

La nuova LOD Platform di SHARE Catalogue: un'evoluzione nel segno delle pratiche collaborative della Share Family

Claudio Forziati, Annalisa Di Sabato, Rossella Molisso, Chiara Mugnano

Abstract: Il contributo descrive la progettazione, lo sviluppo e l'evoluzione di SHARE Catalogue, concepito per integrare, in un catalogo collettivo in linked open data, i dati bibliografici delle istituzioni aderenti alla convenzione SHARE. Sono analizzati, tra i recenti processi che hanno interessato il catalogo, sia la mappatura UNIMARC-BIBFRAME 2.0, curata da un gruppo tecnico interateneo, sia la sua trasposizione, tuttora in corso, in un'istanza di Wikibase dedicata. I diversi e successivi stadi della piattaforma sono descritti per evidenziarne la prossimità alle pratiche di collaborazione promosse nel contesto della Share Family e per affermare quanto queste pratiche siano cruciali per una rappresentazione efficace delle risorse culturali nel web dei dati, compatibilmente con le esigenze di qualità e riuso delle informazioni prodotte dalle istituzioni e di sostenibilità a lungo termine.

Parole chiave: Linked Open Data; UNIMARC mapping; Wikibase; Wikidata; BIBFRAME.

Abstract: *This paper describes the conception, development and evolution of the SHARE Catalogue, which was designed to integrate the bibliographic data of the institutions participating in the SHARE Consortium in a collective catalogue of linked open data. Among the recent processes concerning the catalogue, this article analyses both the UNIMARC-BIBFRAME 2.0 mapping, managed by an inter-university technical group, and its ongoing transposition into a dedicated Wikibase instance. The different and successive stages of the platform are described in order to highlight its proximity to the collaborative practices promoted in the context of the Share Family, and to affirm how crucial these practices are for an effective representation of cultural resources in the web of data, in line with the requirements of quality and re-use of the information produced by the institutions, and long-term sustainability.*

Keywords: *Linked Open Data; UNIMARC mapping; Wikibase; Wikidata; BIBFRAME.*

Claudio Forziati, University of Naples Federico II, Italy, claudio.forziati@unina.it, 0000-0002-8976-0393
Annalisa Di Sabato, @Cult, Italy, annalisa.disabato@atcult.it, 0009-0003-8020-6059
Rossella Molisso, University of Naples Federico II, Italy, rossella.molisso@unina.it, 0009-0001-7213-7941
Chiara Mugnano, University of Salerno, Italy, cmugnano@unisa.it, 0009-0006-9489-5432

Referee List (DOI 10.36253/fup_referee_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Claudio Forziati, Annalisa Di Sabato, Rossella Molisso, Chiara Mugnano, *La nuova LOD Platform di SHARE Catalogue: un'evoluzione nel segno delle pratiche collaborative della Share Family*, © 2024 Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0356-2.14, in *Unione Romana Biblioteche Ecclesiastiche, Parsifal. Un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*, edited by Silvano Danieli, pp. 105-25, 2024, published by Firenze University Press, ISBN 979-12-215-0356-2, DOI 10.36253/979-12-215-0356-2

Acknowledgment

Il presente contributo è frutto della riflessione condivisa tra gli autori nell'ambito dei lavori del Gruppo tecnico di mappatura UNIMARC-BIBFRAME 2.0, operante nella convenzione interuniversitaria SHARE¹. Nello specifico: Claudio Forziati ha curato i paragrafi 1 e 3, il paragrafo 5 con Rossella Molisso, il paragrafo 6 con Annalisa Di Sabato; Annalisa Di Sabato i paragrafi 2 e 4.2; Chiara Mugnano i paragrafi 4.1, 4.3 e 4.4.

Si ringraziano il prof. Roberto Delle Donne e la dott.ssa Tiziana Possemato per la revisione del testo; le colleghe Stefania Castanò, Paola Denunzio, Alessandra Moi e Marta Modugno per l'attiva partecipazione al Gruppo tecnico. Il contributo è dedicato alla memoria della collega Gabriella Muccione, scomparsa nel luglio 2023.

1. La convenzione interuniversitaria e la nascita di SHARE Catalogue

*SHARE Catalogue*² è un progetto nato nell'ambito della convenzione interuniversitaria SHARE (Scholarly Heritage and Access to Research), siglata per la prima volta nel 2015. Oggetto principale della convenzione è la «realizzazione di un sistema integrato di sviluppo, fruizione e gestione dei propri servizi d'informazione, documentazione e supporto alla didattica, alla ricerca e al trasferimento della conoscenza»³, in cui gli Atenei partecipanti si impegnano a mettere in comune un complesso di piattaforme e progetti «orientati all'integrazione delle attività e allo sviluppo di servizi avanzati»⁴ per le proprie comunità di riferimento. Tra gli esiti più importanti della convenzione c'è il riconoscimento degli utenti degli Atenei firmatari come utenti interni comuni, ai quali è garantito un medesimo livello di accesso ai servizi⁵. A questo si affiancano inoltre le piattaforme condivise di editoria ad accesso aperto (SHARE Journals, SHARE Books e SHARE Open Archive) e il discovery bibliografico per la ricerca federata SHARE Discovery⁶.

Dalle iniziali sei, ad oggi le istituzioni aderenti sono diventate undici e, con livelli di partecipazione differenziati sulla base delle proprie specificità, tutte condividono l'oggetto comune in termini di impegno, collaborazione e reciprocità.

¹ Il testo della convenzione, siglato dai partecipanti nel 2015, e i successivi aggiornamenti sono reperibili a partire dall'indirizzo <<https://www.sharecampus.unina.it>>.

² <<https://catalogo.share-cat.unina.it>>

³ Art. 2, *Convenzione interuniversitaria per l'integrazione di servizi bibliotecari e documentari 2015-2017*, <https://www.sharecampus.unina.it/wp-content/uploads/2023/10/b80db-convenzione_interuniversitaria.pdf>.

⁴ *Ibidem*.

⁵ Per la consultazione della biblioteca digitale dei singoli Atenei, si veda *Carta dei servizi (2018-2022)*, p. 4, <https://www.sharecampus.unina.it/wp-content/uploads/2023/10/be680-convenzione_carta-servizi_v07.pdf>.

⁶ SHARE Journals, <<https://www.serena.unina.it>>. SHARE Books, <<https://www.fedoabooks.unina.it/index.php/fedoapress>>. SHARE Open Archive, <<https://elea.unisa.it/handle/10556/2055>>. SHARE Discovery, <<https://discovery.sharecampus.it>>.

In questo contesto consortile, SHARE Catalogue ha rappresentato il tentativo, unico nel suo genere all'epoca della sua ideazione, di portare un progetto di cooperazione bibliografica, basato sull'integrazione dei dati gestiti dai partecipanti, in una nuova dimensione tecnologica (cfr. Possemato e Delle Donne 2017, 21-29).

Pubblicato online nel 2016 come piattaforma organizzata secondo il modello BIBFRAME⁷, ha risposto all'esigenza di dotare le biblioteche di un motore di ricerca innovativo, attraverso la realizzazione di un punto di accesso unificato e integrato all'insieme delle risorse, fornendo agli utenti uno strumento in linea con l'evoluzione delle tecnologie del web semantico.

Nella creazione di SHARE Catalogue si è tenuta presente la funzione che i dati bibliografici possono svolgere oggi nel web, cioè l'aspirazione a essere direttamente reperibili dai motori di ricerca. Per rendere concreto questo cambiamento è necessario un processo complesso che integri più profondamente le ontologie proprie del web, parallelamente a una «integrazione sempre più ampia con il contesto culturale e tecnologico usato dagli altri soggetti della trasmissione della conoscenza registrata» (Guerrini 2019, 65). Punto di partenza è il record MARC che, trasferito nel web semantico, diventa fonte d'informazione preziosa per la grande quantità di dati che contiene e per la ricchezza di significato in esso espressa: «gli stessi metadati codificati nei tag dei record, tradotti in asserzioni, riconquistano il valore semantico originario che diventa comprensibile e usabile da una macchina» (Guerrini e Possemato 2015, 130).

2. Obiettivi, analisi dei dati e processo di clusterizzazione in SHARE Catalogue 1.0

La progettazione di SHARE Catalogue ha richiesto la definizione di obiettivi molto articolati, sintetizzabili nella volontà di realizzare una base di dati comune per la costituzione di una rete di conoscenza aperta e riutilizzabile. Per raggiungere questo obiettivo è stato necessario adottare diverse soluzioni tecnologiche, tra le quali:

- la creazione di una struttura di dati dove fosse possibile identificare, sulla base del modello FRBR:
 - potenziali entità quali Persona, Famiglia, Ente, Opera e Manifestazione;
 - le relazioni che intercorrono fra le entità individuate.

L'identificazione delle entità e la definizione dei loro collegamenti⁸, espressi dalle relazioni, consentono la realizzazione della funzione di navigazione

⁷ <<https://www.loc.gov/bibframe/docs/bibframe2-model.html>>.

⁸ Identificare e collegare le entità sono due degli obiettivi fondamentali di RDA (Resource Description and Access), standard per la descrizione e l'accesso alle risorse bibliografiche che adotta i principi del web semantico e focalizza le linee guida sul concetto di dato e sulla creazione delle relazioni. Vedi Joint Steering Committee for Development of RDA, *RDA: Resource Description and Access*, 2013, <<https://www.rda-rsc.org>>.

nel portale, la cui interfaccia di ricerca avanzata migliora l'esperienza d'uso e fornisce risultati più ampi agli utenti, consentendo l'accesso alla grande quantità di risorse disponibili, solitamente esposte in maniera frammentata nei diversi cataloghi;

- l'arricchimento dei dati tramite il collegamento a fonti esterne, quali: VIAF (Virtual International Authority File), LCNAF (Library of Congress Name Authority File), ISNI (International Standard Name Identifier), Wikidata e Wikipedia;
- la pubblicazione dei dati in Linked Open Data (LOD), per renderli disponibili su differenti siti e portali open, attraverso il modello RDF, proposto dal W3C, utile a caratterizzare semanticamente le risorse e le relazioni che intercorrono tra di esse. Questo modello consente di esprimere le informazioni in forma di triple (soggetto-predicato-oggetto), dove il soggetto si collega per mezzo di una proprietà (predicato), all'oggetto e dove ogni elemento della tripla può essere identificato da un URI. Nell'organizzazione dei dati in RDF, l'utilizzo di ontologie e vocabolari controllati assume grande importanza per la creazione di triple, in quanto attribuisce significato e definisce la semantica del dato.

Per la realizzazione di SHARE Catalogue è stato scelto BIBFRAME⁹, un modello estremamente funzionale, ispirato a FRBR, costituito da classi (Opera, Istanza, Item) e proprietà RDF.

I dati catalografici, in formato UNIMARC e MARC 21, derivanti dai diversi LMS in uso presso le istituzioni partecipanti (Alma, Aleph, Sebina, Millennium, ecc.), sono stati analizzati con lo scopo di individuare delle entità 'potenziali'. È stata eseguita un'indagine mirata ad analizzare ciascun elemento del record, per comprendere il suo utilizzo specifico ed evidenziare le pratiche catalografiche comuni. Questo approccio ha consentito di individuare i punti di convergenza tra i dati provenienti dalle diverse istituzioni e di identificare le discrepanze.

Una prima verifica ha mostrato un uso poco diffuso, nei record bibliografici, di attributi utili alla identificazione dell'entità *Opera* e dell'entità *Espressione* di

⁹ Questo nuovo modello per la strutturazione dei dati bibliografici viene proposto nel 2012 dalla Library of Congress affinché costituisca la base per il futuro della descrizione bibliografica e allo stesso tempo possa essere condiviso da altre istituzioni dei beni culturali. L'obiettivo principale del progetto è la sostituzione del MARC 21, a favore di una struttura bibliografica che funga da ambiente piuttosto che essere vincolata a un formato specifico, garantendo così un approccio neutro rispetto alle regole catalografiche. I principi fondamentali su cui si basa includono la distinzione chiara tra i dati sul contenuto concettuale e quelli sulla manifestazione fisica di un oggetto bibliografico, l'ampio utilizzo di sistemi di controllo d'autorità per identificare in modo inequivocabile gli elementi informativi e la rappresentazione delle relazioni tra le entità per favorire la navigabilità. Allo stesso tempo, questo modello, come il formato MARC, mira a facilitare lo scambio di dati tra le biblioteche e deve essere compatibile con le nuove strutture di codifica dei metadati e gli standard di contenuto emergenti, includendo esplicitamente Resource Description and Access (RDA) tra le linee guida a cui BIBFRAME si riferisce. Vedi Miller et al. 2012.

FRBR, quali il punto di accesso autorizzato per il titolo o il titolo originale. L'insufficiente utilizzo di tali attributi ha reso più complesse le operazioni di identificazione, complicando la *Entity Resolution*, cioè l'insieme dei processi macchina che cerca di risolvere le ambiguità date dalla disomogeneità delle descrizioni riferibili alla medesima entità (Possemato 2022, 16).

Da qui l'idea di seguire un altro tipo di indagine, il cui oggetto fosse rappresentato dalle entità "Persona-Famiglia-Ente" (di seguito indicate come *Persona*) e dai relativi attributi: ci si è concentrati sull'analisi di tutti gli elementi impiegati per descrivere gli autori nei record di origine. Inoltre, per superare una possibile criticità derivante dai casi di scarso o difforme utilizzo di attributi qualificanti e permettere il processo di riconciliazione dei nomi¹⁰, si è deciso di confrontare i diversi punti di accesso con le intestazioni presenti nel VIAF¹¹.

Il cuore tecnologico di SHARE Catalogue, concepito nel contesto del progetto ALIADA¹², è stato riportato all'interno di una infrastruttura più adeguata a gestire quantità di dati rilevanti ed essere ampliata e potenziata. I dati ricevuti dalle istituzioni partecipanti sono stati elaborati secondo processi di *Text analysis* e *String matching* per identificare le entità incluse nei testi 'piatti' (i record) e preparare la creazione dei cluster. Questa funzione di identificazione delle entità è stata potenziata tramite processi di interrogazione delle differenti fonti esterne utilizzate nell'ambito del progetto (VIAF, ISNI, LCNAF, GND, Nuovo soggetto ecc.), attraverso il framework Authify. I dati originali sono stati, così, arricchiti con gli identificatori (URI) e le forme varianti dei nomi proposti sulle fonti esterne per le medesime entità. Questo arricchimento consente poi, in fase di conversione dei dati in linked data, il meccanismo dell'*interlinking*, essenziale per la condivisione e il riuso dei dati nel web.

I cluster così generati sono salvati nella Cluster Knowledge Base (CKB) di progetto (su database PostgreSQL¹³), i cui dati vengono utilizzati per produrre in output un file MARC arricchito. Tale file rappresenta l'input del modulo RESTFul¹⁴, responsabile del processo di conversione e pubblicazione dei dati in RDF secondo l'ontologia BIBFRAME 1.0.

¹⁰ Lo scopo di un processo di riconciliazione è quello di mettere insieme forme diverse di nomi per la realizzazione di un unico 'grappolo' di dati riferiti alla medesima entità, definito anche cluster.

¹¹ Nel VIAF non esiste una forma unica, universalmente riconosciuta come 'accettata' o 'uniforme' del nome di un'entità, né una forma che prevalga sulle altre. Le diverse forme equivalenti, provenienti da varie agenzie nazionali coinvolte nel progetto, insieme a tutte le altre forme con le quali il dato ricercato può essere rappresentato, sono collegate tra loro per essere associate allo stesso identificatore.

¹² Progetto open source finanziato dall'UE, all'interno del Seventh Framework Programme for Research (FP7), il cui obiettivo è stato la realizzazione di un framework di conversione e pubblicazione automatica in RDF dei dati appartenenti ad istituzioni culturali diverse. Cfr. Gazzarini 2015, 161.

¹³ <<https://en.wikipedia.org/wiki/PostgreSQL>>.

¹⁴ È il predecessore del modulo Lodify, attualmente utilizzato nel progetto Share-VDE, secondo l'ontologia BIBFRAME 2.0.

La produzione dei cluster relativi all'entità *Opera* è risultata, per vari motivi, piuttosto complessa, a causa dell'uso di diverse forme del titolo con cui l'*Opera* è stata identificata nei diversi cataloghi, per la mancanza di record di autorità per i titoli e per la carenza, come detto, di punti di accesso autorizzati nei record bibliografici. La costruzione è stata molto laboriosa, ma il risultato ottenuto è stato comunque notevole. Decine di titoli diversi, associati alle pubblicazioni presenti nei cataloghi, sono stati raggruppati a formare un unico cluster per ciascuna *Opera*.

L'identificazione delle entità *Opera* e *Persona*, riconciliate in cluster, ha contribuito alla realizzazione di un portale con tre livelli di navigazione:

- *Persone e Opere*: primo strato informativo di consultazione, arricchito dai dati provenienti da fonti esterne ai cataloghi delle biblioteche, dove possono essere recuperati e visualizzati i titoli delle opere per le quali nel catalogo collettivo in LOD sia presente almeno un titolo di un'istanza o pubblicazione.
- *Istanze o Pubblicazioni*: strato di consultazione realizzato utilizzando i dati bibliografici dei differenti cataloghi che, attraverso interventi di conversione e di matching, sono collegati al primo strato.
- *Item*: relativo alle informazioni sulle copie collegate ai record di Istanza/Pubblicazione (realizzato attraverso un permalink verso l'OPAC locale di ciascuna biblioteca).

3. Esplorare Wikidata per trasformare SHARE Catalogue: una sperimentazione

Il gruppo di lavoro responsabile della convenzione, nell'ambito delle iniziative di sviluppo del catalogo LOD, nel 2017 ha accolto la proposta di un gruppo di 'bibliotecari wikimediani' che ha suggerito di valorizzare il potenziale degli identificatori dei cluster *Persona* di SHARE Catalogue attraverso una maggiore interazione con Wikidata, progetto che iniziava ad affermarsi anche negli studi biblioteconomici, in virtù del momento favorevole alla riflessione sul concetto di integrazione con le realtà produttrici di conoscenza in forma di dati. La letteratura pubblicata fino a quel momento¹⁵, evidenziava la ricchezza di Wikidata in termini di connessione a fonti esterne estremamente variegata e di dati di contesto per le dichiarazioni presenti nella base di conoscenza¹⁶.

Questo approccio della letteratura ha giocato un ruolo centrale nell'indirizzare l'interesse del gruppo di lavoro verso l'interconnessione con una fonte di provenienza non esclusivamente bibliografica. Oggetto di questo interesse era proprio la natura *multipurpose* di Wikidata, una base di conoscenza aperta, arricchita collaborativamente, che consentiva di ipotizzare usi originali dei dati relativi alle

¹⁵ Tra i tanti, vedi Erxleben et al. 2014, 50-65. In italiano, vedi Martinelli 2016.

¹⁶ «The data is highly interlinked and connected to many other datasets» e ancora «Many statements come with provenance information or include additional context data, such as temporal validity» (Erxleben et al. 2014, 50).

Item	RemLabel	Image	share_cluster_ID
Q wdt:Q11609	Shafi Goldwasser		27502
Q wdt:Q15030	Éva Tardos		433692
Q wdt:Q82939	Christos Papadimitriou		291381
Q wdt:Q82964	Sanjeev Arora		480836
Q wdt:Q82927	Noga Alon		66369
Q wdt:Q82957	Avi Wigderson		738674
Q wdt:Q82963	Rajeev Motwani		65488
Q wdt:Q83080	Silvio Micali		746266
Q wdt:Q83124	Ronald Fagin		263320
Q wdt:Q83149	Madhu Sudan		753690
Q wdt:Q357916	László Lovász		41000
Q wdt:Q1585964	Moshe Y. Vardi		464788
Q wdt:Q82896304	Nir Shavit		518240
Q wdt:Q7045525	Noam Nisan		771621

Figura 1. Persone con un identificatore SHARE Catalogue, vincitori di Gödel Prize, di cui esista un'immagine in Wikimedia Commons, collegata tramite la proprietà P18 a Wikidata. Per una visualizzazione in forma di griglia d'immagini si veda all'URL <<https://w.wiki/9EkH>>.

entità potenzialmente connesse tra i due progetti¹⁷. Nello sviluppo originario di SHARE Catalogue c'era l'obiettivo dell'arricchimento dei dati grazie a fonti esterne, le stesse fonti (VIAF, ISNI, ecc.) che l'utente poteva consultare durante le proprie ricerche grazie ai collegamenti presenti nel portale; questa evoluzione intendeva intensificare il legame con Wikidata lavorando direttamente nella *knowledge base* della famiglia Wikimedia.

Tralasciando le fasi operative, altrove descritte nel dettaglio¹⁸, vale la pena di illustrare sinteticamente ciò che si voleva far emergere e gli obiettivi di questa particolare attività, per le sue ricadute sul futuro del progetto.

Innanzitutto la varietà di Wikidata consentiva di valorizzare i metadati prodotti dalle istituzioni facendoli interagire con informazioni reperibili al di fuori di quanto localmente ritenuto rilevante (cfr. Stinson 2018). Grazie al query service SPARQL di Wikidata è possibile combinare i dati frutto dell'*interlinking* con altri difficilmente reperibili nei record catalografici e, inoltre, è possibile generare delle visualizzazioni funzionali a una migliore comprensione delle relazioni intercorrenti tra una determinata entità, o un insieme di esse, e il contesto scientifico, culturale, storico di appartenenza. È il caso della tabella in figura 1: il dato dell'assegnazione del premio Gödel potrebbe essere riportato, anche in forma strutturata, nei rispettivi record di authority nel tag UNIMARC 334, ma di fatto, attualmente, non c'è e non è combinabile nel record con un'immagine della persona, oltre a non essere informazioni e risorse considerate, localmente, come gestibili in maniera costante e rilevanti dal punto di vista bibliografico.

¹⁷ Un esempio molto chiaro, a valle dell'inserimento di oltre duecentomila identificatori di SHARE Catalogue, è la possibilità di generare una *bubble chart*, aggiornata in tempo reale, dei primi venti luoghi di nascita in Italia degli umani presenti in SHARE Catalogue come visionabile all'indirizzo <<https://w.wiki/9Dwk>>.

¹⁸ Si veda in particolare: Forziati e Lo Castro 2018, 109-20, e Possemato e Forziati 2019, 228-38.

Label	Wikidata ID	Occupation	SHARE Cat ID	IRIS UNINA ID	VIAF ID	ORCID ID	SBN author ID	IdRef ID
Aldo Loris Rossi	Q1236941	architect	8584 [Ⓞ]	01804 [Ⓞ]	34713757 [Ⓞ]		CFIV059805 [Ⓞ]	184976553 [Ⓞ]
Donatella Mazzoleni	Q1240613	architect	10361 [Ⓞ]	01085 [Ⓞ]	30313 [Ⓞ]		CFIV095206 [Ⓞ]	096530537 [Ⓞ]
Nicola Fusco	Q1477114	mathematician university teacher	9279 [Ⓞ]	03887 [Ⓞ]	14993937 [Ⓞ]	0000-0001-8215-8351 [Ⓞ]	CFIV333796 [Ⓞ]	070707677 [Ⓞ]
Giancarlo Corsetti	Q1522354	economist university teacher	122773 [Ⓞ]	55122 [Ⓞ]	64160035 [Ⓞ]	0000-0001-8985-9853 [Ⓞ]	TOOV078929 [Ⓞ]	050316532 [Ⓞ]
Marcello Gigante	Q1893332	papyrologist classical philologist university teacher hellenist Byzantinist	36239 [Ⓞ]	12069 [Ⓞ]	109998667 [Ⓞ]		CFIV014251 [Ⓞ]	028370058 [Ⓞ]
Derrick de Kerckhove	Q1904943	sociologist writer linguist	40285 [Ⓞ] 514956 [Ⓞ]	12238 [Ⓞ]	31998360 [Ⓞ]	0000-0002-7359-5872 [Ⓞ]	CFIV332102 [Ⓞ]	02694779X [Ⓞ]
Andrea Graziosi	Q2066672	historian economist university teacher sovietologist	119723 [Ⓞ]	15484 [Ⓞ]	2805080 [Ⓞ]	0000-0002-1178-3085 [Ⓞ]	CFIV066717 [Ⓞ]	05720697X [Ⓞ]
Alberto Izzo	Q2831772	architect	435931 [Ⓞ]	11106 [Ⓞ]	12310157 [Ⓞ]		CFIV029247 [Ⓞ]	026932261 [Ⓞ]
Gabriele Frasca	Q3094072	writer poet	169992 [Ⓞ]	23326 [Ⓞ]	118224903 [Ⓞ]		CFIV077975 [Ⓞ]	029501717 [Ⓞ]
Giovanni Romeo	Q3107330	historian	174425 [Ⓞ]	01670 [Ⓞ]	66506503 [Ⓞ]		LO1V020023 [Ⓞ]	030157293 [Ⓞ]

Figura 2. La lista in figura, creata con il tool *Listeria*, è disponibile all'URL <<https://your.ls/xmm7a>>¹⁹.

Un altro aspetto estremamente vantaggioso proveniva dalla funzione di *hub* di identificatori da fonti esterne svolta da Wikidata (cfr. Neubert 2017, 14-25): come si vede in figura 2 è possibile generare una lista in cui vengono allineati identificatori non presenti originariamente nell'arricchimento pensato per SHARE Catalogue, provenienti da fonti legate alla produzione di ricerca (come il catalogo IRIS dell'Università degli studi di Napoli Federico II o più generali come ORCID) e altri di natura bibliografica come SBN e IdRef.

Wikidata si è rivelato, inoltre, un formidabile strumento per evidenziare alcune tipologie di errore, frequenti in cataloghi che adottano logiche di clusterizzazione simili a SHARE Catalogue, come duplicazioni e conflazioni, aiutando di fatto a migliorare la CKB e suggerendo azioni da adottare direttamente nei cataloghi, in assenza di un *linked data editor*.

È possibile, infatti, generare una lista delle violazioni di valore singolo presenti in Wikidata per la proprietà collegata agli identificatori di SHARE Catalogue: un esempio è mostrato in figura 3 dove, per l'item corrispondente a Giorgio Levi Della Vida²⁰, erano presenti inizialmente nel catalogo collettivo tre differenti cluster.

Dopo il report generato tramite query SPARQL, si è provveduto a normalizzare le forme del nome che creavano le duplicazioni in due cataloghi partecipanti, producendo per i cluster con identificatori 550090 e 550091 in SHARE Catalogue un *redirect* al cluster correttamente aggregato²¹. Una volta effettuate le

¹⁹ Un altro esempio di identificatori presenti in Wikidata è disponibile all'URL <<https://w.wiki/9UBi>>. In esso vengono mostrati i filologi romanzi nati nell'800, con identificatori di SHARE Catalogue, Parsifal e IdRef, che non hanno una voce corrispondente nella Wikipedia in italiano.

²⁰ <https://it.wikipedia.org/wiki/Giorgio_Levi_Della_Vida>.

²¹ <https://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=183503>.

identificativo SHARE Catalogue di un autore P3987	 183503 ...		 0 riferimenti
	 550090 ...		 0 riferimenti
	motivo di classificazione sconsigliata	reindirizzamento / Q45403344	
	 550091 ...		 0 riferimenti
	motivo di classificazione sconsigliata	reindirizzamento / Q45403344	

Figura 3. I valori inseriti in Wikidata per l'item Giorgio Levi Della Vida (<<https://www.wikidata.org/wiki/Q387029>>), comprensivi del qualificatore relativo alla deprecazione.

modifiche nei record, per mantenerne traccia in Wikidata, si è adottata la strategia di evitare la cancellazione e segnalare il reindirizzamento tramite il qualificatore relativo.

Questo segmento di lavoro ha fatto emergere una criticità presente nella prima versione di SHARE Catalogue, risolta nella nuova infrastruttura del progetto: se l'intervento di normalizzazione non avviene nei record bibliografici ma in quelli di authority, ciò che si genera nell'aggiornamento periodico di SHARE Catalogue non è il reindirizzamento ma un cluster orfano di collegamenti, che non è ricercabile ma di fatto non viene invalidato.²²

Gli esempi descritti ci hanno consentito, in corso di sperimentazione, di scoprire vantaggi inattesi e ci permettono ora di guardare oltre il processo di *inter-linking*, con l'obiettivo di utilizzare nel loro pieno potenziale i metadati prodotti e le collezioni rese accessibili all'interno del contesto cooperativo che SHARE rappresenta. Ciò che è stato rapidamente delineato ha anche lo scopo di mostrare alcune delle possibilità offerte dall'interazione di progetti istituzionali con Wikidata, una interazione non focalizzata solo sul servirsi di Wikidata come una delle fonti di riferimento, ma che ambisca a utilizzarne la densità informativa partecipando al processo che la accresce collaborativamente.

²² Si veda: Klotz, Henri Pierre <https://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=299380> e la forma del record di authority eliminato Klotz, H. Pierre <https://catalogo.share-cat.unina.it/sharecat/searchNames?n_cluster_id=432457>.

4. Verso SHARE Catalogue 2.0

4.1 L'evoluzione della convenzione interuniversitaria nel contesto della Share Family

SHARE Catalogue si inserisce nell'ecosistema costituito dalla *Share Family*²³, ambiente che riunisce diversi sistemi collaborativi in LOD supportati dalla medesima tecnologia, sviluppata nel progetto *Share-VDE* e dotata di una scalabilità che le consente di modulare le proprie prestazioni a seconda del contesto di utilizzo. In questo ambito, SHARE Catalogue partecipa delle novità e dell'evoluzione dell'intera iniziativa e trae vantaggio dalla sua particolare flessibilità tecnologica adottata, che ben si adatta anche a realtà consortili, quale quella della rete di biblioteche universitarie del Sud Italia. Questa tecnologia, infatti, si integra con le funzionalità dei differenti LMS utilizzati, consentendo in prospettiva non solo forme di catalogazione collaborativa, ma anche un controllo più dettagliato e puntuale dei titoli bibliografici e degli *authority file* in maniera automatizzata, fino all'arricchimento dei dati con gli URI riconciliati da fonti esterne. Si realizza in questo modo un circolo virtuoso che comporta il controllo e il miglioramento qualitativo dei dati locali, in una triangolazione che presenta indubbi vantaggi a fronte di alcune criticità che sono attualmente all'attenzione del gruppo tecnico (vedi 4.4).

È proprio l'evoluzione dell'infrastruttura a consentire a SHARE Catalogue l'inserimento tra gli altri *tenant* della Share Family con la propria CKB e il proprio *discovery portal*, punto di accesso ai dati di tutte le istituzioni convenzionate, dotato di un'interfaccia di ricerca *entity based*, fedele all'approccio ontologico di FRBR e organizzata secondo RDF (Resource Description Framework) grazie all'adozione del modello BIBFRAME.

Il passaggio dall'attuale alla nuova LOD Platform²⁴ si ricollega all'annuncio ufficiale nell'*Executive Summary* di Share-VDE 2.0 del 2022²⁵ e alle recenti attività del suo team di sviluppo. A settembre 2021 viene lanciata la prima release della nuova versione di Share-VDE, con il nuovo portale web, che ospita un sottoinsieme dei dati delle istituzioni. Confrontando le due versioni dell'*Executive Summary*, approvate dall'Advisory Council²⁶ rispettivamente a settembre 2021 e a dicembre 2022, è facile evidenziare in quale direzione vada l'evoluzione degli strumenti di gestione per gli operatori delle istituzioni aderenti. Dalla iniziale fase di sviluppo, i partecipanti si muovono nel 2022 verso la concreta identificazione di un set di *tool* avanzati per la gestione delle entità e di un *linked data editor* dedicato (JCricket) per l'interazione diretta all'interno della piattaforma, mantenendo l'obiettivo di consentire alle biblioteche di sfruttare l'ambiente Share-VDE continuando a lavorare in locale con i propri formati descrittivi.

²³ Per un'analisi degli altri ambienti collaborativi che costituiscono la Share Family, vedi <https://wiki.share-vde.org/wiki/ShareFamily:Main_Page>.

²⁴ <https://wiki.share-vde.org/w/images/5/54/share_components_EN.pdf>.

²⁵ <https://wiki.share-vde.org/w/images/4/43/Share-VDE_Executive_Summary_2022-12-07.pdf>.

²⁶ Share-VDE Advisory Council, <[https://wiki.share-vde.org/wiki/ShareVDE:Members/Share-VDE_working_groups#SVDE_Advisory_Council_\(AC\)](https://wiki.share-vde.org/wiki/ShareVDE:Members/Share-VDE_working_groups#SVDE_Advisory_Council_(AC))>.

La versione beta di Share-VDE 2.0 presenta nuove funzionalità²⁷, tra le quali:

- l'aggregazione dei dati delle istituzioni in istanza condivisa, con collegamento alla risorsa negli OPAC locali;
- portali che supportano funzionalità specifiche per le singole *skin*²⁸;
- una ricerca semplice *Opac like*;
- una ricerca avanzata per ogni tipologia di agenti, opere e pubblicazioni²⁹;
- pagine delle entità integrate con immagini e descrizioni dai progetti della famiglia Wikimedia;
- l'attivazione di una opzione per la visualizzazione estesa degli identificatori, interni ed esterni, nella propria interfaccia e la possibilità di esportare i risultati in una molteplicità di formati;
- ottimizzazione delle funzionalità di accessibilità con personalizzazione dei caratteri tipografici, dei temi, dei contrasti visivi e del layout.

Con questa versione, si assiste alla nascita di un nuovo *Entity Discovery Portal* e di una nuova infrastruttura di back-end per i linked data, basati su una tecnologia che è condivisa con gli altri portali della Share Family.

I record bibliografici delle varie istituzioni partecipanti a Share-VDE si presentano generalmente in formato MARC 21, la cui transizione all'ambiente dei dati collegati e del *Semantic Web* viene realizzata attraverso BIBFRAME, che è in grado di superare i limiti del formato MARC. In particolare BIBFRAME viene utilizzato in Share-VDE opportunamente integrato con estensioni costituite da entità aggiuntive che ne disegnano la specifica ontologia. Nel paragrafo successivo vedremo come è stata realizzata attraverso il confronto con IFLA LRM e la creazione dell'entità *svde:Opus*, per poi soffermarci sulla mappatura di UNIMARC, necessaria considerando che i dati catalografici delle istituzioni aderenti a SHARE Catalogue seguono, nella maggior parte dei casi, questo formato.

4.2 Come funziona Share-VDE: l'ontologia come ponte tra BIBFRAME, IFLA LRM e RDA

L'adesione alla comunità di Share-VDE da parte di prestigiose istituzioni europee, tra cui la National Library of Norway, National Library of Finland e della British Library, che vantano profonde radici nel contesto culturale europeo e hanno stretti legami con l'IFLA, ha sollevato importanti questioni sulla compatibilità tra il modello BIBFRAME e l'IFLA Library Reference Model (IFLA 2020).

IFLA LRM, ereditando i livelli delle varie declinazioni della famiglia FR, sviluppa, rispetto a BIBFRAME, un modello concettuale più complesso, ripropo-

²⁷ <<https://www.svde.org/about/release-log>>.

²⁸ Ad esempio: <<https://smithsonian.svde.org>> e <<https://penn.svde.org>>.

²⁹ Ad esempio, è possibile cercare selezionando come Agente una *Conference* la cui *location* sia Turin (Italy), il cui *meeting year* sia compreso tra 2000 e 2005, con, a oggi, 23 risultati.

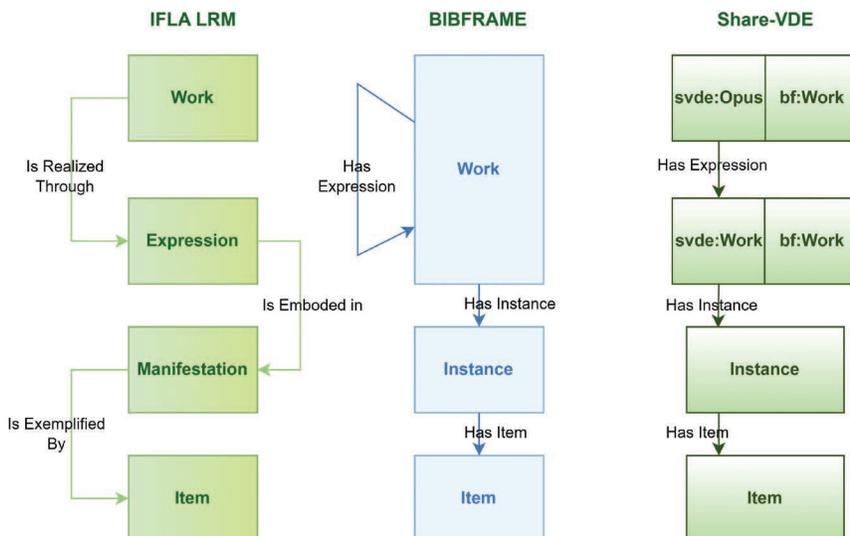


Figura 4. Comparazione dei tre modelli: IFLA LRM, BIBFRAME e Share-VDE.

nendo come entità principali quelle di FRBR: Work, Expression, Manifestation e Item³⁰.

Nel modello BIBFRAME l'entità Work rappresenta il contenuto intellettuale di una risorsa e nello stesso tempo dà indicazione dell'opera creativa che era nella mente del suo autore, poi espressa in tanti modi differenti. In sostanza i Work di BIBFRAME sono tutti sulla medesima linea orizzontale: non è possibile distinguere un'opera originaria dalla sua traduzione, soprattutto nei casi di opere con storie editoriali molto complesse, in cui si perde traccia della fonte ispiratrice e non si riesce a capire se l'opera tradotta è stata prodotta a partire dall'originale o da un'altra traduzione.

Per questo motivo il gruppo Sapiientia Entity Identification Working Group (SEI)³¹, che si occupa di tutti i temi relativi al modellamento dei dati all'interno della comunità Share-VDE, dopo una serie di riflessioni avviate a partire dal 2019, ha introdotto l'entità *svde:Opus*, per raccogliere sotto di sé tutte le diverse espressioni relazionate (qualcosa, dunque, di molto vicino al concetto di Work di FRBR e di IFLA LRM), rappresentando nello stesso tempo quell'elemento apicale che manca in BIBFRAME (cfr. Possemato 2022).

La risorsa di tipo *svde:Opus* è anche un *bf:Work* (quindi è un tipo di *bf:Work*) così come la risorsa *svde:Work* è anche un *bf:Work*.

³⁰ Per approfondimenti su IFLA LRM, vedi IFLA 2020.

³¹ <[https://wiki.share-vde.org/wiki/ShareVDE:Members/Share-VDE_working_groups#SVDE-AC_Sapiientia_Entity_Identification_working_group_\(SEI\)](https://wiki.share-vde.org/wiki/ShareVDE:Members/Share-VDE_working_groups#SVDE-AC_Sapiientia_Entity_Identification_working_group_(SEI))>.

Le due entità modellate per Share-VDE diventano tipi di Work BIBFRAME, essendo così perfettamente compatibili sia con BIBFRAME che con le entità Work ed Expression di IFLA LRM.

Questa estensione di BIBFRAME, adottata nell'ontologia di Share-VDE, funge da tramite con IFLA LRM e RDA e da struttura portante rispetto alle funzionalità dell'intera Share Family, garantendo una corrispondenza senza la necessaria mappatura diretta tra le *core classes* RDA e BIBFRAME (cfr. Hahn e Possemato 2023).

4.3 La mappatura UNIMARC-BIBFRAME come supporto alla transizione verso 2.0

Affinché l'impianto concettuale e la corrispondenza tra le diverse ontologie si tramutino in una più evoluta esposizione dei dati bibliografici in funzione di una reale erogazione di servizi agli utenti, per SHARE Catalogue si sono rese necessarie diverse attività tecniche, tra cui la mappatura UNIMARC-BIBFRAME nella versione 2.0, finalizzata a supportare la transizione alla nuova LOD Platform e a garantire l'allineamento a BIBFRAME dal momento che i metadati bibliografici provenienti dalle diverse istituzioni italiane consorziate si presentano in gran parte in formato UNIMARC.

La mappatura, in particolare, è stata il punto di partenza delle attività e consiste in una rappresentazione delle specifiche di conversione in formato tabellare, realizzata da un gruppo tecnico operativo, costituito nel 2022 e composto da membri appartenenti alle istituzioni della convenzione interuniversitaria e da analisti dell'azienda @Cult³².

Un punto di riferimento fondamentale, in fase di avvio del progetto, è stato il documento realizzato dalla Library of Congress nel 2001 per la conversione dei record bibliografici da UNIMARC nel formato MARC 21³³; sono stati affiancati a questo documento, inoltre, gli aggiornamenti UNIMARC Bibliographic 3rd ed. 2012-2022³⁴. In corso d'opera il gruppo tecnico ha effettuato una comparazione non soltanto astratta e formale ma ha tenuto conto anche del concreto uso di UNIMARC nei cataloghi partecipanti per riuscire a risolvere alcune difficoltà riconducibili al diverso impianto concettuale e strutturale dei due formati (vedi par. 4.4).

In tale mappatura, MARC 21 ha avuto la funzione di strumento di conversione da UNIMARC a BIBFRAME. Perché i dati bibliografici potessero transitare efficacemente in BIBFRAME 2.0 attraverso le specifiche di conversione di MARC 21 pubblicate dalla Library of Congress³⁵, era necessaria una mappatura

³² Vedi Gruppi di lavoro, SHARE Catalogue (LOD), <https://www.sharecampus.unina.it/?page_id=66>.

³³ UNIMARC to MARC 21 Conversion Specifications, Version 3.0 (August 2001), <<https://www.loc.gov/marc/unimarc2tomarc21.html>>.

³⁴ <<https://www.ifla.org/g/unimarc-rg/unimarc-bibliographic-3rd-edition-with-updates>>.

³⁵ Library of Congress, *MARC 21 to BIBFRAME 2.0 Conversion Specifications*, <<https://www.loc.gov/bibframe/mtbf>>.

che tenesse conto delle peculiarità delle istituzioni aderenti a SHARE Catalogue e, non ultimo, delle prassi catalografiche locali.

Parallelamente alla mappatura, il gruppo si è dedicato anche ad altre due attività, cioè l'analisi e la correzione delle duplicazioni di cluster in SHARE Catalogue 1.0 rilevate tramite i report generati da Wikidata (vedi par. 3) e la modellazione della mappatura stessa in forma strutturata per pubblicarla, in formato aperto, in un'istanza Wikibase.

L'attività di analisi e correzione, in particolare, si è rivelata come un utile spazio di collaborazione per individuare alcune discrepanze descrittive ricorrenti presenti nei diversi cataloghi. La trasposizione dei risultati della mappatura in Wikibase.Cloud costituisce invece un'attività successiva, ancora *in itinere*, che si colloca coerentemente nella politica di disponibilità e riusabilità che contraddistingue le attività inserite nel quadro della convenzione interuniversitaria.

4.4 Stato dell'arte e criticità nella mappatura

Esistono sicuramente differenze sostanziali tra MARC 21 e UNIMARC, quest'ultimo integrato con gli aggiornamenti della terza edizione, che rendono in alcuni casi critica la ricerca delle corrispondenze.

Il diverso impianto concettuale sotteso ai due formati non sempre consente di individuare una corrispondenza univoca, spesso a causa di maggiori o minori livelli di profondità nella suddivisione in campi, sottocampi, indicatori o della diversa natura, testuale o codificata, delle informazioni cui essi sono dedicati. Il caso emblematico è quello in cui un campo UNIMARC contenga valori più generali e non strutturati rispetto a quelli del corrispondente campo MARC 21.

Un esempio è costituito dai campi 324 di UNIMARC³⁶ e 534 di MARC 21³⁷. Entrambi rappresentano la “*Original version note*” e le rispettive specifiche sono sostanzialmente sovrapponibili. Nonostante ciò, non è possibile affermare la loro esatta corrispondenza.

UNIMARC 324, infatti, ha un unico sottocampo (*a*) con un contenuto non strutturato (*text*) che si presenterebbe come nel seguente esempio:

324 ##\$a*Facsimile reprint of: 1797 ed. originally published as 'The complaint and the consolation, or, Night thoughts', London : Richard Edwards, 1797*³⁸.

Il campo MARC 21 534, invece, presenta ben 17 sottocampi in cui vengono distribuite in maniera molto granulare le informazioni che, in UNIMARC, sarebbero inserite tutte nel medesimo sottocampo. In questo caso, il gruppo tecnico ha optato per una corrispondenza più generica con il campo MARC 21 500

³⁶ Per una descrizione consultabile liberamente vedi IFLA 2023, in cui il tag è riportato come aggiornato al 2005.

³⁷ <<https://www.loc.gov/marc/bibliographic/bd534.html>>.

³⁸ Vedi EX 1 (IFLA 2023).

(*General note*), per evitare, in output, una commistione di dati qualitativamente non omogenei.

Altri aspetti critici possono essere ravvisati in alcuni casi di seguito:

- la diversa strutturazione dei campi per le informazioni codificate (i rispettivi blocchi 1XX e 00X), che in UNIMARC prevede l'uso di più campi differenziati per tipologia di materiale, mentre in MARC 21 si risolve in larga parte nei campi 006, 007 e 008, di cui l'ultimo particolarmente complesso;
- la diversa gestione delle responsabilità primarie (principale e secondaria), tutte nel blocco del 7XX in UNIMARC, distinte tra quello del 1XX e del 7XX in MARC 21;
- la diversa strutturazione dei titoli uniformi nei campi 130 e 240 di MARC 21 rispetto ai 500 e 576 UNIMARC.

Si tratta, a scopo esemplificativo, solo di alcune delle criticità che il gruppo si è trovato ad affrontare per potersi basare, con una logica omogenea, sulle specifiche di conversione³⁹.

Un altro aspetto critico ha riguardato la disomogeneità relativa agli authority file nelle varie istituzioni. Considerando la particolare attenzione che la nuova infrastruttura tecnologica di SHARE Catalogue presta all'aspetto dell'arricchimento dei record di authority, l'attuale mancanza di una politica comune costituisce, se non un ostacolo, un sicuro rallentamento nella realizzazione di questo aspetto del progetto. Non si tratta soltanto del diverso livello quantitativo dei record locali, ma soprattutto di una notevole differenziazione circa le scelte 'normative' nella loro creazione.

5. Il wiki della mappatura su Wikibase.Cloud

Terminata la mappatura, il gruppo tecnico si è interrogato su come renderla disponibile per un uso diffuso nella comunità professionale. Sulla base dei principi di apertura e interoperabilità, si è cercato di tradurre il lavoro effettuato in una forma potenzialmente riutilizzabile e modificabile. Per fare questo è sembrato utile adottare una strategia che trasformasse la mappatura dalla forma tabellare in cui è stata inizialmente elaborata in un formato strutturato, senza perdita di informazioni rilevanti, con dichiarazioni in forma di triple in cui i soggetti potessero essere i tag e i sottocampi di UNIMARC. L'esperienza maturata con Wikidata ha portato a considerare l'uso di Wikibase, estensione semantica di MediaWiki che ne rappresenta il motore, come la scelta più adeguata a questo scopo.

Wikibase è un software open source progettato per immagazzinare e gestire dati strutturati in un repository centralizzato, la cui realizzazione è pensata in funzione di una *knowledge base* collaborativa sul modello di Wikidata. Il *data model*

³⁹ Library of Congress, *MARC 21 to BIBFRAME 2.0 Conversion Specifications*, <<https://www.loc.gov/bibframe/mtbf/>>.

di Wikibase gestisce in maniera uniforme tutte le entità, facendo una distinzione tra quelle descritte o messe in relazione (*Item*) e le specifiche caratteristiche espresse in forma di predicato (*Property*) dichiarate per le entità in forma di stringhe, altri *Item*, URL ecc.⁴⁰.

Item e *Property* contengono le dichiarazioni (*Statement*), formulate a seconda del tipo di dato scelto per caratterizzare i predicati e che possono essere integrate con qualificatori e fonti, utili a fornire informazioni aggiuntive relative alle dichiarazioni. Alle entità è assegnato un identificatore univoco e progressivo, generato in automatico da Wikibase e preceduto da una “Q” per gli *Item* e da una “P” per le *Property*.

Item e *Property* possono essere creati manualmente, assegnando a ciascuno una label, espressa a partire dalle lingue riconosciute nei progetti Wikimedia, accompagnata da una descrizione ed eventualmente degli alias, visualizzati come “Also known as”. È inoltre possibile creare gli *Item* utilizzando tool sviluppati dalla comunità degli utilizzatori, tra i quali Cradle⁴¹ e QuickStatements⁴².

Il Wikibase *data model* si presenta come un metamodello di astrazione dei dati (Lepore 2017-18, 4) in grado di fornire uno schema semantico non vincolato a un’ontologia di riferimento, rendendo possibile l’integrazione di schemi di metadati differenti. Un modello così strutturato apre alla possibilità di esporre i propri dati nel web semantico per valorizzarli e riutilizzarli, adattandosi alle esigenze descrittive delle singole istituzioni, partendo naturalmente da una ridefinizione del *workflow* tradizionale per la creazione di metadati comprensibili per i motori di ricerca.

Una sperimentazione d’uso di Wikibase è stata elaborata dalla Biblioteca Nazionale Centrale di Firenze (cfr. Bergamin e Bacchi 2018, 35-74), a partire dall’idea di una condivisione dei dati bibliografici procedendo a un’esportazione in LOD dei metadati di un campione di record in formato UNIMARC, limitatamente ad alcuni campi del blocco 2 (200, 210, 215), oltre ai campi 500 e 700 (Lepore 2017-18, 3-4).

Nell’esperienza fiorentina, ogni record descrittivo oggetto di conversione è tradotto come *Item* di Wikibase individuato da un URI, mentre sono espressi come alias gli identificatori univoci quali il BID SBN o il record number BNI. I campi e sottocampi UNIMARC corrispondono a proprietà che si qualificano come predicato per l’oggetto ‘dato bibliografico’ rappresentato come *Item*; i sottocampi sono poi inclusi come sottoclassi, tramite la proprietà ‘*superproprietà di*’, nel tag relativo. È presente, inoltre, una mappatura per tag e sottocampi con le

⁴⁰ Per un elenco di tutte le tipologie di dato utilizzabili, vedi ad es. <<https://unimarc2bibframe.wikibase.cloud/wiki/Special:ListDatatypes>>.

⁴¹ Cradle, <https://www.mediawiki.org/wiki/Wikibase/Creating_and_deleting_data/en#Cradle>. Il repository è disponibile a partire da <<https://github.com/magnusmanske/cradle>>.

⁴² QuickStatements, <https://www.mediawiki.org/wiki/Wikibase/Creating_and_deleting_data/en#QuickStatements>. Il repository è disponibile a partire da <<https://github.com/magnusmanske/quickstatements>>.

principali ontologie (schema.org, Dublin Core, BIBFRAME, ecc.), grazie al predicato ‘*corrisponde strettamente*’, dichiarate tramite URI (Storti 2019).

Il progetto della BNCF non è l’unico ad aver utilizzato Wikibase per la traduzione in linked data dei dati descrittivi delle risorse bibliografiche. Project Passage, sperimentazione portata avanti da un nucleo di ricerca di OCLC con un pilot durato dieci mesi, ha inteso mostrare le potenzialità trasformative dei linked data nella conversione dei record MARC, troppo limitati da un punto di vista della comprensibilità per il Semantic Web, fornendo una *sandbox* nella quale bibliotecari provenienti da sedici istituzioni statunitensi hanno potuto sperimentare con la creazione di linked data per descrivere risorse, articolando le informazioni senza necessariamente dover partire da una solida conoscenza tecnica sui linked data, con un flusso di lavoro particolarmente interessante nella creazione di *Item* per voci di authority (cfr. Godby et al. 2019).

Oltre a poter installare e utilizzare Wikibase localmente e su servizi di hosting dedicati, come nei due esempi precedenti, Wikimedia Deutschland (WMDE) fornisce una piattaforma cloud, attualmente in *open beta*, in modo da facilitare la creazione e lo sviluppo di database collaborativi in LOD. Wikibase.Cloud⁴³ è stata lanciata nel 2022 e il primo passo del suo ampliamento ha riguardato la migrazione, completata a giugno del 2022, degli utenti da WBStack.com in uso dal 2019. Attualmente la piattaforma conta 879 wiki⁴⁴ e si presenta come un utilissimo strumento operativo, in virtù della presenza di tool di modifica preinstallati e dell’assenza di processi e costi per manutenzione e aggiornamento per l’utente finale. Il gruppo tecnico ha perciò reputato vantaggiosa questa soluzione per trasferire la mappatura, scegliendo inoltre di rilasciare i dati in CC0 1.0, salvo dove diversamente espresso.

In *SHARE Catalogue Mapping Knowledge Base*⁴⁵, i tag e i sottocampi sono *Item*, caratterizzati da una label formulata in almeno tre lingue, cui per i tag si accompagnano le descrizioni⁴⁶. In tutti gli *Item* sono inseriti anche gli alias, particolarmente significativi per i sottocampi perché utili a disambiguarli rispetto a tutti gli altri che, nel formato, presentano la stessa denominazione. Per ogni *Item* le dichiarazioni definiscono la posizione ‘gerarchica’, in riferimento ad esempio al blocco o al campo di inclusione, ma anche l’occorrenza, la ripetibilità e il riferimento all’aggiornamento utilizzato. Inoltre sono state create proprietà utilizzabili come qualificatori per esprimere particolari condizioni d’uso (ad esempio, un sottocampo con occorrenza *Mandatory* quando in un record bibliografico devono essere descritti specifici tipi di risorse). Gli *Item* saranno poi completati, oltre che con fonti per alcune dichiarazioni come l’URL dell’aggiornamento uti-

⁴³ <<https://www.mediawiki.org/wiki/Wikibase/Wikibase.cloud>>.

⁴⁴ Dato aggiornato al 9 febbraio 2024, vedi <<https://meta.wikimedia.org/wiki/Wikibase/Wikibase.cloud>>.

⁴⁵ <<https://unimarc2bibframe.wikibase.cloud>>.

⁴⁶ Label e descrizioni sono tratte, generalmente, per l’inglese da UNIMARC Bibliographic (3rd ed.) Updates e per il francese da <<https://www.transition-bibliographique.fr/unimarc/manuel-unimarc-format-bibliographique>>.

lizzato, con dichiarazioni per proprietà grazie alle quali si possa mettere in relazione il valore BIBFRAME mappato, in forma di stringa e di URI; altre proprietà serviranno invece da connessione, a livello di tag, verso pagine descrittive della mappatura all'interno del wiki.

Attualmente il progetto è in fase di arricchimento e il lavoro è concentrato sulla creazione, principalmente tramite QuickStatements, dei diversi tag e sottocampi di UNIMARC e delle proprietà utili a completare il modello realizzato dal gruppo.

6. Conclusioni

L'evoluzione di SHARE Catalogue è l'esito di una strategia di transizione che dal formato MARC, densamente descrittivo ma privo della granularità e della semantica necessaria a renderlo comprensibile nel web, si muove verso una tecnologia che lo rende pienamente utilizzabile nei molteplici ambienti che nel web dei dati non sono definiti a priori. Una transizione reale quindi verso i linked data, nella loro declinazione open, frutto della cooperazione nata tra i bibliotecari e le istituzioni che hanno concepito Share-VDE, sviluppata nella più articolata Share Family e i cui frutti continuano a essere aperti a ulteriori 'contaminazioni'.

La Share Family comprende istituzioni di varie dimensioni e prestigio, che spaziano dalle grandi realtà per estensione e qualità dei servizi offerti, fino a realtà più piccole ed estremamente specializzate⁴⁷, che concorrono con il loro unico contributo alla vitalità della comunità sviluppata attorno a questa iniziativa. Le istituzioni della Share Family partecipano, pertanto, con le proprie peculiarità, rispondendo a esigenze proprie, ma contemporaneamente arricchiscono e stimolano nuove evoluzioni nei progetti fratelli.

Pur differenziandosi per dimensioni, risorse ed esperienze apportate, tutte queste istituzioni condividono i principi fondamentali di condivisione, riutilizzo, arricchimento e una maggiore esposizione delle risorse nel web. Affinché tutto ciò avvenga in modo efficace, è essenziale stabilire alcune pratiche comuni riguardanti l'interoperabilità e il riuso dei dati: l'apertura delle fonti rappresenta il punto di partenza, accompagnata da un'attenzione particolare alla qualità dei dati e dall'adozione di standard e protocolli che agevolino la condivisione⁴⁸.

⁴⁷ Un esempio è *LILLIT, Libri illustrati italiani*, portale in linked open data di un catalogo delle edizioni a stampa prodotte tra il XVI e il XVII secolo, che contengono illustrazioni o incisioni importanti dal punto di vista dell'illustrazione del libro. La base dati di partenza del piccolo gruppo di lavoro, responsabile del lavoro di revisione, analisi e studio delle edizioni descritte, è costituita da circa 700.000 record bibliografici, in formato UNIMARC, messi a disposizione dall'ICCU.

⁴⁸ In questo senso assume un significato di politica generale dell'iniziativa la policy adottata dai partecipanti a Share-VDE, vedi *Share-VDE Open Metadata Policy* (Approved by the Share-VDE Advisory Council on March 15th 2023), <https://wiki.share-vde.org/w/images/7/73/Share-VDE_Open_Metadata_Policy_2023-03-15.pdf>.

È il caso di *Parsifal*, per il quale le biblioteche partecipanti non solo hanno definito delle linee guida che supportino i catalogatori nella normalizzazione dei punti di accesso e nella gestione del flusso di lavoro in generale, ma hanno enfatizzato il concetto di riuso con la realizzazione dell'Authority File Centralizzato (di seguito AFC), generato dalla CKB di URBE.

I record di authority dell'AFC, derivati dalla sinergia e dalla condivisione nella rete, possono essere impiegati direttamente a livello locale, così da fornire alle istituzioni stesse un file di authority che in precedenza non avrebbero potuto produrre, considerata la persistente limitatezza di risorse vissuta quotidianamente dalle biblioteche. In alternativa, questi record possono essere reintegrati nel sistema di authority locale, arricchiti di attributi e URI pertinenti, integrazione questa che, non solo agevola l'accesso ai dati, ma porta anche a un notevole miglioramento globale della qualità delle informazioni disponibili.

Di conseguenza, oltre all'esposizione dei dati e alla condivisione in funzione di riuso, l'interoperabilità, intesa come la capacità di connettere collezioni anche estremamente diverse tra loro, con reciproco arricchimento (uno dei concetti chiave dei linked data), apporta un valore significativo al web. I dati esposti basandosi su questo principio, consentono agli utenti di accedere a informazioni più complete, approfondite, migliorate nella loro accuratezza, in un circuito tecnologico tra locale e centrale che valorizza la rappresentazione dei metadati di un patrimonio unico nel suo genere.

Questa ottima pratica può sicuramente funzionare da stimolo anche per SHARE Catalogue, provocando un impatto positivo proprio su una delle criticità maggiori individuata dal gruppo tecnico, ovvero il coordinamento delle diverse 'abitudini' catalografiche relativamente all'authority.

Da parte sua, anche SHARE Catalogue ha intrapreso una complessa evoluzione tecnologica consapevole di poter dare un apporto significativo: la mappatura ne è un esempio, che richiede costanti affinamenti e adattamenti ma che risponde alla volontà di confrontarsi in maniera problematica e attiva con la reattiva e sfaccettata comunità che anima la Share Family.

Non solo la mappatura però: due ulteriori percorsi riguarderanno l'integrazione di fonti non sufficientemente studiate nella prima versione del progetto, come IdRef⁴⁹ e un lavoro, in gran parte da realizzare, sulla clusterizzazione dei soggetti, per il quale va ampliata l'inclusione di tesauri, poco o per nulla valorizzati nella pratica italiana, che costituiscono il composito mondo della *Critical Librarianship*⁵⁰.

Impegnarsi in una costante evoluzione (dei modelli descrittivi, dei protocolli, delle tecnologie, dei vocabolari) con l'attenzione puntata alla qualità, veridicità, verificabilità e disseminazione dei dati, rappresenta oggi un investimento cru-

⁴⁹ <<https://www.idref.fr>>.

⁵⁰ Tra le tante risorse sull'argomento, vedi ad es. <https://www.ala.org/acrl/publications/keeping_up_with/critlib>, <https://academicworks.cuny.edu/gc_pubs/537> e <<https://subjectguides.library.american.edu/c.php?g=1025915&p=7749829>>.

ciali per il futuro delle risorse culturali nel web. Evoluzione che, come dimostra la Share Family, per essere sostenibile a lungo termine non può fare a meno di azioni coordinate e collaborative, aperte a tutti i soggetti produttori di nuova conoscenza, con un occhio al ruolo che biblioteche, archivi e musei hanno storicamente rappresentato nell'organizzazione delle informazioni.

Riferimenti bibliografici

- Bergamin, Giovanni, e Cristian Bacchi. 2018. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data." *JLIS.it*, IX, 3: 35-74. <<https://www.jlis.it/index.php/jlis/article/view/96>>
- Erxleben, Fredo, Michael Günter, Markus Krötzsch, Julian Mendez, Denny Vrandečić. 2014. "Introducing Wikidata to the Linked Data Web." In *The Semantic Web – ISWC 2014*, a cura di Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig Knoblock, Denny Vrandečić, Paul Groth, Natasha Noy, Krzysztof Janowicz, Carole Goble. 50-65. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-11964-9_4>
- Forziati, Claudio, e Valeria Lo Castro. 2018. "La connessione tra i dati delle biblioteche e il coinvolgimento della comunità: il progetto SHARE Catalogue-Wikidata." *JLIS.it*, IX, 3: 109-20. <<https://doi.org/10.4403/JLIS.IT-12488>>
- Gazzarini, Andrea. 2015. "Aliada un framework di pubblicazione di linked open data." In *Linked data per biblioteche, archivi e musei*, a cura di Mauro Guerrini e Tiziana Possemato. Milano: Editrice Bibliografica.
- Godby, Jean, Karen Smith-Yoshimura, Bruce Washburn, Kalan Knudson Davis, Karen Detling, Christine Fernsebner Eslao, Steven Folsom, Xiaoli Li, Marc McGee, Karen Miller, Honor Moody, Craig Thomas, Holly Tomren. 2019. "Creating Library Linked Data with Wikibase. Lessons Learned from Project Passage." <<https://www.oclc.org/content/dam/research/publications/2019/oclcresearch-creating-library-linked-data-with-wikibase-project-passage.pdf>>
- Guerrini, Mauro, e Tiziana Possemato. 2015. *Linked data per biblioteche, archivi e musei*. Milano: Editrice Bibliografica.
- Guerrini, Mauro. 2019. "La biblioteca integrata: nuovi modelli, nuove tecniche, alcune esperienze europee e italiane." *Biblioteche oggi Trends* V, 1: 59-66. <<https://doi.org/10.3302/2421-3810-201901-059-1>>
- Hahn, Jim, e Tiziana Possemato. 2023. "The Share-VDE ontology: a BIBFRAME extension for linked data discovery" (1.0.0-beta). Zenodo. <<https://doi.org/10.5281/zenodo.8414627>>
- IFLA. 2020. *IFLA Library Reference Model. Un modello concettuale per le informazioni bibliografiche*, a cura di Pat Riva, Patrick Le Boeuf, e Maja Žumer. Roma: ICCU. <https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712-it.pdf>
- IFLA. 2023. *UNIMARC Bibliographic Format Manual*, a cura di Gordana Mazić, Branka Badovinac, con Rosa Galvão, Olga Zhobinskaya e Saeedeh Akbari-Da-

- ryan. <<https://www.ifla.org/unimarc-updates/unimarc-bibliographic-format-manual-online-ed>>
- Lepore, Valentina. 2017-18. "Il Wikibase data model per la pubblicazione dei dati bibliografici sul web semantico. Una sperimentazione presso la Biblioteca Nazionale Centrale di Firenze." Tesi di Master di secondo livello in Archivistica e Biblioteconomia e Codicologia. Università degli studi di Firenze.
- Martinelli, Luca. 2016. "Wikidata: la soluzione wikimediana ai linked open data." *AIB studi*, LVI, 1: 75-85. <<https://doi.org/10.2426/aibstudi-11434>>
- Miller, Eric, Uche Ogbuji, Victoria Mueller, Kathy MacDougall. 2012. *Bibliographic framework as a Web of data: linked data model and supporting services*. Washington, DC: Library of Congress. <<https://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>>
- Neubert, Joachim. 2017. "Wikidata as a Linking Hub for Knowledge Organization Systems?"; in *Proceedings of the 17th European Networked Knowledge Organization Systems Workshop*. 14-25. <<https://ceur-ws.org/Vol-1937/paper2.pdf>>
- Possemato, Tiziana, e Roberto Delle Donne. 2017. "SHARE Catalogue: un'esperienza di cooperazione." *Biblioteche oggi*, XXXV (gennaio-febbraio): 21-9. <<https://doi.org/10.3302/0392-8586-201701-021-1>>
- Possemato, Tiziana, e Claudio Forziati. 2019. "Riuso, interoperabilità, influenza: la cooperazione virtuosa tra i progetti SHARE e Wikidata". In *La biblioteca che cresce: contenuti e servizi tra frammentazione e integrazione*. 228-38. Milano: Editrice Bibliografica.
- Possemato, Tiziana. 2022. "Entity Modeling: tracce di un percorso in evoluzione." *JLIS.it*, XIII, 3: 12-28, <<https://doi.org/10.36253/jlis.it-48>>
- Stinson, Alex. 2018. "Wikidata in Collections: Building a Universal Language for Connecting GLAM Catalogs." *Down the Rabbit Hole* (blog). <<https://medium.com/freely-sharing-the-sum-of-all-knowledge/wikidata-in-collections-building-a-universal-language-for-connecting-glam-catalogs-59b14a-a3214c>> 9 aprile 2018.
- Storti, Chiara. 2019. "La condivisione dei dati bibliografici legacy in ambiente LOD Wikibase" (slides). <https://commons.wikimedia.org/wiki/File:La_condivisione_dei_dati_bibliografici_legacy_in_ambiente_LOD_WikiBase.pdf>