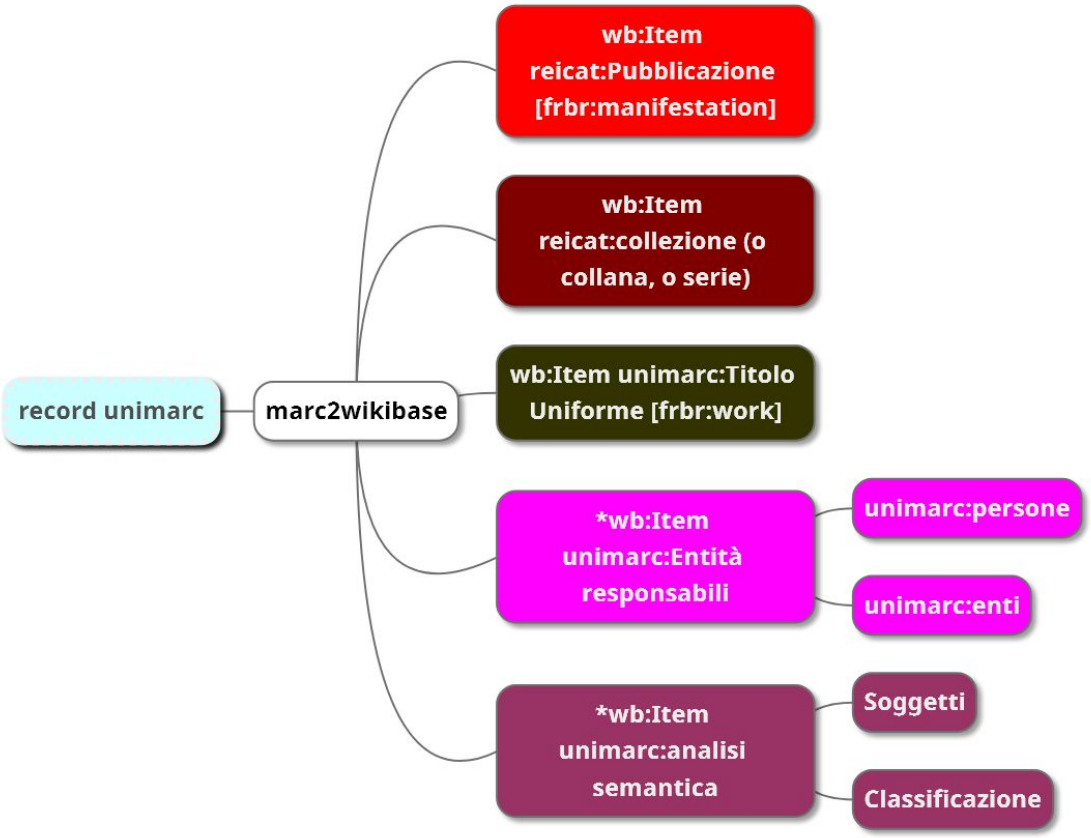

Figura 22: schema grafico della ristrutturazione del record UNIMARC in *Item* distinti per tipologie di dato

Dati quantitativi della sperimentazione

I dati quantitativi della sperimentazione permettono di rendere un'idea dell'articolazione in diverse *Entity* necessaria a rappresentare un record UNIMARC e gli elementi che esso richiama all'interno del *Wikibase Data Model*. Un record UNIMARC di media complessità - "Lo Stato innovatore" (BID IT\ICCU\CFI\0893220) - in Wikibase ha dato luogo a un *Item* collegato a più di 160 altre *Entity* (fra *Item* e *Property*): oltre 100 *Property* per gli elementi UNIMARC mappati; circa 40 *Item* per i dati codificati (quali i codici di relazione, di lingua o dei paesi); e poi altri *Item* per i record bibliografici collegati (come la serie) o per le voci di autorità (come i punti di accesso ai nomi, al soggetto, alla classificazione o al titolo uniforme).[17]

[17] Le *Entity* sono state create in una installazione temporanea di Wikibase (allestita solo ai fini di questa sperimentazione). La schermata dell'intero *Item* corrispondente al record UNIMARC principale (così come viene visualizzato nell'interfaccia grafica di Wikibase) si può trovare all'indirizzo https://archive.org/details/Q156It; mentre un listato di tutte le *Entity* (codificate nel formato JSON prodotto da Wikibase) create per rendere la struttura UNIMARC del record principale e i dati ad esso collegati, si può trovare all'indirizzo https://archive.org/details/list_of_used_Properties_and_Items.

In dettaglio: il record di esempio ha il proprio *Item*

- Record bibliografico: Lo Stato innovatore (si veda la Figura 15)

Tutte le *Entity* collegate al record sono destinate a essere riutilizzate in connessione con altre *Entity*.
Item per i record collegati e per le voci di autorità:

- Record bibliografico (serie): Anticorpi (si veda la Figura 19)
- Responsabilità (voce di autorità): Galimberti, Fabio <1972- > (si veda la Figura 21)
- Responsabilità (voce di autorità): Mazzucato, Mariana (si veda la Figura 23)
- Voce di classificazione: 338.45 (ed. 23) - INDUSTRIE SECONDARIE E SERVIZI. EFFICIENZA PRODUTTIVA
- Voce di soggetto: Innovazione tecnologica - Interventi statali (si veda la Figura 24)
- Titolo uniforme: The entrepreneurial State

Item e *Property* per la mappatura della sintassi UNIMARC:

- Circa 40 *Item*, utilizzati per archiviare i coded-data (si veda la Figura 13)
- 116 *Property* per gli elementi strutturali dell'UNIMARC (si vedano le Figure 10, 11 e 12)

Questioni aperte individuate con la sperimentazione

La sperimentazione ha permesso anche di individuare alcune questioni utili da considerare nella prospettiva del caricamento di grandi quantità di dati, con processi completamente (o quasi completamente) automatizzati.

Acquisizione di liste di autorità e vocabolari controllati

Nella sperimentazione, l'acquisizione delle voci di liste di autorità e di vocabolari controllati (oltre ai record bibliografici) è stata operata testando differenti fonti: l'OPAC della Biblioteca Nazionale Centrale di Firenze per i record bibliografici, e l'OPAC di Indice SBN per i record di autorità dei nomi.
Per le voci di autorità dei nomi, in particolare, oltre al caso di nomi provenienti da authority record UNIMARC (come la voce "Galimberti, Fabio <1972- >" già illustrata nella Figura 21), nella sperimentazione si è testata la situazione di nomi per i quali non esiste un record di autorità: in questo caso l'*Item* corrispondente al nome è stato generato in Wikibase con i soli dati desunti dal punto di accesso 7xx del record-bibliografico.
È il caso del nome Mazzucato, Mariana

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Figura 23: *Item* creato per l'autrice Mazzucato, Mariana

Analogamente, si è sperimentata la generazione di *Item* per voci di autorità quali il titolo uniforme ("The entrepreneurial State"), il soggetto ("Innovazione tecnologica - Interventi statali") e la classificazione ("338.45 (ed. 23) - INDUSTRIE SECONDARIE E SERVIZI. EFFICIENZA PRODUTTIVA") a partire dai dati registrati nel record bibliografico che li richiama.

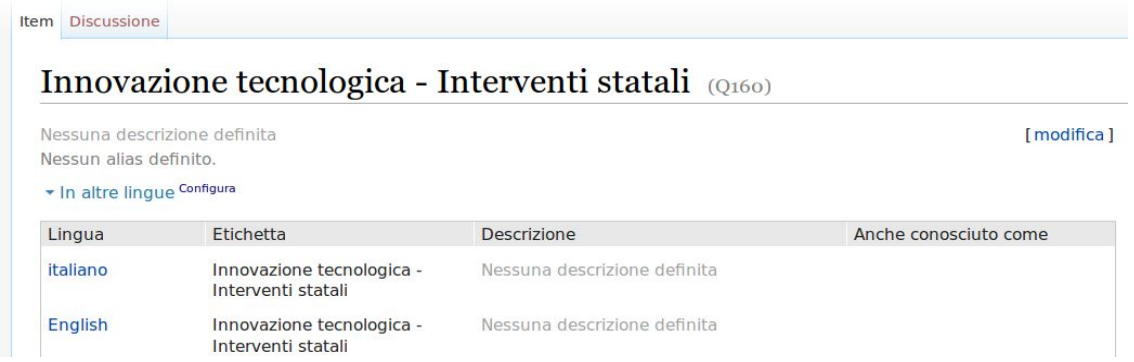

Figura 24: *Item* creato per la voce di soggetto

Ma certamente, nella prospettiva del caricamento di grandi quantità di dati, dovrebbe essere pianificata preliminarmente un'acquisizione di tutte le voci di vocabolari controllati e di liste di autorità.

Trasformazione del formato bibliografico verso il Wikibase Data Model

Nella sperimentazione sono state adottate alcune scelte rispetto al *Target* testuale degli *Statement* corrispondenti a campi UNIMARC:

- quando il campo trattato corrisponde a un'area ISBD, nel *Target* è stata ricostruita la punteggiatura ISBD.

Figura 25: *Statement* corrispondente al campo UNIMARC 210 del record BID IT\ICCU\CFI\0893220; il *Target* contiene la punteggiatura originale

- quando il campo trattato non corrisponde a un'area ISBD con relativa punteggiatura prescritta, nel *Target* è stato semplicemente inserito l'elenco di tutti i contenuti dei sottocampi (nel loro ordine originale) separati da spazio trattino spazio (" - ").

Figura 26: *Statement* corrispondente al campo UNIMARC 181 del record BID IT\ICCU\CFI\0893220

La *Label* dell'intero *Item* corrispondente al record bibliografico, contiene solo il titolo proprio

Figura 27: *Fingerprint* dell'*Item* corrispondente all'intero record UNIMARC BID IT\ICCU\CFI\0893220.
La *Label* contiene il titolo proprio del record UNIMARC.

In ogni caso è ovvio che nel *Fingerprint* dell'*Item* dovranno, con ogni probabilità, confluire alcuni elementi o alcune interpolazioni di elementi che il catalogo di biblioteca tradizionalmente usa per identificare il record bibliografico (area del titolo, identificatori, etc); e, analogamente, i *Target* degli *Statement* corrispondenti ad aree ISBD, potranno ospitare la sequenza dei sottocampi completa della punteggiatura prevista. A tale scopo la pianificazione di un caricamento di grandi quantità di dati dovrebbe prevedere anche un processo di ri-generazione della punteggiatura convenzionale usata nelle visualizzazioni del record bibliografico, supportando dunque tutto il complesso sistema di casi di punteggiatura in base alle combinazioni di campi e sottocampi.

Inoltre, sempre nella prospettiva di gestire numerosi record UNIMARC come *Item* in Wikibase, dovrebbe essere attentamente valutata la quantità di dati del record da riportare nel *Fingerprint* dell'intero *Item* (*Label*, *Description*, *Alias*), tenendo presente che il solo titolo proprio non è sufficiente né a identificare una manifestazione (o una pubblicazione) né a garantire che si evitino, all'interno della medesima istanza di Wikibase, casi di *Label* uguali in differenti *Item*.

Data-type di Wikibase

La sperimentazione ha evidenziato la possibilità di convertire alcuni dati UNIMARC verso specifici data-type del *Wikibase Data Model*: ad esempio le date registrate nei dati codificati dell'UNIMARC possono essere normalizzate verso il data-type "Point in time".

Per la gestione delle stringhe di codici con valore posizionale, possono essere studiati specifici metodi di archiviazione: nella sperimentazione, ad esempio, la label del record UNIMARC (oltre ad essere tradotta in tutte le sue parti componenti mediante specifici *Qualifier*) viene archiviata nel *Target* dello *Statement* fra doppie virgolette, per conservare gli eventuali spazi alla fine della sequenza di caratteri

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Guida del record [bib]	"01429nam0 2200313 i 450 " [modifica]	
	Stato del record [bib]	Stato del record: nuovo record
	Tipo record [bib]	Materiale a stampa
	Livello bibliografico [bib]	Monografia
	Livello gerarchico [bib]	Livello gerarchico: senza relazione gerarchica
	Livello di codifica [bib]	Livello di codifica: completo
	Tipo catalogazione descrittiva [bib]	Record parzialmente conforme a ISBD

Figura 28: *Statement* corrispondente alla "record label" del record UNIMARC BID IT\ICCU\CFI\0893220

Implementazione di procedure automatiche

Nella sperimentazione, il caricamento dei dati è stato operato coadiuvando l'input manuale delle diverse soluzioni in fase di test con l'implementazione di molteplici procedure di upload automatico - chiamate *bot* (diminutivo di robot) - mirate, in particolare, a caricare porzioni di dati uniformi fra loro in una installazione temporanea di Wikibase. Nello specifico, per la scrittura delle procedure è stato utilizzato il linguaggio di programmazione Python, avvalendosi della libreria Pywikibot[18] che supporta il dialogo con le API (Application Programming Interface) del software Wikibase,[19] a loro volta un'estensione delle MediaWiki API.[20]

La sperimentazione dunque, sebbene fosse concentrata sullo studio del modello di dati del sistema e non sullo sviluppo di un software di caricamento, ha permesso di verificare che per l'import dei dati in Wikibase esiste una grande varietà di tool open source mantenuti dalla comunità di sviluppo che circonda i sistemi MediaWiki. In tale ambito, fatte salve le verifiche sulle performance da vagliare su grandi quantità di dati, si possono certamente trovare gli strumenti per affrontare le questioni più consuete nei processi di trasferimento dati da un sistema a un altro.

A titolo d'esempio: la replica dell'impostazione relazionale dei dati UNIMARC all'interno del *Wikibase Data Model* richiede, ovviamente, la creazione di nuovi puntatori fra gli elementi descrittivi (ad esempio un record bibliografico che punta a un record di autorità dell'autore) in cui si sostituiscono i "vecchi" identificativi dei dati originali (il BID del record bibliografico e il VID della voce di autorità del nome) con i "nuovi" identificativi assegnati alle *Entity* da Wikibase; ma i "nuovi" identificativi vengono creati solo dopo il caricamento di ciascuna *Entity* e non possono essere previsti in anticipo. Questa questione (risolvibile con le consuete elaborazioni intermedie dei dati da caricare) potrà certamente essere affrontata testando con grandi quantità di dati le diverse opzioni già offerte dalle librerie e dai comandi API: nel caso di un caricamento di dati pre-elaborati organizzati in lotti pre-ordinati, si potrà sfruttare l'opzione dell'upload di record interi (codificati in JSON); nel caso di un aggiornamento dei puntatori fra le *Entity* successivo al loro caricamento in Wikibase, si potrà sfruttare una grande varietà di opzioni di modifica via API

[18] Il sito di riferimento della libreria è https://www.mediawiki.org/wiki/Manual:Pywikibot, mentre la pagina di valutazione in Mediawiki è https://www.mediawiki.org/wiki/API:Client_code/Evaluations/Pywikibot.

[19] Le Wikibase API sono documentate in https://www.mediawiki.org/wiki/Wikibase/API.

[20] Le Mediawiki API sono documentate in https://www.mediawiki.org/wiki/API:Main_page.

dei dati già caricati nel sistema.[21] In ogni caso, l'elaborazione dei puntatori (così come il trattamento di ogni elemento delle *Entity* che si vanno a creare in Wikibase) può sfruttare una completa serie di messaggi di risposta restituiti dalle API a ogni intervento sui dati: per ciascun *Item* o *Property* creato o modificato, viene restituito l'identificativo della *Entity* e il dettaglio completo dei dati modificati; in questo modo diventa possibile tenere traccia dei nuovi identificativi che vanno a sostituire i BID (o i VID) dei record e pianificare il più efficace metodo di aggiornamento dei puntatori fra le *Entity* create.

Nuove mappature della semantica UNIMARC

Con la sperimentazione si è verificato che nel *Wikibase Data Model* le relazioni fra *Entity* sono gestite con un grado di flessibilità tale da poter immaginare che una "mappatura" della semantica UNIMARC realizzata in Wikibase possa essere poi messa in relazione con molteplici altri schemi di metadati (o ontologie) attraverso riferimenti aggiunti negli *Alias* delle *Property* create per rappresentare l'UNIMARC stesso, oppure attraverso relazioni (di equivalenza, di inclusione o altro) con le *Property* create per il nuovo schema (lo stesso può essere fatto per gli *Item* utilizzati per le voci dei vocabolari controllati).
Per il collegamento fra *Property* (e fra *Entity* in generale) sono già disponibili in Wikidata diverse "relazioni" come "proprietà equivalente" (*Property* P1628)[22] o "classe equivalente" (*Property* P1709) [23] che possono essere utilizzate per "mappare" le *Entity* di uno schema con quelle di un altro.
Un esempio di "mapping di proprietà" può essere trovato in Wikidata nella *Property* "titolo" (P1476),[24] che viene collegata, mediante la relazione "proprietà equivalente" (*Property* P1628), sia al campo "title" del Dublin-Core[25] sia al campo "name" di Schema.org.[26]

[21] Si vedano i comandi per la modifica delle *Entity* esemplificati nella documentazione: https://www.wikidata.org/wiki/Wikidata:Creating_a_bot#API.
[22] https://www.wikidata.org/wiki/Property:P1628
[23] https://www.wikidata.org/wiki/Property:P1709
[24] https://www.wikidata.org/wiki/Property:P1476
[25] http://purl.org/dc/terms/title
[26] https://schema.org/name. Per la documentazione sullo schema: https://schema.org.

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

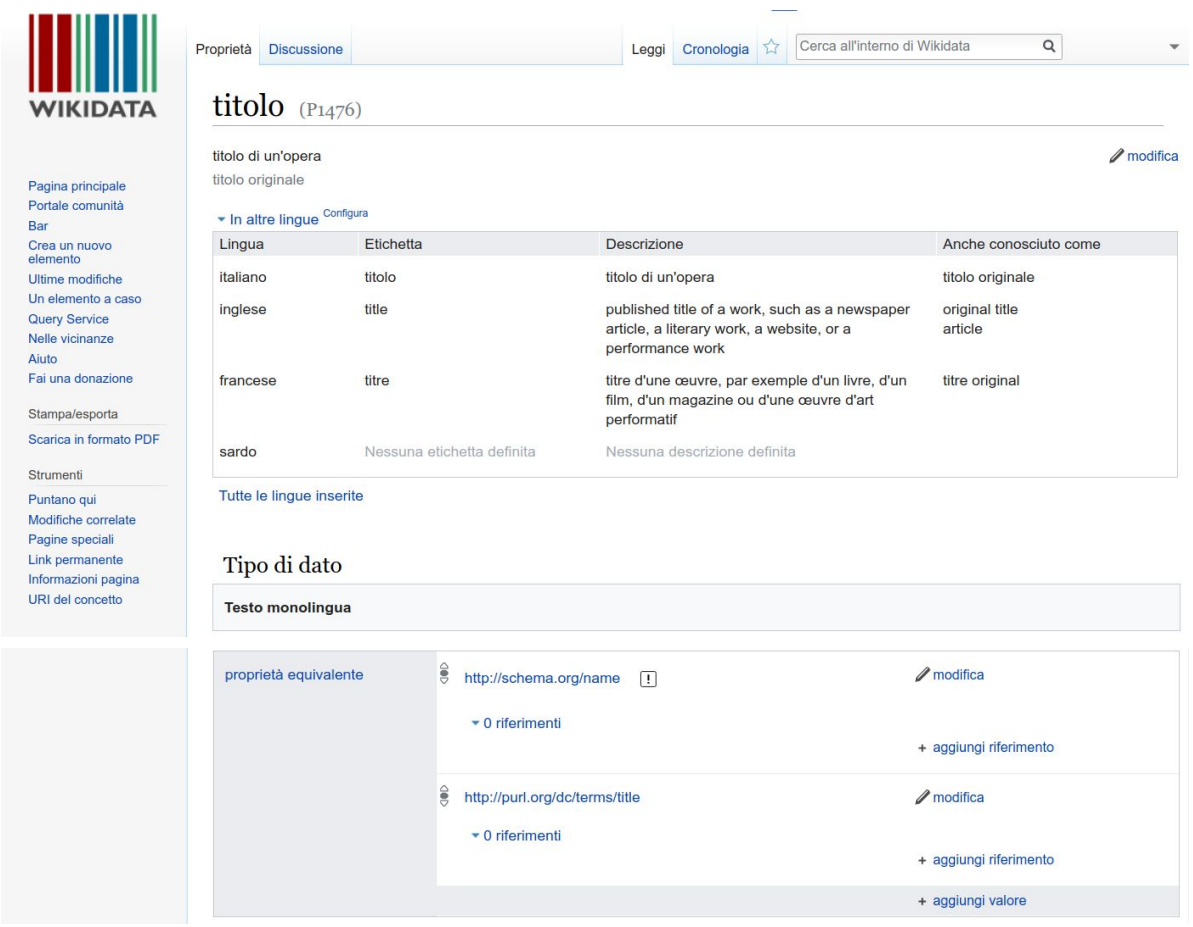

Figura 29: *Property* "Titolo" (P1476), collegata al campo "title" del Dublin-Core mediante la *Property* "proprietà equivalente" (P1628)

Un esempio di "mapping di classi" può essere trovato nell'*Item* "creative work" di Wikidata (Q17537576),[27] che viene collegato, mediante la relazione "classe equivalente" (P1709), sia alla classe "CreativeWork" di Schema.org[28] sia alla classe "work" della ontologia GND.[29]

[27] https://www.wikidata.org/wiki/Q17537576
[28] http://schema.org/CreativeWork
[29] https://d-nb.info/standards/elementset/gnd#Work

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

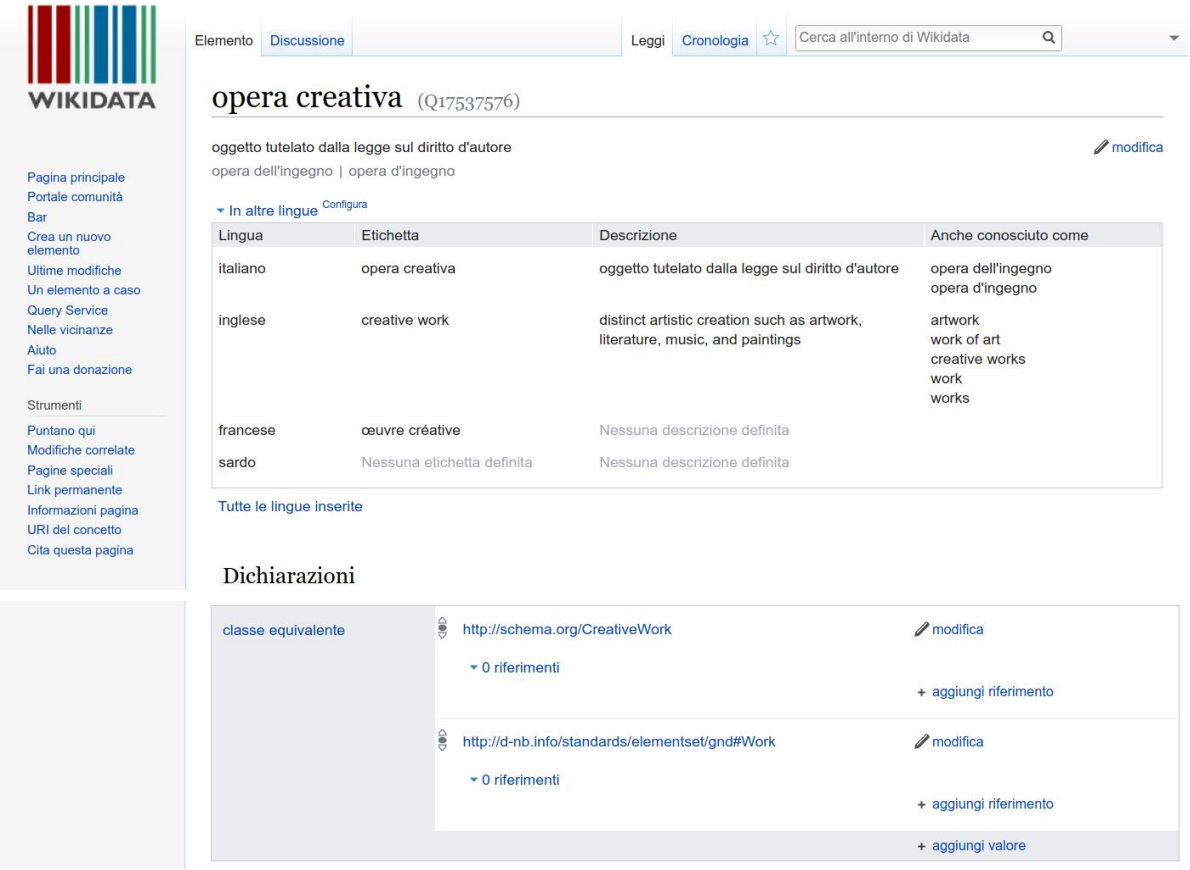

Figura 30: *Item* "Creative work" (Q17537576), collegato alla classe "CreativeWork" di Schema.org e alla classe "work" della ontologia GND mediante la relazione "classe equivalente" (P1709).

Relazioni fra *Entity* come quelle descritte possono poi essere sfruttate in fase di ricerca per utilizzare indifferentemente qualunque schema mappato: rimanendo agli esempi citati, con una semplice query nell'endpoint SPARQL di Wikidata[30] si può recuperare la *Property* "titolo" a partire dal campo "title" del Dublin-Core[31] o l'*Item* "creative work" sia a partire dalla classe "CreativeWork" di Schema.org[32] sia a partire dalla classe "work" della ontologia GND.[33]
Inoltre, nei contenuti di Wikidata, e quindi anche nei meccanismi di Wikibase, è prevista la gestione delle relazioni gerarchiche fra *Property*: fra gli esempi di query SPARQL nella documentazione[34] si trova l'interrogazione che raccoglie tutte le "Subproperties" della *Property* "location" (P276) anche a molteplici gradi di separazione;[35] questo tipo di relazione gerarchica può essere utilizzato, ad esempio, per collegare i campi MARC ai relativi sottocampi o ai singoli caratteri dei campi a lunghezza fissa, e anche per ricondurre (nel caso delle mappature) più campi MARC a medesimi concetti più generici (come "title" o "creator").

[30] Il "Query Service" di Wikidata è https://query.wikidata.org/, e la pagina iniziale della documentazione è https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/Wikidata_Query_Help.
[31] Si veda la query SPARQL http://tinyurl.com/yca73bgo.
[32] Si veda la query SPARQL http://tinyurl.com/ycxnecuo.
[33] Si veda la query SPARQL http://tinyurl.com/y83vu8v4.
[34] Si veda in particolare, ai fini del ragionamento qui esposto, la sezione "Wikibase predicates" https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples#Wikibase_predicates.
[35] Si veda la query SPARQL http://tinyurl.com/ydh877z5.

Tutti questi meccanismi di rimando fra le *Entity* possono essere inseriti in interrogazioni più complesse per ricercare i dati in base ai diversi schemi mappati: un esempio immediato può essere la ricerca di tutti gli *Item* di Wikidata che presentano la stringa "Alice's Adventures in Wonderland" come "title" Dublin-Core[36]

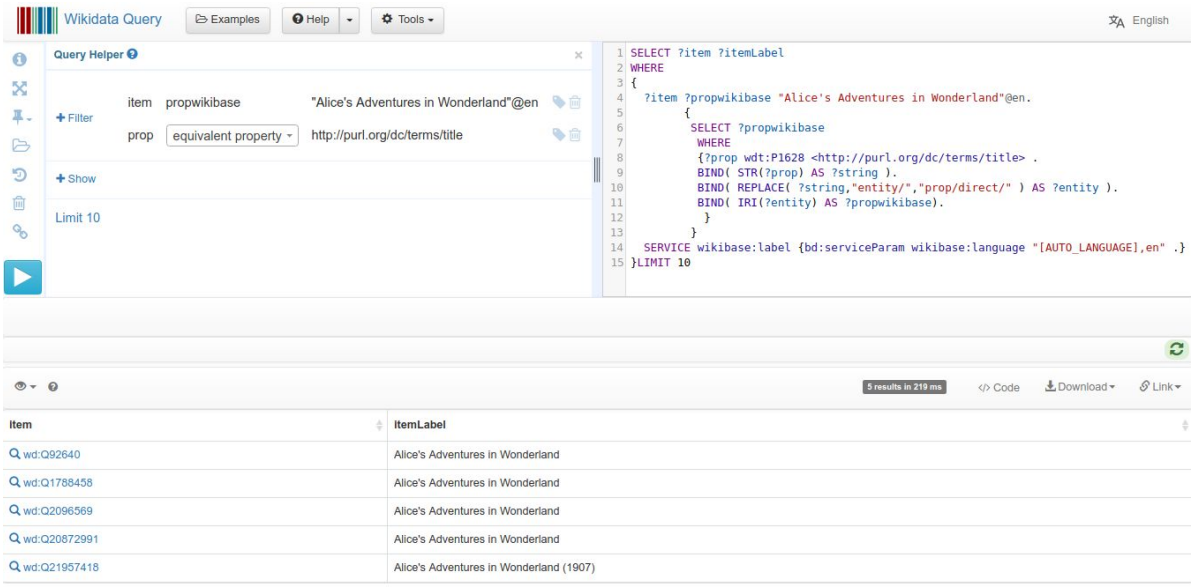

Figura 31: Query SPARQL che cerca, in base al "title" Dublin-Core, gli *Item* di Wikidata contenenti la stringa "Alice's Adventures in Wonderland".

In pratica lo scenario che può essere prefigurato è quello di grandi quantità di dati bibliografici presenti in Wikibase (con il dettaglio informativo verificato in questa sperimentazione) che possono essere visti e interrogati oltre che con la semantica UNIMARC anche ad esempio con la semantica *schema.org* o *BIBFRAME*.

Principali potenzialità tecniche di Wikibase, per l'informazione bibliografica

Una volta tradotto tutto il dettaglio informativo del MARC nel *Wikibase Data Model*, diventa possibile sfruttare per il dato bibliografico le soluzioni tecniche e i servizi implementati per Wikibase.

[36] Si veda la query SPARQL http://tinyurl.com/y79l2sjt.

La conversione del record UNIMARC nel Wikibase Data Model può facilitare l'uso dei dati

Il grado di reale "usabilità" dei dati strutturati in MARC, specie in quest'epoca di grandi innovazioni nei formati di archiviazione e nei protocolli di interrogazione e scambio, è un tema ampiamente dibattuto nella letteratura di settore e costituisce spesso un argomento di critica verso lo standard.

Il caricamento di tutto il dettaglio MARC nel *Wikibase Data Model* (a partire da uno dei suoi dialetti, come l'UNIMARC) può probabilmente rispondere a un buon numero di necessità che si evidenziano nella reale pratica d'uso dello standard. Un rapido elenco delle più significative necessità per le quali è utile cercare risposta, si può certamente trarre dalle autorevoli analisi già proposte per il problema, come l'articolo "MARC21 as Data: A Start" in cui Karen Coyle evidenzia la necessità di superare i limiti del formato MARC nella prospettiva di fornire nuovi servizi bibliografici nel contesto del web, sottolineando però che è necessario prima avere una cognizione completa di tutto il dettaglio informativo che il MARC veicola. Nella sua analisi, Karen Coyle elenca una serie di limiti tecnici del formato MARC che ostacolano la decodifica completa del suo contenuto. Questi limiti tecnici possono forse trovare una soluzione con la traduzione nel *Wikibase Data Model*: di seguito si illustrano alcuni esempi significativi tratti dall'articolo citato e da un altro della stessa autrice.

Una "usable version" del MARC: con esplicitazione dei valori codificati

La trasformazione nel *Wikibase Data Model* può fornire una "usable version" del MARC: è noto (ed è anche ovvio) che chiunque o qualunque sistema sia chiamato a lavorare sui dati trasferiti mediante il MARC deve fare una propria rielaborazione del formato;[37] in Wikibase le definizioni e le descrizioni del dato (*Label*, *Description* e *Alias* dei *Fingerprint*) coesistono con esso nel medesimo sistema mediante l'uso delle *Property* e delle relazioni fra *Entity*.[38] Sfruttando questo meccanismo, i coded-data del MARC possono essere registrati con puntatori alle *Entity* in cui si trovano le definizioni dei singoli dati (si veda sopra il paragrafo "Impostazione relazionale"); lo stesso può essere fatto per i dati registrati in MARC come segnaposto nei campi a lunghezza fissa: tutti i codici previsti per i casi di ambiguità (il "fill character" per i casi in cui non è possibile definire il valore corretto, il "blank" da usare come semplice segnaposto e gli altri "Unknown", "Combination", "Not applicable", "Not present" e "Other") possono essere gestiti con apposite *Entity* che contengono le definizioni corrispondenti.[39] Analogamente, possono essere create specifiche *Entity* per tutti i valori previsti negli indicatori dei campi MARC, compreso il valore "Undefined".[40]

[37] *Anyone wishing to develop applications for MARC21 must create their own usable version* (Coyle 2011)

[38] *My own database still lacks definitions and descriptions, and those will probably need to be added by screen-scraping hundreds of screens from the LC web site.* (Coyle 2011)

[39] *The use of values for "unknown" and "no attempt to code" are directly related to the characteristic of fixed length fields made up of positional data elements, where all positions must be filled in to retain the positioning.* (Coyle 2011)

[40] *In addition to the fixed field values like "Unknown," almost 60% of indicator positions (206 out of 350 in my database) have the value "Undefined." These truly represent empty positions in the record format and they can be ignored.* (Coyle 2011)

Individuazione dei dati univoci e delle rispettive fonti

Le relazioni fra *Entity* supportate dal *Wikibase Data Model* e la conseguente articolazione in *Entity* univoche nella trasformazione dal formato MARC, permettono di garantire l'individuazione dei dati univoci, strutturali e di contenuto. Questo rende completamente trasparente la traduzione dei dati codificati con i rispettivi nomi e le definizioni in linguaggio naturale: le voci di liste di autorità e vocabolari controllati sono registrate in *Item* autonomi, e vengono automaticamente sostituiti da Wikibase con la *Label* (nell'interfaccia grafica) e con l'identificativo (negli export di dati) dell'*Item* che ospita il dato univoco.[41] Tutte le *Entity* del *Wikibase Data Model* sono univocamente identificate con delle URI, così come richiesto nel contesto del Semantic Web;[42] viene garantita la possibilità di registrare il puntatore alla fonte del dato, al livello di ogni singolo *Statement* (corrispondente a campo UNIMARC) mediante la *Source* e, con la soluzione descritta, si garantisce anche la possibilità di ricostruire l'esatta sequenza dei dettagli originali.[43]

Statistiche degli elementi descrittivi

Una volta caricati in Wikibase, tutti gli elementi della semantica del MARC (campi, sottocampi, indicatori, valori posizionali in campi a lunghezza fissa e voci dei vocabolari controllati) sono disponibili non solo in elenco completo e interrogabile, ma anche con le statistiche del loro utilizzo. Le quantificazioni si possono ottenere mediante le interrogazioni supportate dal sistema stesso oppure mediante quelle supportate dagli endpoint SPARQL: ad esempio, fra le interrogazioni proposte nella documentazione del Wikidata Query Service,[44] una query raccoglie tutti gli *Item* con una specifica *Property* (la *Property* "nome del database Wikimedia", P1800);[45] la stessa query può essere facilmente adattata per restituire il conteggio di tali *Item*,[46] fornendo così una statistica dell'utilizzo della *Property*.[47]

---

[41] *The fixed fields are fixed-length strings with positional data elements that take coded data that is presumably useful for machine processing. These are primarily in the form of controlled term lists with the terms represented by 1-2 character codes* (Coyle 2011)

[42] *All of the data elements in my study must be assigned an identifier, and I use http URIs for this under the registered domain name "marc21.info".* (Coyle 2011)

[43] *It is also convenient if the description of the data elements contains information that would lead back to the original encoding of that data in MARC21.* (Coyle 2011)

[44] Si veda la documentazione all'indirizzo https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples#All_items_with_a_property

[45] Si veda la query SPARQL http://tinyurl.com/yatx3324.

[46] Si veda la query SPARQL http://tinyurl.com/ya8d52zr.

[47] *Having the MARC21 fixed fields, tags and subfields in a database first allowed me to do some quick statistics based on the names of fields and fixed field values.* (Coyle 2011)

Condivisione dei dati bibliografici fra sistemi di differente natura

Wikibase può facilitare la condivisione di parti componenti del record, anche partendo dalle più semplici: in Wikibase potrebbero confluire le parti condivise del record, anche prodotte a differenti livelli di approfondimento e in base a differenti standard (si veda di seguito, più in dettaglio).[48] I dati bibliografici condivisi potrebbero risiedere in una installazione centrale comune.[49]

Il Wikibase Data Model è predisposto per la modifica concorrente dei medesimi dati

Wikibase può gestire la semantica del MARC in un nuovo formato di archiviazione nativamente ideato per la condivisione e la modifica partecipata, anche di singole parti del record.[50] Il *Wikibase Data Model* è un formato di archiviazione aperto e condiviso con una vasta comunità di utilizzatori, anche in altri ambiti.[51] Wikibase (anche se utilizza DBMS relazionali come MySQL o MariaDB) si basa sull'archiviazione dei dati in triple.[52]

Reindirizzamento fra Item

Wikibase gestisce il reindirizzamento nel caso di "sostituzioni" di un *Item*: il vecchio *Item* rimarrà conservato, con il suo *Fingerprint* e tutte le sue informazioni, ma conterrà il rimando al nuovo *Item* che lo sostituisce. Questa funzionalità fornisce una soluzione tecnica al problema definito, in gergo nell'ambiente SBN, degli "schiacciamenti": in pratica l'esigenza di fondere due entità quando si verifica che una di queste è stata creata per errore.

---

[48] *Output rates of intellectual and cultural products is increasing. Libraries have already responded to this through shared cataloging and purchase of cataloging from product vendors. However, the records produced in this way are then loaded into thousands of individual catalogs in the MARC-using community.* (Coyle 2017)

[49] *Those records are often edited for correctness and enhanced. Thus they are costing individual libraries a large amount of money, potentially as much or more than libraries save by receiving the catalog copy.* [...] *The repeated storing of the same data in thousands of catalogs means not being able to take advantage of true sharing. In a cloud solution, records would be stored once (or in a small number of mirrors), and a record enhancement would enhance the data for each participant without being downloaded to a separate system.* (Coyle 2017)

[50] *"Sharing" in today's environment means exporting data and sending it as a file. Since MARC records can only be shared as whole records, updates and changes generally are done as a "full record replace" which requires a fair amount of cycles.* (Coyle 2017)

[51] *Moving more toward open source would be facilitated by moving away from a library-centric data standard and using at least a data structure that is commonly deployed in the information technology world.* (Coyle 2017)

[52] *NoSQL databases and triple stores. The current batch of databases are open source, fast, and can natively process data in a variety of formats (although not MARC)* [...] *This makes deployment of systems easier and faster.* (Coyle 2017)

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Figura 32: esempio di reindirizzamento fra due *Item* in Wikibase

Cronologia di tutte le versioni degli Item

Wikibase permette di preservare lo storico degli interventi di modifica su tutte le *Entity* (sia *Item* sia *Property*) registrate nel sistema: tutte le revisioni delle *Entity* sono conservate nella cronologia con i riferimenti di data e autore delle modifiche e, nella migliore tradizione Wiki, con il supporto delle discussioni riguardo ciascuna versione.
Le versioni possono essere singolarmente ripristinate a posteriori.

Figura 33: cronologia dell'*Item* corrispondente all'autore Galimberti, Fabio

Esposizione dei dati in RDF

Benché nel *Wikibase Data Model* sia implementata una versione "arricchita" della tripla semantica, Wikibase fornisce anche un export di tutti gli elementi descrittivi delle *Entity* in RDF standard.
La soluzione è implementata nei Wikidata Toolkit (una libreria di strumenti aggiuntivi connessa al software Wikibase),[53] e si basa sul meccanismo della "reification": quel processo con il quale, nell'ambito del Web Semantico, si rappresentano strutture di dato complesse (come gli *Statement* del *Wikibase Data Model* che contengono anche *Qualifier* e *Sources*) introducendo nuovi elementi che permettano di utilizzare solo triple semantiche elementari (Erxleben, Günther, Krötzsch, Mendez, Vrandečić 2014).
Mediante questo servizio già presente in Wikibase, dunque, il caricamento di tutto il dettaglio informativo dell'UNIMARC testato nella sperimentazione ottiene anche l'effetto di una mappatura dello standard verso l'RDF e, conseguentemente, la possibilità di esporre i dati con i consueti metodi del Semantic Web
A titolo di esempio, riprendendo il principale record UNIMARC testato nella sperimentazione, l'export Turtle dell'RDF (nella parte subito successiva alle dichiarazioni iniziali dei prefissi, *@prefix*) presenta i contenuti del *Fingerprint* dell'*Item* e, di seguito, l'elenco dei *Target* di tutti gli *Statement* con i quali si sono mappati i campi del record: si noti, nella Figura 34, la stringa completa "Lo stato innovatore..." per lo *Statement* con *Property* "P754" (area del titolo), e la presenza di puntatori ad altri *Item* per i campi dei punti di accesso, ad esempio "Q152" (l'*Item* di Galimberti, Fabio) per la *Property* "P1195" (Nome di persona: responsabilità secondaria).

[53] Il sito di riferimento della libreria è https://www.mediawiki.org/wiki/Wikidata_Toolkit.

```
wdata:Q156 a schema:Dataset ;
    schema:about wd:Q156 ;
    cc:license <http://creativecommons.org/publicdomain/zero/1.0/> ;
    schema:softwareVersion "0.0.1" ;
    schema:version "2504"^^xsd:integer ;
    schema:dateModified "2016-08-02T22:58:11Z"^^xsd:dateTime .

wd:Q156 a wikibase:Item ;
    rdfs:label "Lo  Stato innovatore"@en ;
    skos:prefLabel "Lo  Stato innovatore"@en ;
    schema:name "Lo  Stato innovatore"@en ;
    rdfs:label "Lo  Stato innovatore"@it ;
    skos:prefLabel "Lo  Stato innovatore"@it ;
    schema:name "Lo  Stato innovatore"@it ;
    skos:altLabel "CFI0893220"@en,
        "IT\\ICCU\\CFI\\0893220"@en,
        "CFI0893220"@it,
        "IT\\ICCU\\CFI\\0893220"@it ;
    wdt:P644 "IT - 2014-2498" ;
    wdt:P779 "Roma ; Bari : GLF editori Laterza, 2014" ;
    wdt:P657 "20140514d2014    ||||0itac50      ba" ;
    wdt:P923 "IT - IT-FI0098 - 20140514" ;
    wdt:P754 "Lo Stato innovatore : sfatare il mito del pubblico contro il privato / Mariana Mazzucato ; traduzione di Fabio Galimberti" ;
    wdt:P793 "Anticorpi ; 43" ;
    wdt:P620 "9788858111703" ;
    wdt:P618 "CFI0893220" ;
    wdt:P673 "it" ;
    wdt:P619 "20160201124421.0" ;
    wdt:P671 "ita" ;
    wdt:P788 "XXVI, 351 p. ; 21 cm" ;
    wdt:P1195 wd:Q152 ;
    wdt:P1193 wd:Q155 ;
    wdt:P1218 wd:Q157 ;
    wdt:P1224 wd:Q159 ;
    wdt:P1226 wd:Q160 ;
    wdt:P1228 wd:Q161 ;
    wdt:P603 "\"01429nam0 2200313 i 450 \"" ;
    wdt:P1255 "z01 - i - xxxe" ;
    wdt:P1261 "z01 - n" ;
    p:P644 wds:Q156-d549bbc7-46cd-da5f-95f7-e7fc0d24eef8 .
```

Figura 34: vista parziale dell'esposizione RDF dell'*Item* corrispondente al record UNIMARC BID IT\ICCU\CFI\0893220

Grazie alle scelte effettuate nell'implementazione del *Wikibase data Model*, l'export RDF può gestire un determinato contenuto informativo sia come insieme (per esempio Area 1 ISBD o etichetta UNIMARC 200) sia come elemento atomico (per esempio Titolo proprio o sottocampo a): questa prima parte dell'esposizione RDF (Figura 34) può garantire l'accesso (in interrogazione o estrazione dei dati) al contenuto dei campi UNIMARC "per intero" (quali l'Area 1 ISBD, mappata con la Property P754).

Più avanti nella medesima esposizione testuale RDF (Figura 35), si trova anche il dettaglio di *Qualifier* e *Source* per ciascuno *Statement*: ad esempio quello corrispondente all'UNIMARC 200 (Area 1 ISBD), dove si possono riconoscere i contenuti di tutti i sottocampi (compreso il sottocampo a, mappato con la *Property* P755) e della *Source* "Biblioteca Nazionale Centrale di Firenze" che è stata tradotta con la *wikibase:Reference* "wdref:872bd272cb87cb64b2cc120c5ec14406a1546efb".

```
wds:Q156-43f2e678-4ac0-048e-003f-217c61585270 a wikibase:Statement,
        wikibase:BestRank ;
    wikibase:rank wikibase:NormalRank ;
    ps:P754 "Lo Stato innovatore : sfatare il mito del pubblico contro il privato / Mariana Mazzucato ; traduzione di Fabio Galimberti"
    pq:P1250 wd:Q180 ;
    pq:P755 "Lo Stato innovatore" ;
    pq:P759 "sfatare il mito del pubblico contro il privato" ;
    pq:P760 "Mariana Mazzucato" ;
    pq:P761 "traduzione di Fabio Galimberti" ;
    pq:P11 "a[1] e[1] f[1] g[1]" ;
    prov:wasDerivedFrom wdref:872bd272cb87cb64b2cc120c5ec14406a1546efb .

wd:Q156 p:P793 wds:Q156-79885b6a-46bf-c78e-f24b-8226b526ed69 .
```

Figura 35: altra vista parziale dell'esposizione RDF dell'*Item* corrispondente al record UNIMARC BID IT\ICCU\CFI\0893220

Questo livello più dettagliato dell'esposizione RDF può garantire l'accesso più mirato al contenuto di sottocampi, di indicatori e dei dati registrati nelle stringhe delle informazioni codificate (come i campi 1xx dell'UNIMARC).

Wikibase può gestire molteplici modelli descrittivi nel medesimo Item

Sfruttando la possibilità di qualificare ciascuno *Statement* e di specificarne la fonte, diventa possibile registrare anche lo standard adottato per ciascun campo archiviato: questo rende tecnicamente possibile gestire, all'interno del medesimo *Item*, descrizioni prodotte in base a molteplici standard e provenienti da molteplici fonti, lasciando sempre la facoltà di riconoscere e quindi filtrare i campi di interesse.
Questo concetto di "compresenza" di molteplici standard nel medesimo *Item*, costituisce un'opzione diversa e aggiuntiva rispetto a quella (descritta sopra) di mappare a posteriori gli elementi strutturali dei dati già caricati, e si riferisce alla possibilità di creare lo stesso elemento seguendo standard o regole catalografiche diverse: per esempio, la data di pubblicazione può essere descritta seguendo sia le regole di trascrizione delle informazioni previste da ISBD, sia quelle previste da RDA.

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

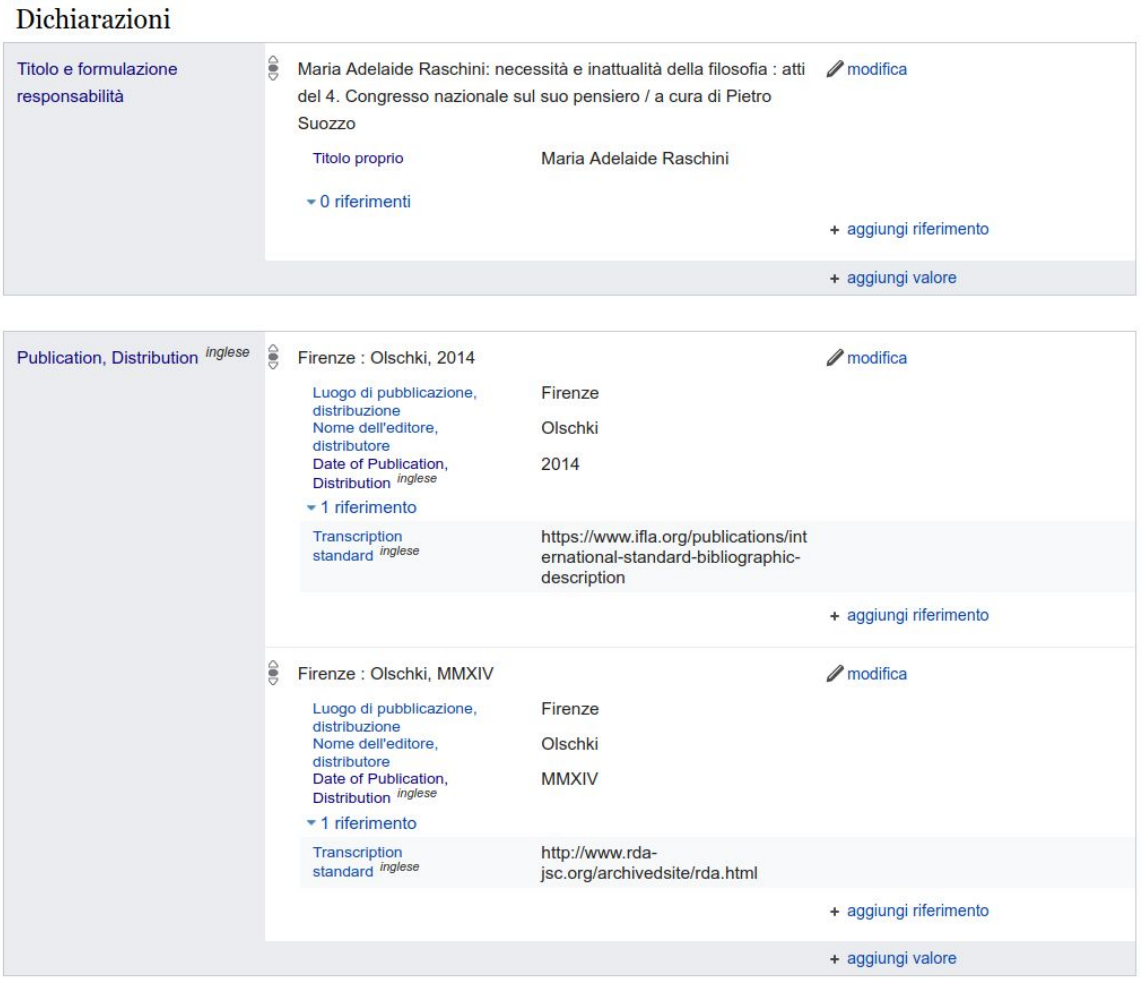

Figura 36: esempio di 2 *Statement* corrispondenti ai dati di pubblicazione compilati seguendo rispettivamente le regole di trascrizione delle informazioni previste da ISBD, e quelle previste da RDA (sperimentato su https://test.wikidata.org/wiki/Q166411) -- *nota per la versione italiana: l'immagine è stata aggiornata rispetto alla versione inglese https://test.wikidata.org/w/index.php?title=Q166411&oldid=396598*

Wikibase può gestire molteplici flussi di modifica

Wikibase propone interfacce di colloquio (API) aperte e documentate, che consentono a sistemi esterni di interrogare e modificare i dati gestiti; tali interfacce, assieme alla presenza di una grande quantità di librerie di software di base prodotte dalla comunità Wikimedia per il colloquio con Wikibase, danno la possibilità di esporre i dati a molteplici flussi di modifica dall'esterno.

Possibili direzioni

Come ricordato, nella maggior parte delle iniziative esistenti la pubblicazione come LOD (Linked Open Data) dei dati bibliografici diventa una attività *aggiuntiva* rispetto al flusso tradizionale di creazione e fruizione del record bibliografico: in pratica si pubblicano come LOD i dati

bibliografici esistenti e la loro pubblicazione non influenza la loro creazione. In queste iniziative la pubblicazione dei dati avviene mediante un processo - di solito automatico - di conversione semantica (non solo sintattica): l'ontologia (o lo schema di metadati) tradizionalmente in uso viene sostituita da una nuova ontologia ritenuta più adatta al W*eb dei dati*. Uno dei punti di riferimento di queste iniziative è il *modello concettuale FRBR*.[54] Fa eccezione BIBFRAME che - come ricordato - si pone sia il problema di nuove modalità di gestione del record bibliografico, sia le problematiche della conversione semantica completa del formato MARC21.

La proposta presentata in questo contributo parte dalla considerazione che l'UNIMARC (e in generale il MARC) è di fatto una ontologia e si propone di renderla esplicita e di convertirla - solo a livello sintattico - in strutture Linked Data/RDF mediante l'uso del *Wikibase data model*. I dati bibliografici ospitati in Wikibase potrebbero quindi diventare non solo un modo per pubblicare i dati come LOD, ma anche un ambiente per la produzione dei dati bibliografici. Molti potrebbero essere i vantaggi di questo nuovo ambiente, tra i quali:

- le potenzialità di collaborazione del modello Wiki;
- la capacità nativa di trattare i dati in ambiente multilingue;
- la disponibilità dei dati (in lettura e in scrittura) con protocolli aperti (SPARQL, API Wiki, ecc);
- la disponibilità di software e di programmi di utilità aperti con una larga base di sviluppo e sperimentazione a livello mondiale.

A questi vantaggi se ne deve aggiungere un altro: l'offerta di strumenti molto agili - in particolare la capacità di ospitare differenti ontologie - che possono fare da supporto anche all'evoluzione della riflessione in corso sul controllo bibliografico (da FRBR a LRM[55]) senza avere - come punto di partenza - conversioni di natura semantica dei dati pregressi che, in un ambiente di produzione, possono essere davvero costose e problematiche.

In oltre mezzo secolo di applicazione del MARC abbiamo già assistito a un trasferimento della semantica dello standard (mantenuta sostanzialmente inalterata) da una "sintassi di base" - ISO 2709 - a un'altra - XML - con la relativa evoluzione dei protocolli di interrogazione, dallo z39.50 ai diversi protocolli dei servizi web (web service) che supportano il linguaggio XML via HTTP (per esempio SRU[56] e SRW[57]). Su questa strada questo contributo propone di veicolare la semantica del MARC (preservandone tutto il dettaglio informativo) attraverso una nuova modalità sintattica, il *Wikibase Data Model*, che permette di sfruttare un sistema di gestione già ampiamente utilizzato, il software Wikibase, completo dei protocolli, le API, per l'interrogazione e la modifica dei dati.

Il duplice effetto di questo aggiornamento della "sintassi di base" del MARC vuole essere da una parte quello di traghettare i dati bibliografici prodotti nel formato MARC (nella completezza della loro versione originale) verso il *mainstream* dei più diffusi metodi di condivisione dei dati, e dall'altra parte fornire uno strumento operativo che faciliti la mappatura del MARC verso le nuove proposte di natura semantica (o ontologica) per la gestione dei dati bibliografici.

[54] Il punto su FRBR si può trovare in (Karen, 2016)

[55] Il sito di riferimento per IFLA RLM è: https://www.ifla.org/publications/node/11412.

[56] http://www.loc.gov/standards/sru/

[57] http://www.loc.gov/standards/sru/companionSpecs/srw.html

Il contesto italiano

Per quanto riguarda lo specifico caso italiano, utilizzare il *Wikibase data model* come modalità di pubblicazione dei LOD, ma soprattutto sperimentare il nuovo ambiente di produzione potrebbe risultare una proposta interessante anche per il Servizio Bibliotecario Nazionale. In particolare il protocollo SBNMARC[58] (basato su messaggi XML trasmessi via HTTP) che viene attualmente utilizzato dai Poli locali per aggiornare il catalogo della cooperazione o Indice SBN, potrebbe leggere e scrivere - con una opportuna interfaccia - anche sul "nuovo ambiente di produzione Wikibase" implementato come descritto in questo contributo: in questo modo (per esempio per gestire una fase di transizione) i software attualmente in uso nei poli SBN potrebbero aggiornare i dati del nuovo ambiente di produzione. Questa ipotesi andrebbe naturalmente approfondita, dettagliata e sperimentata, ma dal punto di vista tecnologico è una strada percorribile. In estrema sintesi il protocollo di colloquio SBNMARC veicola essenzialmente informazioni codificate di due tipi. Il primo tipo di informazioni è relativo ai servizi applicativi che l'Indice centrale mette a disposizione dei sistemi locali (o poli): si tratta di uno scambio di messaggi di richiesta dei poli (per esempio: Cerca, Crea, Modifica, Cancella, Allinea) e di risposta dell'Indice. Il secondo tipo riguarda le informazioni bibliografiche che vengono scambiate nei messaggi e queste ultime - come è noto - sono interamente conformi alla semantica UNIMARC. Per permettere ai Poli locali di colloquiare con un "nuovo ambiente di produzione Wikibase" sarebbe, dunque, necessario mappare i servizi applicativi del protocollo SBNMARC con le interfacce di colloquio (API) aperte e documentate di Wikibase (richiamate nel paragrafo precedente).
In ogni caso, vista la mancanza in Italia di una iniziativa su larga scala riguardante i dati bibliografici come LOD, anche la sola pubblicazione dei dati SBN usando il *Wikibase Data Model* - e quindi senza un cambio immediato di ontologia - sarebbe un passo significativo.

[58] Il sito di riferimento del protocollo SBNMARC è http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/evoluz_indice_sbn/pagina_144.html.

Versione italiana dell'articolo: Bergamin, G., Bacchi, C. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". JLIS.it 9, 3 (September 2018): 35-74. DOI: 10.4403/jlis.it-12458.

Riferimenti bibliografici

Baker, Tom. 2013. "La traduzione dei dati nel linguaggio del web semantico". Jlis, 4 (1). https://www.jlis.it/article/view/6308/7866.

Berners-Lee, Tim. 2006. "Linked data" http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, Tim, James Hendler e Ora Lassila. 2001. "The semantic web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". *The scientific American*. http://www.scientificamerican.com/article.cfm?id=the-semantic-web

Caplan, Priscilla. 2003. "Metadata fundamentals for all librarians". Chicago:ALA.

Coyle, Karen. 2011. "MARC21 as Data: A Start". Code4Lib Journal, 14.

Coyle, Karen. 2012. "Linked data tools: connecting on the Web". ALA Library Technology Reports , 48 (4).

Coyle, Karen. 2016. "FRBR, Before and After: A Look at Our Bibliographic Models". Chicago:ALA. http://www.kcoyle.net/beforeAndAfter/978-0-8389-1364-2.pdf

Coyle, Karen. 2017. "Precipitating Forward" http://kcoyle.blogspot.it/2017/04/precipitating-forward.html

Erxleben, Fredo, Michael Günther, Markus Krötzsch, Julian Mendez, e Denny Vrandečić. 2014. "Introducing Wikidata to the Linked Data Web". In *Proceedings of the 13th International Semantic Web Conference*: 50–65. Springer.

Hallo, María, Sergio Luján-Mora, Alejandro Maté, and Juan Trujillo. 2016. "Current state of Linked Data in digital libraries". *Journal of Information Science* 42: 117-127, first published on July 21, 2015 doi:10.1177/0165551515594729.

Svenonius, Elaine. 1992. "Bibliographic Entities and Their Uses." in *Proceedings of the Seminar on Bibliographical Records, held in Stockholm, 15-16 August 1990*. Munchen: K.G. Saur: 3-18.

Tennant, Roy. 2017. "'MARC Must Die' 15 Years On" https://hangingtogether.org/?p=6221 .

Vrandečić, Denny and Markus Krötzsch. 2014. "Wikidata: A Free Collaborative Knowledgebase". In *Communications of the ACM,* 57 (10): 78–85.