

Carlo Bianchini - Stefano Bargioni - Camillo Carlo Pellizzari di San Girolamo*

*Le voci di autorità dei nomi di persona in SBN e Alfabetica:
problemi e prospettive*

1. Introduzione

Il lavoro di autorità (o *authority work*) è un'attività centrale e cruciale nel processo di catalogazione e, di conseguenza, nel flusso dei dati e delle informazioni necessarie per il corretto funzionamento di una biblioteca in tutte le sue attività: informazione bibliografica e reference, progettazione delle collezioni, acquisizioni, gestione e mantenimento delle collezioni.¹

Essa consiste nel «processo grazie al quale viene assicurata l'omogeneità formale dei punti di accesso» di un catalogo (Petrucciani - Turbanti 2021, p. 265), laddove con punto d'accesso si intende un «nome, termine, codice, ecc. mediante il quale i dati bibliografici o di autorità sono ricercati e identificati» (IFLA Cataloguing Section

* Il testo è stato scritto in totale collaborazione e condivisione: tuttavia vanno ascritti a Stefano Bargioni i paragrafi 2 e Appendice, a Carlo Bianchini i paragrafi 1 e 5 e a Camillo Pellizzari i paragrafi 3 e 4.

¹ Per una panoramica nazionale e internazionale sull'authority control, si vedano gli atti del convegno di Firenze Guerrini 2003.

e IFLA Meetings of Experts on an International Cataloguing Code 2017, p. 17). Dal momento che i lettori cercano i dati nelle forme che conoscono – che possono essere tanto varie tra loro quanto lo sono i lettori – per assicurare che qualunque sia la forma utilizzata dal lettore il catalogo sia in grado di restituire tutte le informazioni rilevanti (in forma di registrazione bibliografica o di autorità) è necessario il controllo di autorità (o *authority control*). Questo processo, svolto dal catalogatore, consiste nella verifica della corrispondenza tra le forme autorizzate e varianti dei diversi tipi di nomi associati a un'entità (per esempio, forme varianti del nome in una medesima lingua, forme varianti in lingue diverse, identificatori assegnati da agenzie diverse ecc.) e una loro aggregazione sotto forma di dati (o registrazioni o voci) di autorità (IFLA Cataloguing Section e IFLA Meetings of Experts on an International Cataloguing Code 2017, p. 14).

Il notevole costo in termini di risorse umane e finanziarie richiesto dallo svolgimento del lavoro di autorità ha spinto da tempo alla realizzazione di progetti di cooperazione su scala sempre più ampia. Il progetto del Controllo Bibliografico Universale si fonda proprio sulla possibilità di scambiare i dati bibliografici e di autorità prodotti dalle agenzie bibliografiche nazionali.² Se a livello internazionale ha svolto un ruolo indubbiamente centrale il progetto del VIAF (Virtual International Authority File),³ in Italia il punto di riferimento indiscusso è da sempre stato SBN e, in particolare, le sue voci di autorità. Da questo punto di vista, le voci di autorità create e rese disponibili da SBN attraverso il suo OPAC⁴ costituiscono un patrimonio di estremo valo-

² Sul controllo bibliografico universale si vedano in proposito gli atti del convegno “The Bibliographic Control in the Digital Ecosystem. International Conference, February 8-12, 2021”, pubblicati integralmente su JLIS.it, vol. 13, no. 1 (2022), <<http://jlis.it/index.php/jlis/issue/view/34>>.

³ <<https://viag.org>>. Si vedano inoltre a titolo esemplificativo: Tillett 2006 e Guerrini 2020, pp. 127-129.

⁴ Dal 28 febbraio disponibili all'indirizzo <<https://opac.sbn.it/voci-controllate-nomi>> e in precedenza all'indirizzo: <https://opac.sbn.it/opacsbn/opac/iccu/authority.jsp?db=solr_auth>.

re per tutte le biblioteche italiane e, in prospettiva, mondiali. Infatti, è ormai accettata l'idea che la catalogazione, «in età contemporanea, si inserisce all'interno dell'Universal Bibliographic Control (UBC), un'iniziativa importante promossa dall'IFLA e da altri soggetti a partire dagli anni Settanta del secolo scorso con lo scopo di favorire la condivisione dei dati bibliografici eliminando le ridondanze e incoraggiandone il riuso» (Guerrini 2020, p. 41).

Il VIAF e SBN sono nati in un contesto in cui il web semantico non era ancora comparso e quindi oggi si trovano a operare in uno scenario che è profondamente cambiato e nel quale i punti di riferimento e gli ambiti di interesse sono marcatamente diversi (Willer - Dunsire 2013).

Infatti, non solo è diventato più evidente il nesso tra il lavoro di autorità svolto dalle biblioteche e l'importanza dei loro dati per l'identificazione delle entità nel web semantico, ma la centralità di quei dati è diventata un aspetto chiave per il raggiungimento del controllo bibliografico universale in una forma ampiamente cooperativa e distribuita nel web semantico (Dunsire - Willer 2014).

Come osserva Alessandra Moi, «la costruzione di questo contesto è basata strettamente sul cambiamento delle tecnologie e sulla capacità delle biblioteche di utilizzarle e allinearsi ad esse [...]. L'evoluzione tecnologica che si spera [...] dovrebbe portare al 'rilascio' dei dati bibliografici al di fuori del ristretto mondo dei cataloghi» (Moi 2020, p. 2).

Si tratta di un passaggio cruciale, perché l'ampliamento del contesto di riferimento in cui i dati vengono creati e devono essere utilizzati richiede un'attenta riconsiderazione e valutazione dei dati stessi. Se ciò è vero per le registrazioni bibliografiche (Deana 2021), diventa cruciale e irrinunciabile per le registrazioni di autorità che devono svolgere principalmente una funzione *identificativa*. Per esempio, in un bacino contenente un numero ridotto di voci di autorità, può essere più che sufficiente identificare una persona con il suo nome e cognome. Ma, non appena il numero di voci di autorità si amplia, perché il bacino di dati interni cresce in modo naturale e inevitabile o perché

i dati prodotti si confrontano con un bacino di dati molto più grande (come quello necessario per il Controllo Bibliografico Universale), diventa indispensabile affiancare al solo nome e cognome qualche altro elemento distintivo, allo scopo di ottenere un punto d'accesso realmente identificativo.

ICCU ha sempre dato importanza all'interoperabilità dei dati di SBN; già nel 2013, Rossella Caffo scriveva: «Il nodo centrale dello scenario evolutivo dell'interoperabilità e dell'accessibilità nel web dei dati catalografici e dei contenuti digitali, è rappresentato dal ruolo degli standard nell'ambito della descrizione e accesso del patrimonio culturale. Uno dei principali compiti istituzionali dell'ICCU è promuovere e coordinare l'elaborazione delle normative catalografiche per le biblioteche italiane e diffondere gli standard internazionali» (Caffo 2013, p. 10).

Sul piano internazionale, l'ICCU ha promosso una integrazione tra SBN e il VIAF; sempre Rossella Caffo nel 2013 annota: «L'ICCU dal 2009 ha avviato la partecipazione al VIAF fornendo voci di autorità con *livello 97*, cioè *voci d'autorità controllate ed esaustive*. Prossimamente si invieranno anche le voci di *livello 95*, per lo più inseriti dalle Biblioteche nazionali centrali di Firenze e Roma che, pur non contenendo tutti i dati previsti per il livello 97, sono omogenei rispetto alle voci presenti nel VIAF e contengono elementi sufficienti per identificare l'autore in modo univoco. In totale saranno così presenti nel VIAF ca. 200.000 voci di autorità» (Ibidem).⁵

Per tutti i lettori che non sono avvezzi alla terminologia di SBN, ma anche per la sua centralità nell'analisi svolta in questo lavoro, va chiarito il concetto di *livello di autorità* in SBN.

Il livello di autorità è un codice numerico crescente che rappresenta il «livello di affidabilità e di completezza della catalogazione fornita e si applica a qualsiasi documento, titolo di accesso o voce di autorità» (ICCU 2017). I codici rappresentano quindi un'indicazione crescente della qualità delle voci create, secondo questo schema:

⁵ Il corsivo è stato aggiunto.

05 indica che la catalogazione è stata effettuata senza verifica sul documento; si applica per la conversione in SBN di dati ricavati dal catalogo cartaceo o comunque da precedenti catalogazioni.

51 indica che la catalogazione è stata effettuata esaminando il documento, ma fornendo i soli dati essenziali e necessari all'individuazione del documento.

71 indica che la catalogazione è stata effettuata esaminando il documento e fornendo tutti i dati e i collegamenti richiesti, sia obbligatori che facoltativi se applicabili.

90 indica che i dati obbligatori per questo livello sono tutti quelli desumibili dalla pubblicazione e da fonti esterne. Per tale livello è inoltre obbligatorio, ove necessario, il controllo dei medesimi dati sui repertori bibliografici normalmente in uso, quali p.es. le bibliografie nazionali dei vari paesi, cataloghi di grandi biblioteche, dizionari biografici, authority file nazionali ed internazionali, etc. e il collegamento con tutti i legami previsti e applicabili.

95 indica un livello di autorità che non si differenzia dal livello 90 per contenuti, ma che, essendo attribuito soltanto ai Poli delle due biblioteche nazionali centrali, di Roma e Firenze, consente di tutelare la qualità dei dati destinati a essere pubblicati sulla Bibliografia nazionale italiana.

96 e 97 riservati ai record di autorità in fase di lavorazione o già lavorati, possono essere attribuiti ai record solo da utenze personali che intervengono sulle voci di autorità non attraverso l'aggiornamento contestuale della base dati di Polo, ma agendo direttamente sul catalogo collettivo con modalità di lavoro centralizzate. (ICCU 2017).

Risulta quindi chiaro che, riferendosi ai livelli 95 e 97, Rossella Caffo esprimeva l'intenzione di inviare voci di qualità accertata, «*controlled ed exhaustive*» e, perciò, adatte allo scambio con le altre agenzie nazionali che hanno creato o collaborano al VIAF.

L'invio delle voci di autorità al VIAF (grafico 1), che fino al 2013 era proceduto relativamente a rilento, avviene in effetti in modo massiccio nel 2014 (+107.724) e negli anni successivi il numero delle voci di SBN nel VIAF continua a crescere: +7.770 nel 2015, +3.038 nel 2016, +881 nel 2017, +4175 nel 2018, +13.124 nel 2019 e +1302 fino al settembre 2020, quando era 178.348 (Bianchini - Bargioni - Pellizzari 2021, p. 10). A dicembre 2021 le voci di autorità inviate da SBN al VIAF erano arrivate a un totale di 185.185.

Quantità di voci inviate da SBN al VIAF per anno

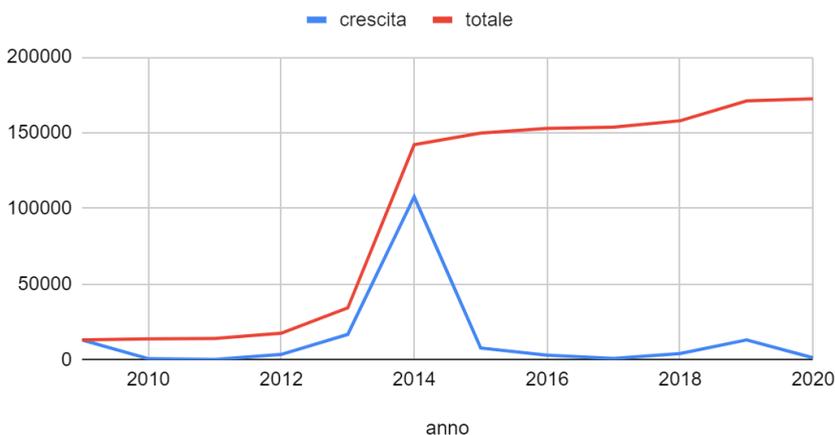


Grafico 1. Quantità di voci inviate da SBN al VIAF per anno

Il contributo di SBN al VIAF in termini quantitativi può essere confrontato con il numero delle voci di autorità di tipo persona da parte di altre agenzie nazionali; entro il settembre 2020, la Library of Congress aveva inviato 7.266.628 voci, la Deutsche Nationalbibliothek 5.677.731, il SUDOC (Système Universitaire de Documentation) 3.278.189 e l'olandese Koninklijke Bibliotheek 2.754.036. Sono

altre otto le agenzie nazionali che avevano inviato al VIAF più di un milione di voci di tipo persona (Bianchini - Bargioni - Pellizzari 2021, p. 10). L'ISNI e Wikidata (nata nel 2012), che non sono contributori ufficiali del VIAF ma 'altre fonti scelte' dal VIAF stesso, avevano contribuito rispettivamente con 8.455.814 (la fonte più presente in VIAF) e 2.148.680 voci di autorità (figura 2).

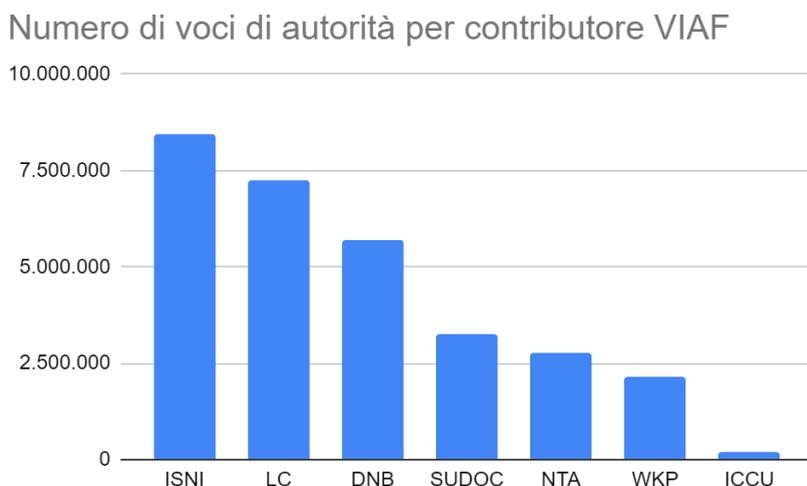


Grafico 2. Numero di voci di autorità (persona) per contributore VIAF

La sproporzione quantitativa è evidente e spinge inevitabilmente a chiedersi le ragioni di questa differenza. L'ipotesi più immediata è quella del ricorso da parte dell'ICCU a parametri molto stringenti per l'invio delle voci, corrispondenti ai livelli 95 e 97, come ha spiegato Rossella Caffo, per garantire l'invio di voci di profilo qualitativo molto alto.

Questa ipotesi plausibile apre immediatamente altre questioni: le voci di autorità pubblicate su SBN e inviate al VIAF sono davvero di altissima qualità, ovvero corrispondenti ai requisiti dei livelli 95 e 97? La qualità delle voci è davvero tale da giustificare la pubblicazione e l'invio di un numero tanto ridotto? L'uso dei «livelli di autorità» delle voci di autorità (d'ora in poi, livelli gerarchici) in SBN e la pubblica-

zione così parsimoniosa delle voci, impostati più di trent'anni fa, sono ancora pratiche vantaggiose in generale e, in particolare, nel contesto del web semantico? I livelli gerarchici delle voci di autorità e le limitazioni alla loro modifica basate sui livelli di autorità dei catalogatori sono una pratica utile e sostenibile nello scenario bibliografico attuale?

Le questioni che si aprono sono rilevanti per il passato, certo, ma riguardano molto anche il futuro di SBN, alla luce della pubblicazione e presentazione del portale *Alfabetica* da parte di ICCU il 16 dicembre 2021.⁶ Infatti, in concomitanza, se non in ragione, della pubblicazione del nuovo portale, SBN si è arricchito di 141.717 nuove voci di autorità, presumibilmente corrispondenti al livello 90.⁷ Questo suscita ulteriori domande: la qualità dei dati delle voci di autorità è migliorata o peggiorata in questo passaggio? Le voci di autorità oggi disponibili sono adeguate, in termini di quantità e qualità, alle nuove funzioni che *Alfabetica* si propone di raggiungere?

2. *Materiali e metodi*

Una risposta fondata ai quesiti posti presuppone l'accesso e l'analisi delle voci di autorità di SBN, in termini di maggiore o minore completezza dei metadati, nonché della loro struttura, ovvero della modalità con la quale i singoli dati sono registrati all'interno di ciascuna voce. SBN tuttavia non offre la possibilità di effettuare un download massiccio delle proprie voci di autorità, sebbene vi siano diversi importanti esempi di pubblicazione completa e aperta dei dati, quali la British Library, la National Diet Library of Japan o la Library of Congress.⁸

⁶ Per il portale si veda: <<https://alfabetica.it/web/alfabetica/>>. Per il lancio, <<https://www.beniculturali.it/comunicato/21880>>.

⁷ I dati relativi al livello non sono accessibili agli utenti esterni di SBN e non vengono pubblicati. Si veda *infra*, paragrafo 2.

⁸ Si vedano rispettivamente: <<https://www.bl.uk/collection-metadata/downloads>>, <https://id.ndl.go.jp/information/download_en/>, <<https://id.loc.gov/>>

Per disporre degli indispensabili dati da analizzare, nel caso di SBN sono stati adoperati due diversi metodi di recupero delle voci di autorità. Il primo è stato applicato alle voci di autorità di SBN inviate al VIAF prima della pubblicazione di Alfabetica (e recuperate a inizio dicembre 2021); il secondo alle voci di autorità pubblicate sul nuovo OPAC SBN in concomitanza con Alfabetica (e recuperate a metà dicembre 2021, dopo la pubblicazione di Alfabetica). Per questo motivo, d'ora in poi vengono utilizzate due sigle: PRE indica il dataset delle 185.132 voci di autorità di SBN recuperate nel dicembre 2021 e quindi di livello 95 e 97; POST indica il dataset delle 326.849 voci di autorità di SBN recuperate nel dicembre 2021 e quindi di livello 95, 97 e, presumibilmente, 90. Il dataset POST indica quindi lo stato di fatto delle voci di autorità presenti nel nuovo OPAC SBN a dicembre 2021.

Il dataset PRE è stato recuperato mediante uno dei dataset che il VIAF espone mensilmente alla pagina a ciò deputata.⁹

Nel dicembre 2021 è stato scaricato il file *viaf-20211206-links.txt.gz*,¹⁰ che contiene i link ai cluster delle voci ricevute dalle biblioteche partecipanti al VIAF, e pertanto anche i «VID», gli identificatori delle voci di autorità di SBN. I VID si trovano raccolti nel file *pre-alfabetica_VID.list*. In base a questa lista sono state prelevate le voci di autorità in formato UNIMARC con una procedura di *scraping*.¹¹

Il file *viaf-20211206-links.txt.gz* è stato cancellato dal VIAF un mese dopo la sua pubblicazione, come da prassi. Tuttavia, in un'ottica di open science, la copia scaricata è stata conservata e pubblicata su Zenodo insieme a tutti i file di dati e i programmi di elaborazione (script) adoperati nell'analisi e descritti in questo testo.¹²

download/>.

⁹ <<http://viaf.org/viaf/data/>>.

¹⁰ Era disponibile all'indirizzo <<http://viaf.org/viaf/data/viaf-20211206-links.txt.gz>>.

¹¹ Il web scraping (detto anche web harvesting o web data extraction) è una tecnica informatica di estrazione di dati da un sito web per mezzo di programmi software. Nel caso specifico, è stato utilizzato uno script in linguaggio Perl.

¹² Si veda Bargioni - Bianchini - Pellizzari 2022.

Il programma utilizzato per lo *scraping* è lo script *download_unimarc_pre-alphabetica.pl*, che riproduce sostanzialmente – in modo automatico e quindi molto più veloce – l’operazione manuale di salvataggio di una voce dall’OPAC SBN denominata “Scarico Unimarc”, ripetendola per ogni VID.¹³

Il dataset POST è stato recuperato attraverso un’estrazione degli identificatori direttamente dal nuovo OPAC SBN tramite lo script *estrazione_VID_post-alphabetica.pl*.¹⁴ Questo script si avvantaggia di un’utile possibilità di interrogazione offerta dal nuovo OPAC SBN e recupera dati in formato JSON, a blocchi di 10 identificatori per chiamata. Successivamente, sono stati prelevati i record UNIMARC con lo script *download_unimarc_post-alphabetica.pl*, che opera in modo molto simile allo script utilizzato per il dataset PRE, compresa la citata correzione dell’errore del formato ISO-2709.

I due dataset PRE e POST¹⁵ sono stati sottoposti ad alcune analisi quantitative e qualitative – esclusivamente in riferimento alle voci di autorità di tipo persona – allo scopo di verificare quanti dati fossero contenuti in ciascuna voce di autorità (attraverso il conteggio del numero dei campi UNIMARC), quali dati fossero contenuti (attraverso l’analisi della frequenza di specifici campi UNIMARC), quanto e quando i dati fossero stati creati, modificati o aggiornati, i poli di provenienza delle voci ecc. Le analisi hanno portato alla creazione di un gruppo di tabelle, dedicate ciascuna a una specifica analisi quanti-

¹³ Lo script include una correzione al record MARC che presenta un errore sistematico del formato ISO-2709. Senza tale correzione diventa impossibile utilizzare le librerie (p. es.: *MARC::Record.pm*) e i tool più comuni disponibili (p. es.: *yaz-marcdump*) per la lettura dei record MARC scaricati. Si tratta di un errore nella parte *leader* del record, di cui si auspica la correzione. Si veda lo standard in <<https://www.loc.gov/marc/specifications/specrecstruc.html#leader>>, paragrafo Entry map. Di questo errore non risentono le voci di autorità di opere e luoghi pubblicate sul nuovo OPAC SBN.

¹⁴ <<https://nuovo-opac.sbn.it/>>

¹⁵ I due dataset sono disponibili rispettivamente nei file *pre-alphabetica.mrc* e *post-alphabetica.mrc*, pubblicati su Zenodo.

tativa o qualitativa.

Un importante limite dell'analisi riguarda il livello di autorità delle voci di autorità prese in esame. Il livello di ciascuna voce di autorità non fa parte dei dati pubblicati nella voce stessa, ma è possibile dedurre il livello di autorità delle voci contenute nei due dataset basandosi sull'articolo di Rossella Caffo e sulla ragionevole possibilità che, con l'avvento di Alphabetica, siano state pubblicate le voci con livello di autorità 90.

Di conseguenza, lo studio si è necessariamente limitato all'analisi di voci di autorità di livello 90 o superiore, che sono quelle che maggiormente interessano dal punto di vista qualitativo. Non si sono potute invece analizzare le voci di autorità di livelli inferiori a 90, in quanto non disponibili pubblicamente.

2.1 *Tabelle di dati*

I risultati delle analisi compiute sono state raccolti e organizzati nelle seguenti tabelle:

*Tabella 1.*¹⁶ *Tipo e numero di campi UNIMARC presenti nelle voci di autorità.* La tabella mostra il conteggio dei campi UNIMARC presenti in ogni voce di autorità, allo scopo di verificare se vi sono i campi UNIMARC richiesti dal livello delle voci pubblicate (90, 95 o 97) e di valutare il grado di completezza delle voci stesse.

*Tabella 2.*¹⁷ *Provenienza delle voci di autorità in base al polo SBN.* La tabella mostra il conteggio delle voci di autorità per polo di provenienza (utilizzando il campo UNIMARC 001).

*Tabella 3.*¹⁸ *Voci di autorità minimali.* Questa tabella mostra il numero di voci di autorità nelle quali sono presenti i soli dati essenziali

¹⁶ Gli script usati per generare i dati sono *conteggio_singole_occorrenze_per_tag.sh* e *conteggio_occorrenze_per_tag.sh*.

¹⁷ Lo script usato per generare i dati è *conteggio_per_polo.sh*.

¹⁸ Lo script usato per generare i dati è *conteggio_minimali.sh*.

di una voce di autorità. In base allo standard UNIMARC per voci di autorità, in ogni voce sono attesi il campo 001 con relativo identificatore univoco e formato previsto per un VID costruito secondo l'espressione regolare *ITNCCU\poloV\6 cifre* (per esempio, *ITNCCU\MILV\266512*), un campo 005 con il *timestamp* dell'ultima modifica, un solo campo 100 (dati generali della voce, compresa la data di creazione) e un solo campo 200 con la forma scelta per il nome di persona.¹⁹

Tabella 4²⁰ e 4bis.²¹ Aggiornamenti e creazioni per anno. Queste tabelle mostrano il numero di voci di autorità aggiornate o create per anno da ciascun polo SBN (basandosi sui dati presenti nei campi UNIMARC 001, 005 e 100). La tabella 4bis mostra il dettaglio del solo polo CUB.

Tabella 5.²² Presenza e distribuzione delle date nei campi UNIMARC 200\$f e 300. La tabella 5 mostra la presenza nelle voci di autorità delle date di nascita e morte degli autori rappresentati (attraverso l'analisi dei campi UNIMARC 200\$f e 300).

Tabella 6.²³ Regole di catalogazione. La tabella mostra la presenza del dato relativo alle regole di catalogazione utilizzate per la redazione delle voci di autorità.

Tabella 7.²⁴ Fonti utilizzate. La tabella mostra la presenza e il valore dei dati relativi alle fonti scientifiche utilizzate per il lavoro autorità (repertori bibliografici, bibliografie nazionali dei vari paesi, cataloghi di grandi biblioteche, dizionari biografici, authority file nazionali ed internazionali ecc.) (ICCU 2020e).

¹⁹ <<https://www.ifla.org/wp-content/uploads/2019/05/assets/uca/unimarc-authorities-format.pdf>> (consultato il 17 gennaio 2022); cfr. anche gli aggiornamenti in <<https://www.ifla.org/g/unimarc-rg/unimarc-authorities-3rd-edition-with-updates/>>.

²⁰ Lo script usato per generare i dati è *conteggio_date_creazione_modifica.sh*.

²¹ Lo script usato per generare i dati è nuovamente *conteggio_date_creazione_modifica.sh*.

²² Lo script usato per generare i dati è *conteggio_date_200_300.sh*.

²³ Lo script usato per generare i dati è *conteggio_regole_catalogazione.sh*.

²⁴ Lo script usato per generare i dati è *conteggio_fonti.sh*.

*Tabella 8.*²⁵ *Indicazioni sullo stato di redazione della voce.* La tabella mostra la presenza e il valore dei dati relativi allo stato di redazione della voce (per esempio, «definitivo», «provvisorio» ecc.).

*Tabella 9.*²⁶ *Analisi del campo locale 999.* La tabella mostra la presenza dei collegamenti a EDIT16 e a Manus Online (MOL) introdotti nel dataset POST.²⁷

*Tabella 10.*²⁸ *Dichiarazioni di omonimia nel sottocampo 200\$c.* La tabella mostra il numero di voci di autorità che dichiarano di raccogliere sotto un unico accesso più entità diverse, cioè le voci di autorità non realmente identificative.

Tabella 11. *Presenza dei campi UNIMARC più rilevanti nei dataset PRE e POST.* La tabella riepiloga, da altre tabelle, i dati relativi all'effettiva presenza dei dati obbligatori se applicabili per le voci di autorità di SBN.

*Tabella 12.*²⁹ *Confronto sulla presenza di ISNI, date, paese e lingua in WD e SBN.* La tabella confronta la presenza, in percentuale, di specifici attributi in item (o elementi) Wikidata rispetto alla presenza degli stessi attributi nelle corrispondenti voci di autorità di SBN.

3. *Analisi dei dati*

I risultati delle analisi svolte sui due dataset PRE e POST scaricati dal VIAF e dal nuovo OPAC SBN possono essere utilizzati per creare tabelle che mostrano interessanti aspetti qualitativi e quantitativi dei

²⁵ Lo script usato per generare i dati è *conteggio_status.sh*.

²⁶ Lo script usato per generare i dati è *conteggio_999.sh*.

²⁷ Si vedano rispettivamente <<https://edit16.iccu.sbn.it/>> e <<https://manus.iccu.sbn.it/>>.

²⁸ Lo script usato per generare i dati è *conteggio_200c_omonimie.sh*.

²⁹ Gli script usati per generare la tabella sono *confronto.sh*, *confronto.sparql*, *conteggio_date_200_300_confronto.sh*, *tabella12.sparql*. I file di record UNIMARC ricavati dagli script sono *confronto.mrc* e *confronto_personal_names.mrc*. L'elenco dei VID estratti da Wikidata è *confronto_lista_vid.txt*.

dati presenti nelle voci di autorità di SBN. Le tabelle di dimensioni più piccole sono state inserite all'interno del testo. Le tabelle di dimensioni maggiori sono state invece raccolte in una apposita Appendice e i dati sono stati illustrati nel testo con l'aiuto di grafici.

3.1. Tipo e numero di campi UNIMARC presenti nelle voci di autorità

Dal momento che la presenza di molti campi è indice di maggiore completezza descrittiva della voce, la tabella 1 (vedi Appendice) mostra quali altri campi UNIMARC sono presenti nelle voci di autorità, oltre a quelli essenziali in quanto indispensabili all'esistenza della voce (campi 001, 005, 100, 200, 801). Essendo le voci analizzate di livello alto (95 e 97 per PRE e anche 90 per POST), ci si aspetterebbe voci di autorità che tendono alla completezza delle informazioni relative all'entità descritta e alle fonti utilizzate per il lavoro di autorità.

Solo il campo UNIMARC 300 Nota informativa è presente in più della metà delle voci (54,2%) in PRE, e in meno della metà (48,6%) in POST.

Il secondo campo più diffuso, coerentemente con il lavoro di authority control, è il campo UNIMARC 810 Fonti, che è presente nel 40,7% delle voci PRE, ma solo nel 28,1% delle voci POST. Il terzo dato più diffuso è il campo UNIMARC 010, che contiene l'ISNI della persona rappresentata nella voce di autorità. La presenza di questo dato passa dal 40,3% in PRE al 24,9% in POST.

La tabella 1, inoltre, evidenzia un uso limitato del campo UNIMARC 010 per l'ISNI, che sarebbe ripetibile (Permanent UNIMARC Committee 2022). Rispetto all'obiettivo di inserire nelle voci di autorità di livello 90 «dati sui repertori bibliografici normalmente in uso, quali p.es. le bibliografie nazionali dei vari paesi, cataloghi di grandi biblioteche, dizionari biografici, authority file nazionali ed internazionali, etc. e il collegamento con tutti i legami previsti e applicabili» (ICCU 2017), l'analisi effettuata sui due set di dati denota l'assenza totale del campo UNIMARC 017 che potrebbe invece ospitare identificatori esterni diversi dall'ISNI (per esempio quelli del VIAF, della Library of

Congress, o di Wikidata) (Permanent UNIMARC Committee 2022). Il campo 120 (*Coded Data Field: Personal Name*), destinato alla registrazione del genere della persona in forma codificata, non viene usato.

Tipo e numero di campi UNIMARC presenti

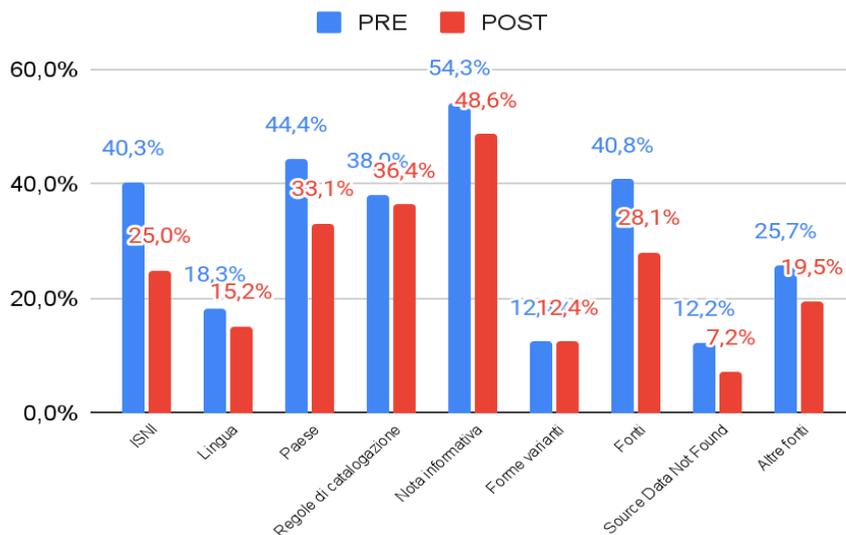


Grafico 3. Tipo e numero di campi UNIMARC presenti

Il grafico 3 evidenzia le occorrenze dei campi UNIMARC nei due dataset e mostra che, nel passaggio da PRE (blu) a POST (rosso), la percentuale dei diversi dati diminuisce costantemente. L'unica eccezione è il campo UNIMARC 999, che fa parte dei campi personalizzabili a disposizione delle agenzie bibliografiche e che è stato introdotto nel nuovo OPAC SBN per contenere gli identificatori di EDIT16 e di MOL per la medesima entità, e quindi per integrare le tre basi di dati su cui si fonda Alphabetica.

3.2. Provenienza delle voci di autorità in base al polo SBN

La tabella 2 (vedi Appendice) presenta i risultati delle analisi re-

lative alla quantità delle voci di autorità secondo il polo SBN che le ha prodotte: suddividendo le voci in base ai poli che le hanno create (l'ordinamento è in base a POST), si ottengono 103 poli in PRE, 112 poli in POST.³⁰ La lunghezza della tabella testimonia una distribuzione molto ampia della provenienza delle voci di autorità. Si può notare che i contributi più consistenti alla costruzione dell'archivio di autorità provengono da un numero ristretto di poli: infatti, solo 5 poli (CFI, CUB, BVE, RAV, SBL) superano le 5000 voci create in entrambi i dataset e solo altri 13 poli superano le 1000 voci create in entrambi i dataset.

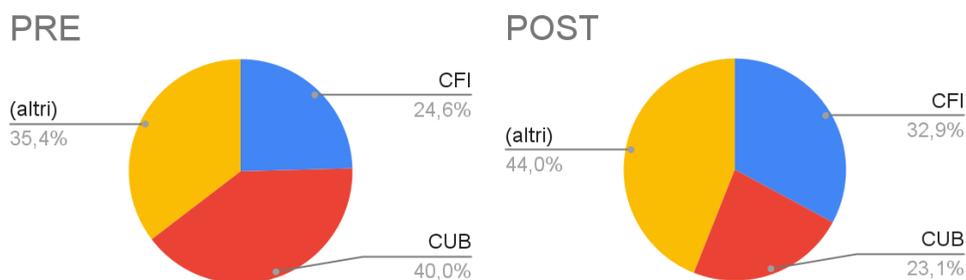


Grafico 4. Provenienza (%) delle voci di autorità per polo, in PRE e POST

Il grafico 4 consente di confrontare anche la percentuale delle provenienze delle voci di autorità in PRE e in POST. In PRE emerge che il polo denominato CUB (che rappresenta probabilmente una importazione di dati dal *Catalogo cumulativo 1886-1957 del Bollettino delle pubblicazioni italiane ricevute per diritto di stampa dalla Biblioteca nazionale centrale di Firenze*) costituisce il 40% delle voci presenti, seguito dal polo CFI (Polo biblioteca nazionale centrale di Firenze), che costituisce un ulteriore 24,6% delle voci (Biblioteca nazionale centrale 1968; 2002). In POST, invece, cresce notevolmente soprattutto la quota del polo CFI, che arriva

³⁰ La lista ufficiale (<<https://www.iccu.sbn.it/it/SBN/poli-e-biblioteche/>>) comprende 103 poli (consultata il 17 gennaio 2022). Non si è riusciti a trovare una lista dei poli comprensiva dei poli non più attivi.

a costituire il 32,9% delle voci e che da solo rappresenta il 43,7% dell'incremento quantitativo, così come crescono di più di 5000 voci i poli UFI (Polo Università degli studi di Firenze), BVE (Polo Biblioteca nazionale centrale di Roma), PAL (Polo regionale di Sicilia) e IEI (Polo degli Istituti culturali di Roma). In POST, la somma del polo CUB (23,1%) e del polo CFI continua a costituire la maggioranza assoluta delle voci.³¹

3.3. Voci di autorità minimali

La tabella 3 (vedi Appendice) presenta i dati relativi alle voci di autorità minimali, cioè quelle estremamente povere dal punto di vista del contenuto, perché contengono esclusivamente i dati minimi indispensabili all'esistenza della voce. Quindi, con voci di autorità minimali si indicano quelle per le quali sono presenti soltanto i campi UNIMARC 001, 005, 100, 200 e 801.

Nel grafico 5 è rappresentato il numero di voci minimali aggiornate per anno.

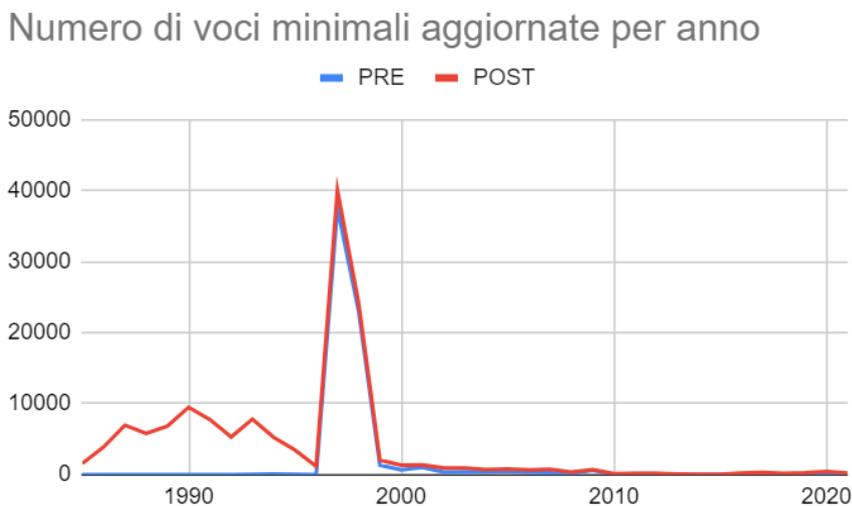
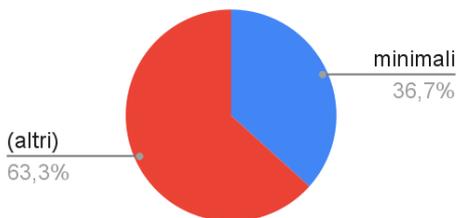


Grafico 5. Numero di voci minimali aggiornate per anno

³¹ Il polo UFI risulta come polo proposto in <<https://www.iccu.sbn.it/it/SBN/organ-di-governo-sbn/documenti-di-proposta-dei-poli-sbn/index.html>>.

PRE (conteggio totale)



POST (conteggio totale)

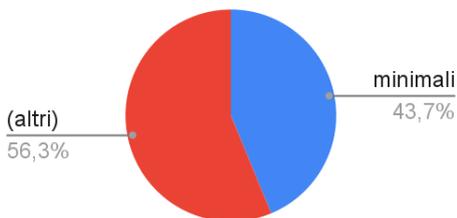
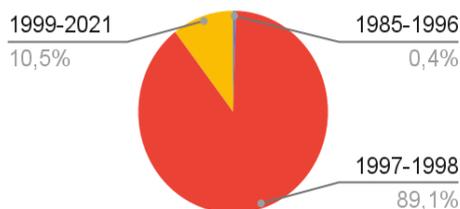


Grafico 6. Percentuale di voci minimali sul totale

PRE (conteggio anni)



POST (conteggio anni)

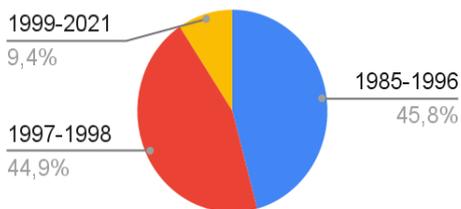
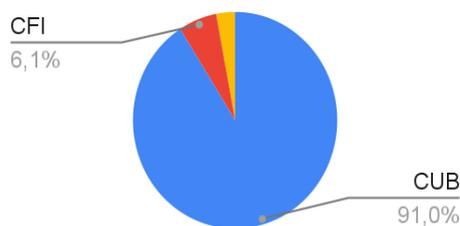


Grafico 7. Percentuale di voci minimali per periodo di aggiornamento

PRE (conteggio poli)



POST (conteggio poli)

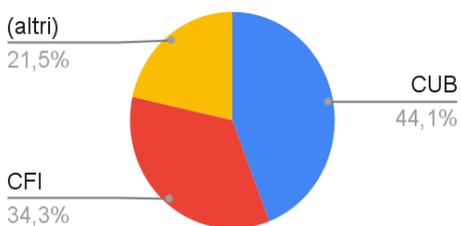


Grafico 8. Percentuale di voci minimali per polo

Si può notare dal grafico 6 che le voci minimali costituiscono il 36,6% delle voci PRE e il 43,7% delle voci POST.

La differenza tra le due situazioni è evidente. In PRE il 91% delle voci minimali proviene dal polo CUB (grafico 8) e l'89,1% (60478

voci) del totale delle voci minimali è stato aggiornato nel 1997-1998 (grafico 7). In POST c'è maggiore "frammentazione": considerando i poli, la prevalenza di CUB (il 44,1% del totale) è molto meno netta, soprattutto per la crescita di CFI (il 34,3% del totale) (grafico 8). Considerando le date di aggiornamento (grafico 7), anche la prevalenza del 1997-1998 (64107 voci, il 44,9% del totale) diminuisce significativamente, soprattutto a causa dell'incremento del periodo 1985-1996.³²

3.4. Aggiornamenti e creazioni per anno

La tabella 4 (vedi Appendice) analizza le voci di autorità rispetto a quando sono state aggiornate (valore presente nel campo UNIMARC 005) o create (valore presente nel campo UNIMARC 100); dall'incrocio dei due dati si ricavano le voci mai modificate dopo essere state create.



Grafico 9. Linea del tempo degli aggiornamenti e delle creazioni per anno

³² L'incremento del 1985-1996 costituisce l'86,8% dell'incremento totale delle voci minimali.

Dal grafico 9, si vede che l'aggiornamento delle voci avviene in modo sostanzialmente corrispondente in PRE e POST e che sono presenti due picchi di aggiornamento: il primo in corrispondenza del biennio 1997-1998 (60910 voci PRE, 66328 voci POST) legato all'inserimento del CUB (vedi tabella 4bis in Appendice) e il secondo nel 2021 (73348 voci PRE, 82990 voci POST) legato all'arricchimento delle voci di autorità con il numero ISNI.³³ La principale differenza tra i due dataset consiste nel numero di voci aggiornate nel 1985-1996, solo 421 in PRE ma 72876 in POST (questo incremento di 72455 voci costituisce il 51,1% dell'incremento totale tra le due fasi).

Le date di creazione presentano una distribuzione per certi aspetti simile a quella delle date di aggiornamento: un picco nel 1997-1998 (74835 voci create in PRE, 82171 voci create in POST), che corrisponde a quello delle date di aggiornamento (ci sono infatti 60750 voci in PRE e 64926 voci in POST create nel 1997-1998 e da allora mai modificate); un calo delle creazioni nel periodo 2019-2021 (1810 in PRE e 4978 voci in POST), che non trova invece riscontro nelle date di aggiornamento; un netto aumento da PRE a POST per quanto riguarda il periodo 1985-1996 (da 15630 voci create a 115599 voci create), che trova riscontro nelle date di aggiornamento (ci sono infatti solo 300 voci PRE, ma 70487 voci POST create nel 1985-1996 e da allora mai modificate).

³³ L'inserimento dell'ISNI è probabilmente avvenuto in modo automatico o semiautomatico; ciò ha determinato un aggiornamento della data della voce, ma non ha implicato necessariamente l'aggiornamento complessivo della voce di autorità, i cui restanti dati non sono stati controllati né aggiornati. Per esempio, la voce IT\ICCU\RAVV\000849 risulta aggiornata nell'aprile 2021; l'aggiornamento è collegato all'aggiunta dell'ISNI, ma la data di morte dell'autore (2019) non è stata aggiunta; si veda <<https://www.artribune.com/arti-visive/arte-contemporanea/2019/10/morto-nel-suo-abruzzo-ettore-spalletti-aveva-79-anni/>>.

4bis. Aggiornamenti e creazioni per anno nel polo CUB

La tabella 4bis (vedi Appendice) fornisce le medesime informazioni della tabella 4 ma l'analisi dei dati è limitata alle voci di autorità del polo CUB.

Aggiornamenti e creazioni CUB per anno

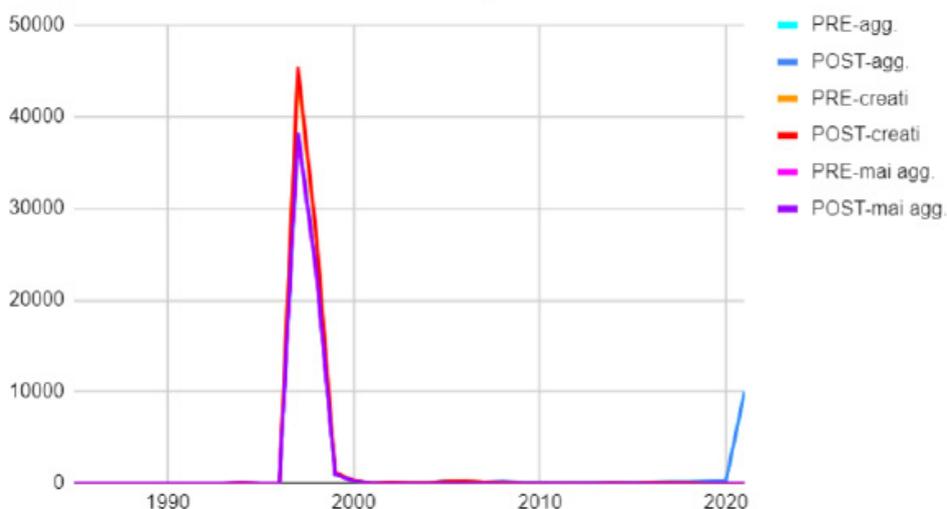


Grafico 10. Linea del tempo degli aggiornamenti e delle creazioni per anno per il polo CUB

Il grafico 10 evidenzia che il polo CUB ha creato voci dal 1985 fino al 2018.

Le date di aggiornamento delle voci (campo UNIMARC 005) mostrano che la maggioranza assoluta³⁴ di esse in entrambi i dataset è stata creata nel 1997-1998 e mai modificata da allora; si tratta sempre di voci di autorità minimali.

³⁴ Rispettivamente l'81,2% (60246 voci) PRE e l'81,1% (61380 voci) POST.

3.5. Presenza e distribuzione di date nei campi UNIMARC 200\$f e 300

La tabella 5 (vedi Appendice) prende in esame la presenza di date relative alla persona rappresentata dalla voce di autorità. Le date di nascita e di morte di una persona sono un elemento essenziale per la sua identificazione, al punto che molte regole di catalogazione internazionale ne richiedono la registrazione sistematica. Il formato UNIMARC prevede la registrazione di tali date nel sottocampo 200\$f, per esempio con una precisione all'anno o con indicazioni come *floruit* ecc.

L'analisi dei due dataset PRE e POST ha rivelato che nelle voci di autorità le date – anche con precisione al mese o al giorno – sono registrate sistematicamente non nel sottocampo 200\$f, bensì nel campo UNIMARC 300 Nota informativa. In questo campo, le voci di autorità di SBN contengono anche informazioni relative all'occupazione, a luoghi e di altra natura, e le date vengono distinte tramite un separatore che consente di individuare la stringa come riferibile a una data (si tratta di uno spazio, doppia barra, spazio).³⁵

Considerando l'uso del sottocampo 200\$f (cioè le date usate per disambiguare omonimi), si nota un lieve calo percentuale da PRE (16,1%) a POST (15,5%); considerando invece l'uso di inserire le date nel campo 300, seguite da una // e dalla descrizione verbale delle attività della persona, anche in questo caso si ha un calo percentuale, dall'86,3% al 78,9%.³⁶ In POST aumenta significativamente il numero di casi in cui una data presente nel sottocampo 200\$f non viene ripetuta nel campo 300 prima della // (diversamente da quanto previsto dalle norme), passando dal 21,4% al 37% dei casi.

³⁵ Un esempio si trova in IT\ICCU\CFIV\103849. L'uso è stabilito in (ICCU 2019, p. 23).

³⁶ Le percentuali sono calcolate sui casi in cui il campo 300 è presente, non sul totale delle voci.

3.6. Regole di catalogazione

La tabella 6 analizza i dati nel campo UNIMARC 152 relativo alle regole di catalogazione utilizzate per redigere la voce di autorità. Le voci che possiedono tale campo sono in percentuale piuttosto bassa e molto simile nei due dataset: 37,6% in PRE e 36,3% POST. Le voci di autorità basate sulle RICA sono in quantità trascurabile (23 e 53 rispettivamente); in tutti gli altri casi, le voci sono basate sulle REICAT.

	PRE	% sul totale (185132)	POST	% sul totale (326849)
REICAT	69739	37,6%	118765	36,3%
RICA	23	0,0%	53	0,0%

Tabella 6. Regole di catalogazione

3.7. Fonti utilizzate

La tabella 7 mostra la presenza di dati relativi alle fonti utilizzate per la compilazione delle voci di autorità, un aspetto di particolare rilievo per il lavoro di autorità. Nel formato UNIMARC le fonti possono essere indicate nei campi 810 Fonti rilevanti e 815 Fonti con esito negativo;³⁷ nelle voci di autorità in formato UNIMARC di SBN, però, le fonti possono essere state indicate anche nel campo UNIMARC 830 dedicato alla nota del catalogatore. Secondo le linee guida del 2019, nel campo UNIMARC 810 «si registrano i codici relativi ai repertori consultati e codificati a livello centrale (p. es. DBI, DEUMM, EDIT16 ...)», mentre nel campo UNIMARC 830 «si registrano i repertori consultati ma non codificati a livello centrale» (tipicamente il campo 830 contiene link a pagine web) (ICCU 2019).

Il dato più significativo è che le voci di autorità sono prive di qual-

³⁷ Questo campo è presente nel formato di esportazione ma non è visibile nella visualizzazione a etichette nell'OPAC SBN (es. IT\ICCU\CFIV\003748).

siasi indicazione di fonte nel 52,2% delle voci PRE e nel 65,4% delle voci POST.

Tra le voci prive di fonti, si trova una percentuale molto alta di voci di autorità minimali: il 70,2% delle voci prive di fonti PRE e il 43,7% delle voci prive di fonti POST.

PRE					POST				
810	815	830	Conteggio	% sul totale	810	815	830	Conteggio	% sul totale
0	0	0	96637	52,2%	0	0	0	213509	65,4%
0	0	1	8140	4,4%	0	0	1	16161	4,9%
0	1	0	224	0,1%	0	1	0	475	0,1%
0	1	1	4683	2,5%	0	1	1	4860	1,5%
1	0	0	34636	18,7%	1	0	0	42607	13,0%
1	0	1	23160	12,5%	1	0	1	30894	9,5%
1	1	0	5967	3,2%	1	1	0	6464	2,0%
1	1	1	11685	6,3%	1	1	1	11879	3,6%
TOTALE			185132		TOTALE			326849	

Legenda: 0 assente; 1 presente (anche più occorrenze)

Tabella 7. Presenza nelle voci di autorità dell'indicazione delle fonti utilizzate

3.8. Indicazioni sullo stato di redazione della voce

La tabella 8 verifica, in base al campo UNIMARC 100, posizione 8, lo status di redazione della voce di autorità, che ha lo scopo di indicare, in un ambiente cooperativo, il grado di affidabilità della voce di autorità in relazione al tipo di lavoro svolto per la sua redazione. Le voci di autorità sono prive di questa indicazione («not applicable») rispettivamente nel 72% dei casi per PRE nell'83,4% dei casi per POST. La percentuale arriva a sfiorare la totalità delle voci se si restringe l'analisi al solo polo CUB (97,9% in entrambi i dataset). Un esame a campione delle voci di autorità che invece possiedono lo status di voci definitive («established») mostra che queste voci sono

di qualità superiore alla media. Non è stato possibile ricostruire un insieme di campi la cui presenza contemporanea consenta di prevedere con certezza l'assegnazione dello status definitivo alla voce.

campo 100, posizione 8	valore	PRE	PRE-CUB	POST	POST- CUB
a	established	51864	1591	54101	1613
c	provisional	0	0	0	0
x	not applicable	133268	72539	272748	74018
TOTALE		185132	74130	326849	75631

Tabella 8. Indicazioni sullo stato di redazione della voce

3.9. *Analisi del campo locale 999*

La tabella 9 evidenzia un'importante novità introdotta nel dataset POST all'interno dei campi UNIMARC riservati alle agenzie per inserire dati di valore locale. Nelle voci di autorità esaminate il campo 999 contiene infatti gli identificatori delle voci di autorità di EDIT16 e di Manus Online (MOL) che rappresentano la medesima entità. EDIT16 e MOL sono due componenti basilari, oltre a SBN, del nuovo portale Alphabetica e la presenza di queste informazioni nel campo UNIMARC costituisce un tassello essenziale per il suo funzionamento. In base ai dati analizzati, le voci collegate con EDIT16 risultano 6390 e 13720 con MANUS; sono infine 18053 le voci collegate con almeno uno dei due.

Identificatori	POST
totale link a EDIT16 e a MOL	20110
di cui EDIT16	6390
di cui MOL	13720
totale voci contenenti almeno un link a EDIT16 o a MOL	18053

Tabella 9. Analisi del campo 999

3.10. Dichiarazioni di omonimia nel sottocampo 200\$c

La tabella 10 mostra come vengono qualificati i nomi utilizzando informazioni diverse dalla data di nascita o di morte attraverso l'uso del sottocampo UNIMARC 200\$c *Aggiunte al nome diverse dalle date* come titoli, epiteti, compiti ecc.

Un aspetto interessante su cui si è concentrata l'analisi si basa sulla regola delle REICAT e applicata in SBN per la quale: «se uno stesso nome si riferisce sicuramente a due o più persone, ma mancano informazioni sufficienti a identificarle e a collegare all'intestazione per ciascuna persona le opere di cui è responsabile, può essere eccezionalmente usata un'intestazione indifferenziata. Si aggiunge un'avvertenza in proposito, sotto forma di qualificazione» (Commissione permanente per la revisione delle regole italiane di catalogazione 2009, par. 15.3.4).

Perciò, l'analisi svolta sul sottocampo UNIMARC 200\$c era mirata a verificare la presenza nella voce di autorità dei seguenti valori: «omonimi», «autore indifferenziato» e «raggruppa autori diversi».

Tipo di dichiarazione	PRE	POST
omonimi	2	388
autore indifferenziato	11	17
raggruppa autori diversi	1	2
TOTALI	14	407

Tabella 10. Dichiarazioni di omonimia nel sottocampo 200\$c

I risultati mostrano in un numero di casi ridotto, ma comunque significativo (soprattutto in POST), come tra le voci di autorità di livello 90 e superiore siano presenti almeno 407 casi che raggruppano persone omonime indifferenziate (388 casi di omonimi, 17 autori indifferenziati e 2 raggruppamenti di autori diversi).

4. *Discussione*

L'analisi dei dati delle voci di autorità si è basata su dati che sono stati ottenuti dal VIAF o dal nuovo OPAC SBN con tecniche non immediatamente accessibili a tutti. Infatti, diversamente da quanto accade per altre agenzie nazionali (per esempio DNB, British Library, la National Diet Library of Japan, Library of Congress), i due dataset analizzati si possono ottenere solo superando alcune difficoltà tecniche che richiedono capacità non elementari. Ciò significa che i dati non sono accessibili e immediatamente disponibili alla comunità bibliotecaria italiana.

I risultati delle analisi presentati nelle tabelle e nei grafici consentono di evidenziare tre principali aspetti problematici nella qualità delle voci di autorità di SBN:

1. la completezza dei dati nelle voci non corrisponde pienamente alle ragionevoli aspettative suscitate dal livello di autorità associato alle voci pubblicate (livello 90 o superiore);
2. la forma di registrazione dei dati non è pienamente adeguata alle indicazioni del formato UNIMARC, in particolare modo in relazione a elementi di informazione rilevanti come le date e gli identificatori;
3. il sistema dei livelli gerarchici delle voci di autorità non è in grado di garantire effettivamente una qualità adeguata ai requisiti dei livelli 90 o superiori, ma è probabilmente di ostacolo alla possibilità che le voci minimali vengano arricchite in modo cooperativo dall'intera comunità.

Se si guarda al lavoro di autorità accessibile al pubblico da una prospettiva esterna come quella del VIAF o di Wikidata, la comparazione con le voci di autorità delle altre agenzie nazionali presenti nel VIAF mostra che il contributo italiano è del tutto sproporzionato in termini quantitativi. Inoltre, è presumibile che tale proporzione rimanga significativa nel breve e medio periodo, anche se si considerano le dimensioni del dataset POST e se si ipotizza (o meglio si auspica) un suo invio al VIAF.

4.1 Incoerenza tra linee guida e qualità globale effettiva

Nella *Guida alla catalogazione SBN - Musica* (ICCU 2020a) è possibile trovare una indicazione più precisa, rispetto a quanto previsto dalle *Norme per il trattamento di informazioni e dati comuni a tutte le tipologie di materiale* (ICCU 2020e), riguardo ai metadati che le regole di SBN considerano obbligatori, se applicabili, in relazione ai livelli di catalogazione.

Per il livello 90 sono obbligatori l'ISNI, la Descrizione, le Fonti, la Nota del Catalogatore e tutte le forme varianti rilevate nei repertori (oltre ai legami, che non sono stati presi in esame nell'analisi). Per il livello 95 sono obbligatori anche la Lingua, il Paese e la Nota informativa. Infine, il livello 97 viene definito *livello di authority* ed è "previsto per quelle notizie per cui sono state operate tutte le attività di controllo sui repertori e di validazione dei dati" (ICCU 2020a).

Grazie a queste indicazioni, è possibile interpretare con maggiore sicurezza i dati relativi al tipo e al numero di campi UNIMARC (tabella 1) e, in parte, quelli relativi alla presenza e alla distribuzione delle date e alle fonti utilizzate (tabelle 5 e 7), riepilogati di seguito:

Metadato	campo	PRE	POST	fonte (tabella)
ISNI	010	40,3%	25,0%	1
lingua	101	18,3%	15,2%	1
paese	102	44,4%	33,1%	1
nota informativa	300	54,2%	48,6%	1
datazione	300 prima della //	46,8%	38,4%	5
fonti o altre fonti	810 o 830	47,7%	34,6%	7

Tabella 11. Presenza dei campi UNIMARC più rilevanti nei dataset PRE e POST

I dati elencati in tabella 11 sono obbligatori, *se applicabili*; tuttavia, la percentuale di presenza di ciascun dato non supera mai il 55% in nessuno dei due dataset PRE e POST. Si può ragionevolmente supporre che una certa percentuale di voci di autorità prive di uno o più di questi dati siano dovute all'impossibilità di reperire facilmente il dato. Ma da un lato il paese e la lingua non sono dati particolarmente difficili da trovare; dall'altro, è sorprendente la percentuale davvero bassa (34,6%) di voci di autorità con l'indicazione di una fonte.

Una verifica sull'applicabilità dei dati nelle voci di autorità è stata fatta utilizzando gli item di Wikidata e le loro proprietà. Esaminando le voci SBN e i corrispondenti item Wikidata (grazie all'identificazione presente in Wikidata) si è calcolata la percentuale di ricorrenze dei campi UNIMARC obbligatori se applicabili e delle proprietà Wikidata, in un campione di 46666 entità.³⁸

I risultati sono riassunti in tabella 12; in Wikidata l'ISNI (-3,9%) e la nazionalità (-7,7%) sono meno presenti che in SBN, mentre la datazione (+8,9%) e la lingua (+43,9%) sono più rappresentate. Inoltre questi quattro attributi sono tutti presenti nel 59,9% delle entità di Wikidata contro il 25,5% delle entità di SBN. Le percentuali di Wikidata mostrano quindi che i dati in oggetto sarebbero stati facilmente applicabili almeno in un altro 34,4% dei casi.

Dal confronto delle percentuali di completezza delle voci di autorità di SBN di tabella 11 e tabella 12 emerge inoltre che le voci di autorità connesse al corrispondente elemento Wikidata sono molto più complete rispetto alla media di SBN, sia PRE sia POST (tabella 11). Ciò dimostra che le voci di autorità di SBN possono essere interconnesse più facilmente quando sono più complete, in particolare rispetto alla presenza di date.

³⁸ In Wikidata 46666 elementi di persone linkano a SBN; tuttavia, 712 VID SBN non sono considerati nella tabella perché inaccessibili (sono i casi in cui Wikidata punta a VID di livello inferiore a 90).

	in Wiki- data	proprietà Wikidata	% in Wiki- data	in SBN	campo UNIMARC	% in SBN	diff % WD-SBN
N. ele- menti con ISNI	46666	P396	100,0%	45954	001	100,0%	0,0%
con data- zioni nazio- nalità lingua	36860	P213	79,0%	38092	010	82,9%	-3,9%
		P569 o					
		P570 o					
	46065	P1317	98,7% ³⁹	41290	300 prima di //	89,9%	+8,9%
	36425	P27	78,1%	39414	102	85,8%	-7,7%
	35347	P1412	75,7%	14626	101	31,8%	+43,9%
tutti i dati qui sopra		P396 + P213 + P27 + P1412 + (P569 o P570 o P1317)					
	27948		59,9%	11740	001 + 010 + 101 + 102 + 300 prima di //	25,5%	+34,3%

Tabella 12. Confronto sulla presenza di ISNI, date, paese e lingua in Wiki-
data e SBN

Le percentuali di voci minimali (tabella 3: 36,7% PRE e 43,7% POST) sono alte rispetto alle attese, considerato che in realtà dovrebbero essere classificate – secondo la normativa interna di SBN – con livello di autorità 51.

Una parte significativa di tali voci minimali (44,1% in POST e addirittura 90,9% in PRE) è costituita da voci del polo CUB e la loro data di aggiornamento più frequente è il periodo 1997-1998 (89,1% PRE e 44,9% POST). L'importazione delle voci del CUB ha quindi falsato significativamente la qualità delle voci di livello 95 (o superio-

³⁹ Di cui: 64,3% con P569 e P570; 98,1% con P569; 64,6% con P570.

re) e tali voci sono state poi arricchite solo in piccola percentuale. È verosimile che lo scarso intervento su queste voci sia dovuto al livello di autorità molto alto che è stato loro attribuito, che ha impedito alla comunità dei catalogatori di SBN di aggiornarle anche in casi in cui i dati fossero facilmente accertabili.

4.2 Forma di registrazione dei dati inadeguata

Quasi la metà delle voci di autorità di POST contiene descrizioni particolarmente ricche di informazioni nel campo UNIMARC 300 e l'analisi mostra che lo sforzo del controllo di autorità è concentrato particolarmente in questo campo. Purtroppo, però, se si tenta di esportare i dati e di riutilizzarli, ci si scontra con problemi di insufficiente granularità dei dati.

Ci sono in particolare due punti critici: i dati anagrafici e i codici identificativi.

I dati anagrafici (date e luoghi di nascita e morte) sono attualmente registrati nei seguenti campi:

- 200\$f per datazioni usate come qualificazione cronologica, soltanto per disambiguare; si usano in prima istanza l'anno di nascita, in seconda istanza il luogo di nascita, in terza istanza il giorno di nascita (ICCU 2020b);

- 300 prima della doppia barra (//) per le datazioni in generale (soltanto anno) (ICCU 2020c);

- 300 dopo la //, in forma discorsiva, per le date di nascita e morte (precisione al giorno, quando possibile) e i luoghi di nascita e morte (ICCU 2020d).

In SBN quindi la registrazione della data non avviene in maniera coerente e costante nel sottocampo 200\$f, ma subordinatamente alla presenza, nell'indice SBN, di una voce da disambiguare. Si tratta di un problema la cui rilevanza cambia completamente a seconda del punto di vista. Internamente a SBN, questa prassi ha ragioni di carattere funzionale e di economia; esternamente, SBN manca di identificare

adeguatamente le proprie entità rispetto alle entità censite da tutte le altre agenzie nazionali. Infatti, la prassi di SBN differisce da quella di tutte le altre agenzie che catalogano in UNIMARC e contribuiscono a VIAF, che riportano sistematicamente gli anni di nascita e morte nel sottocampo 200\$f, proprio a scopo identificativo.⁴⁰

Infine, può essere utile segnalare come due biblioteche nazionali che usano l'UNIMARC provano a strutturare le date di nascita e di morte con precisione al giorno e a indicare in campi appositi i luoghi di nascita e di morte:

- la Bibliothèque nationale de France (codice BNF nel VIAF) usa il campo 103 per le date di nascita e morte con precisione al giorno e i sottocampi 301\$a per luogo di nascita e 301\$b per luogo di morte (ma entrambi non sono menzionati nel manuale UNIMARC dell'IFLA);⁴¹

- la Lietuvos nacionalinė Martyno Mažvydo biblioteka (Biblioteca Nazionale della Lituania Martynas Mažvydas, codice LIH nel VIAF) usa il sottocampo 340\$a⁴² per le date di nascita e morte con precisione al giorno, riservando la descrizione biografica in forma discorsiva al sottocampo 340\$b;⁴³ l'uso del sottocampo 340\$a per le date con precisione al giorno sembra una decisione autonoma della Biblioteca Nazionale della Lituania.

L'unico codice identificativo presente in forma strutturata nelle voci SBN è l'ISNI (campo UNIMARC 010), mentre altri identificato-

⁴⁰ Per esempio, nel cluster VIAF <<https://viaf.org/viaf/97105654/>> si può notare che, tra le sei agenzie che contribuiscono al VIAF inviando i dati in formato UNIMARC, SBN è l'unica a non utilizzare il sottocampo 200\$f.

⁴¹ Si vedano, per esempio, <<https://viaf.org/processed/BNF%7C11911631>> e <<https://viaf.org/processed/BNF%7C11897013>>.

⁴² Il manuale UNIMARC dell'IFLA prescrive l'uso del sottocampo 340\$a per la descrizione biografica in forma discorsiva e non prescrive l'uso del sottocampo 340\$b; un aggiornamento del 2016 introduce l'uso del sottocampo 340\$b come "activity note", riservando il sottocampo 340\$a a una "biographical or historical note"; si veda Permanent UNIMARC Committee 2022.

⁴³ Si vedano, per esempio, <<https://viaf.org/processed/LIH%7CLNB:-V-35849;=BP>> e <<https://viaf.org/processed/LIH%7CLNB:-V-261357;=BK>>

ri potenzialmente utili, come VIAF, LCNAF, GND, BNF e Wikidata vengono sporadicamente riportati, in forma non strutturata, nel campo UNIMARC 830.

Il manuale UNIMARC dell'IFLA non menziona né il campo 010 né alcun altro campo adibito a contenere identificatori (se non il campo UNIMARC 015, riservato all'ISADN). Tuttavia, l'IFLA ha introdotto il campo UNIMARC 010 per indicare l'ISNI nel 2016 (Permanent UNIMARC Committee 2022) ed esso è recepito dal VIAF e usato anche da BNF ed Ethnikí Vivliothíki tis Elládos (Biblioteca nazionale greca, codice GRATEVE nel VIAF), oltre che da SBN.⁴⁴

Inoltre, nel 2017 l'IFLA ha introdotto anche il campo UNIMARC 017, ripetibile, da usarsi per identificatori diversi dall'ISNI (es. ORCID): si indicano l'identificativo nel \$a e il nome dell'identificatore nel \$2. (Permanent UNIMARC Committee 2022) Sarebbe dunque importante che SBN cominciasse a usare anche il campo 017 per gli identificatori diversi dall'ISNI, in modo da evitare la loro dispersione in forma non strutturata nel campo 830.

4.3 I livelli gerarchici delle voci di autorità

La presenza del dato del livello di autorità di una voce di per sé non è un aspetto negativo, soprattutto perché è stata progettata per costituire un utile strumento di valutazione della qualità delle voci di autorità stesse. Tuttavia il livello di autorità viene determinato dal livello del catalogatore che crea e/o modifica il dato e non dalla qualità della voce stessa e anche, come si è riscontrato, indipendentemente dalla completezza raggiunta, dal numero di fonti controllate o dal controllo effettivo della qualità e dell'aggiornamento dei dati. Questo spiega perché più di un terzo delle voci attualmente a livello 90 o superiore dovrebbe avere in realtà un livello 51.

⁴⁴ Per il VIAF si veda <<https://viaf.org/viaf/unimarcTags.xml>>; per BNF <<https://viaf.org/processed/BNF%7C11911631>> e per GRATEVE <<https://viaf.org/processed/GRATEVE%7C19393>>.

La gestione gerarchica dei livelli di autorità ha altre due importanti conseguenze:

- l'impossibilità di modificare e migliorare direttamente voci povere o errate quando, come catalogatori, non si possiede un livello adeguato;⁴⁵

- la visibilità della voce nell'OPAC SBN (e quindi anche nel VIAF) solo a partire dal livello 95 (PRE) o 90 (POST);

Entrambi questi aspetti hanno un impatto negativo sul miglioramento qualitativo delle voci di autorità, la prima in termini di cooperazione interna al sistema SBN, la seconda in termini di interoperabilità con altri database esterni.

4.4 Interoperabilità con altri dataset internazionali

I livelli gerarchici delle voci di autorità così come sono strutturati e utilizzati oggi in SBN costituiscono un intralcio pesante all'interoperabilità dei dati completi di SBN con altri dataset. Infatti, si può ragionevolmente supporre che in termini quantitativi le voci di autorità di livello 90 e superiore siano la punta di un iceberg, la cui base è costituita da un numero di voci probabilmente molto più alto di livello inferiore al 90.⁴⁶ Tuttavia, queste voci di autorità non sono visibili all'utente di SBN né, soprattutto, sono rese disponibili a livello internazionale per promuovere l'interoperabilità di SBN con altri dataset simili.

Nessuna delle principali biblioteche nazionali (per esempio, Library of Congress, BNF, DNB) presenti nel VIAF adotta una politica

⁴⁵ Se è vero che il meccanismo della proposta di correzione che il catalogatore di livello inferiore può avanzare è in via teorica corretto, all'atto pratico costituisce una procedura i cui esiti sono evidenti e che va a detrimento dell'aspetto cooperativo del lavoro di autorità.

⁴⁶ Gli autori non hanno accesso ai dati interni di SBN e, come indicato nella sezione Materiali e metodi, la ricerca si è basata esclusivamente sull'analisi di ciò che di SBN è pubblicamente e gratuitamente accessibile. Si può ragionevolmente ipotizzare che il numero totale delle voci di autorità sia superiore di un ordine di grandezza rispetto a quelle visibili su POST.

di occultamento delle voci di autorità di qualità media o bassa. Tutte le voci sono rese visibili nel web e consentono di identificare le entità che rappresentano tramite il collegamento tra identificatori locali e il VIAF ID. In questo modo, ogni agenzia si avvantaggia reciprocamente del lavoro svolto dalle altre agenzie bibliografiche nazionali, soprattutto per le voci di più difficile identificazione. La clusterizzazione delle voci di autorità nel VIAF si avvale infatti sia dei dati disponibili in una specifica voce di autorità (anche minimale) sia delle relazioni di quella voce con le registrazioni bibliografiche. Perciò, anche quando la voce di autorità è di qualità bassa o media, l'enorme quantità di informazioni disponibili tramite l'insieme dei contributi delle agenzie nazionali consente all'algoritmo del VIAF di tentare un abbinamento molto spesso con esito positivo.

Infine, se i livelli gerarchici fossero applicati in base a una reale autorevolezza e completezza delle voci di autorità si potrebbe giustificare – ma non condividere nell'ottica del web semantico – che le voci di livello 90 o superiore possano essere modificate soltanto da un ristretto nucleo di bibliotecari esperti. Ma i dati mostrano che l'attuale applicazione contemporanea di livelli delle voci di autorità in modo non coerente e la gerarchia dei livelli dei catalogatori non garantiscono che le voci siano di qualità e allo stesso tempo impediscono che vengano migliorate. In sintesi, il doppio livello gerarchico costituisce il presupposto ideale per impedire che nel breve e medio termine il numero di voci minimali di livello 90 o superiore cali significativamente.

Il numero di voci con dichiarazione di omonimia (tabella 10) costituisce inoltre un segnale che è concreto il rischio che le voci minimali con livello alto vengano utilizzate, in mancanza di soluzioni alternative, per raccogliere persone omonime indifferenziate.

4.5 Il riuso di Wikidata in Alphabetica

Il portale Alphabetica rappresenta un indubbio «passo avanti per l'attività di valorizzazione di beni culturali complessi come quelli che

appartengono all'universo bibliografico» grazie a uno strumento «concepito per avvicinare ancora di più i cittadini al patrimonio culturale attraverso nuovi percorsi di fruizione e strumenti digitali» (ICCU 2021).

Una funzione particolarmente interessante relativa alle voci di autorità è svolta dalla innovativa sezione Protagonisti. La visualizzazione dei dati di un «protagonista» si basa sulla voce di autorità di SBN, ma può includere anche informazioni che non fanno parte della voce di autorità originale di SBN. Infatti, se in Wikidata è presente un elemento (item) tra le cui proprietà è stato registrato anche l'«identificativo SBN di un autore» (P396), Alphabetica completa l'informazione da inviare al browser riutilizzando anche alcuni dati che provengono dalle proprietà dell'elemento di Wikidata corrispondente.

Per esempio, in data 30 gennaio 2022, la visualizzazione offerta per il «protagonista» con l'identificatore IT\ICCU\RAVV\095683⁴⁷ era quella che compare in Figura 1.



Figura 1 – Visualizzazione di un ‘protagonista’ prima dell’interlinking con Wikidata.

L’aggiunta, nella proprietà P396 dell’elemento Wikidata

⁴⁷ <<https://alphabetica.it/protagonisti-risultati/-/s/results?input=RAVV095683>>.

Q110734071⁴⁸, del valore IT\ICCU\RAVV\095683 ha completato il processo di identificazione reciproca delle due voci e ha creato i presupposti affinché alcuni dati di Wikidata potessero essere riusati nella visualizzazione della medesima entità in un dataset diverso (Alfabetica), che ora compare arricchita come in figura 2 e 3.



Figura 2 – Vista di un “protagonista” con il riuso di dati provenienti da Wikidata (Bio).

La figura 2 mostra come l’entità sia presentata ora con luoghi e date di nascita e di morte. Compare anche la sezione “Riferimenti esterni” (figura 3), che ordinariamente contiene il link a ISNI (dal campo 010 della voce di autorità) e link non registrati nella voce di autorità come VIAF e Wikidata. In altri casi (figura 4), viene proposta l’eventuale immagine presente in Wikimedia Commons e associata all’elemento in Wikidata tramite la proprietà P18 «immagine». Alfabetica ha anche attivato un’opzione in più per l’esplorazione delle informazioni, sotto forma di «Radar»; è un servizio che mostra in modo pratico ed elegante le relazioni del «protagonista» con altri «protagonisti», opere, luoghi e anni.

⁴⁸ <<https://www.wikidata.org/wiki/Q110734071>>.



Figura 3 – “Protagonista” con collegamento da Wikidata a SBN (Riferimenti esterni).

La fiducia che, grazie al riuso, Alfabetica accorda ai dati provenienti da Wikidata richiede una riflessione per la rilevanza delle sue implicazioni. L'aspetto critico non è tanto nella scelta – del tutto condivisibile, a parere di chi scrive – di riutilizzare i dati di Wikidata, che è ormai riconosciuta come una risorsa di importanza cruciale nel web semantico, anche in relazione al contributo di alto profilo che può fornire allo sviluppo dell'ecosistema dei dati bibliografici (Agenjo-Bullón - Hernández-Carrascal 2018; Allison-Cassin - Scott 2018; Association of Research Libraries 2019; Linked Data for Production 2020; Coladangelo - Ransom 2021; Dostál 2021; Bianchini - Bargioni - Pellizzari 2021; Bianchini - Sardo 2022). Si tratta piuttosto di rilevare che la scelta di avvalersi dei dati di una piattaforma aperta e collaborativa e di fornirli come servizio aggiuntivo agli utilizzatori di Alfabetica sia, di principio, in aperta contraddizione con la rigida gerarchia dei livelli delle voci di autorità e dei catalogatori che contraddistingue SBN come sistema di catalogazione partecipata nel panorama italiano, o forse anche internazionale. Da un lato le voci di autorità vengono pubblicate – cioè rese disponibili anche agli utilizzatori di SBN – con una parsimonia non comune nel panorama internazionale, dall'altro si riutilizzano dati la cui creazione avviene in forma aperta, libera e

collaborativa, ma anche secondo un caratteristico e distintivo approccio dal basso, cioè diametralmente opposto a quello adottato da SBN. Dal punto di vista di Wikidata, si può essere tentati di pensare che si dia più credito ai dati prodotti dai volontari di Wikidata – tra i quali, in ogni caso, ci sono molti bibliotecari – che a quelli prodotti da chi lavora in SBN. Il giusto equilibrio si trova nel riconoscimento del reciproco valore dei dati prodotti e nella valorizzazione piena e senza ostacoli del capitale umano di SBN.

Inoltre, è opportuno segnalare alcuni aspetti problematici del riuso, che probabilmente verranno risolti nei prossimi mesi di evoluzione del servizio.

Per esempio, le date mutuate da Wikidata vengono mostrate con precisione al giorno anche se in Wikidata hanno precisione inferiore; di conseguenza può succedere che venga visualizzata, per esempio, la data di nascita o di morte nella forma “1 gennaio 1962” anche se il dato originale in Wikidata è soltanto “1962” (figura 4).⁴⁹



The screenshot shows the Alphabetica website interface. At the top, there is a navigation bar with 'alphanumeric' logo and links for 'INFORMAZIONI', 'PERCORSI', and 'CONTATTI'. The main content area features a profile for 'Stefania Maria Ciminelli'. It includes a portrait photo, a 'Radar' button, and a bio section with the text: 'Laureata in Filologia romana all'Università di Roma. Attualmente vive e lavora a Barcellona. Traduttrice dallo spagnolo e dal catalano, collabora con diverse case editrici e istituzioni culturali. Nata a Roma.' Below the bio, the birth date is displayed as 'Nascita: Roma, 01/01/1962'. There is also a 'Personaggio PUVV341729' label and a search bar at the bottom right.

Figura 4 – Data di nascita completata automaticamente con giorno e mese fittizi.

⁴⁹ Le date in Wikidata sono modellate come entità (come in IFLA LRM), cosicché oltre al valore si possono registrare anche altri attributi come il calendario (giuliano o gregoriano), la precisione (secolo, decennio, anno, mese, giorno...) e i qualificatori, specialmente utili per le date incerte. Quando si recuperano e riusano questi valori, occorre recuperarne e gestirne anche la precisione.

Un problema non secondario è legato alla dichiarazione della provenienza dei dati. Sebbene la licenza CC0 di Wikidata (Wikimedia Foundation 2021) permetta il massimo riutilizzo dei dati, compreso il riuso senza citare la fonte, la loro inclusione in dati di altra provenienza andrebbe comunque segnalata per diversi motivi: prima di tutto per evitare di lasciare intendere che si tratti di dati disponibili nella voce di autorità o in una sua eventuale estrazione; in secondo luogo, per autotutela, ovvero per non assumersi incondizionatamente la responsabilità di dati prodotti da terzi; infine, per consentire di indirizzare correttamente eventuali suggerimenti o richieste di aggiunte o correzioni, con evidente vantaggio anche per Alphabetica stessa, oltre che per tutti gli altri sempre più numerosi servizi che riusano i dati di Wikidata.⁵⁰ Per esempio, una soluzione pratica è quella del sistema CoBiS⁵¹ e un modello più strutturato di un corretto riuso delle informazioni – con un disclaimer sulla responsabilità – si può leggere sul sito della BBC.⁵²

L'indicazione della fonte dei dati è estremamente rilevante in un contesto di lavoro di autorità e la mancanza di tale indicazione (come si è visto nella tabella 7) compromette l'autorevolezza della voce stessa, a prescindere dalla sua correttezza; ciò vale anche per i «protagonisti» di Alphabetica, naturalmente.

5. Conclusioni

L'analisi dei dati estratti dal VIAF e dal nuovo OPAC SBN ha con-

⁵⁰ Si è discusso del delicato problema del feedback sui dati riusati da Wikidata ai *Data Quality Days, 8-15 September 2021*; si veda in particolare il contributo di Lydia Pintscher e Manuel Merz, *Mismatch finder and beyond: How can we incorporate feedback from our biggest data re-users at scale?*, <https://www.wikidata.org/wiki/Wikidata:Events/Data_Quality_Days_2021>.

⁵¹ Un esempio: <<https://dati.cobis.to.it/agent/95a5yja38danymj1atb5yc1n6c-vk6d8>>. Il bottone “i” accanto a “LINKED DATA” permette di distinguere l'informazione locale dall'informazione esterna.

⁵² <<https://www.bbc.co.uk/editorialguidelines/guidance/feeds-and-links>>.

sentito di evidenziare alcuni aspetti problematici della produzione e della pubblicazione delle voci di autorità di SBN, in relazione alla loro qualità attesa ed effettiva. Tali aspetti problematici, per certi versi, emergono di riflesso anche dalle scelte, coraggiose e condivisibili, che l'ICCU ha compiuto di fatto con la creazione dell'innovativo portale Alphabetica.

5.1 I livelli gerarchici delle voci di autorità e dei catalogatori

L'analisi del dataset POST, che contiene le voci di autorità di livello 95, 97 e 90 pubblicate nel nuovo OPAC SBN, ha mostrato che a dicembre 2021 la qualità dei dati pubblicati non corrisponde nel 43,7% dei casi ai requisiti stabiliti e richiesti a quegli stessi livelli di catalogazione (tabella 3, voci minimali). Una percentuale ancora più alta – il 65,4% in POST – è costituita da voci di livello 90 e superiore prive di qualunque fonte. Può sembrare un dato non molto rilevante; tuttavia, si deve sottolineare da un lato che «documentare le fonti e le decisioni prese in merito ai punti d'accesso» (Gorman 2003, p. 18) è una delle cinque funzioni fondamentali dell'authority control e, dall'altro, che SBN prevede per le voci di autorità di livello 90 o superiore che i dati obbligatori sono tutti quelli desumibili dalla pubblicazione e da fonti esterne.

Se quindi il criterio che ha guidato per trent'anni la pubblicazione di una quantità di voci sproporzionatamente ridotta – se confrontata con quelle delle altre agenzie nazionali che contribuiscono al VIAF e con la quantità di registrazioni bibliografiche pubblicate nell'OPAC – è stato ed è quello di inviare solo voci di alto livello qualitativo, si deve riconoscere che, oggi almeno, non ha più alcuna ragione d'essere e che ha bisogno di essere rivisto.

L'analisi della provenienza per polo (grafico 8) delle voci minimali, cioè quelle inadatte ad essere qualificate di livello 90 o superiore per contenuto e qualità, suggerisce un'altra considerazione in merito al sistema della gerarchia dei livelli dei catalogatori (che è strettamente

legata a quella dei livelli di catalogazione delle voci di autorità). Se in PRE le voci minimali provengono per la maggior parte dal polo CUB (91%), in POST questa percentuale si riduce in modo drastico (a poco meno della metà), a indicare che la responsabilità delle voci minimali, in POST, non è più quasi esclusivamente del polo CUB. Oggi il 34,3% delle voci minimali proviene dal polo CFI (cioè *Polo biblioteca nazionale centrale di Firenze*) mentre solo il 21,5% delle voci minimali proviene da tutti gli altri poli di SBN insieme. I dati mostrano che il problema della bassa qualità delle voci di autorità di alto livello non sembra affatto causato da una presunta scarsa preparazione e capacità dei catalogatori 'di periferia'. Al contrario, anche da questo punto di vista, il sistema della rigida assegnazione gerarchica dei livelli dei catalogatori mostra tutta la sua debolezza. La catalogazione partecipata dovrebbe essere affrontata con una logica diversa da quella top-down, prendendo atto che la necessità di un approccio *dal basso* alla cooperazione nella catalogazione, anche internazionale, è da tempo al centro della discussione e ha una ragion d'essere ancora maggiore in un'ottica di interoperabilità nel web semantico (Willer - Dunsire 2013; Dunsire - Willer 2014; Illien - Bourdon 2014; Riemer 2020).

5.2 *L'apertura dei dati di autorità*

L'applicazione dei livelli gerarchici alle voci di autorità ha comportato la pubblicazione di una quantità di dati di autorità decisamente inferiore rispetto a quella di altre agenzie nazionali e, molto probabilmente, anche alle molto maggiori potenzialità di SBN.

Come si è in parte visto, la politica di apertura dei dati bibliografici e di autorità di SBN attuata dall'ICCU si distingue nettamente da quella di altre agenzie nazionali bibliografiche europee. Per esempio, la British Library invita a «scaricare un'ampia gamma di dataset e di documentazione di supporto da usare per la propria biblioteca o per scopi di ricerca. I dataset contengono la Bibliografia Nazionale Britannica in numerosi formati» (British Library 2022). I dati libera-

mente scaricabili riguardano le monografie, i periodici e le pubblicazioni di prossima uscita, e possono essere ottenuti come LOD (in formato NT e RDF/XML), con file ad aggiornamento mensile e talvolta settimanale. Alcune collezioni tematiche sono scaricabili in formato 'comma-separated value' (CSV) per consentire di analizzare i dati utilizzando strumenti e applicazioni come OpenRefine e i dati vengono prodotti in risposta a richieste provenienti dai ricercatori, per progetti e mostre della British Library (British Library 2022). Tutti i dati sono distribuiti con licenza Creative Commons CC0 1.0 Universal Public Domain Dedication. Una situazione simile vale per la Bibliothèque nationale de France, i cui dati «sono gratuiti e aperti per qualsiasi uso, compreso quello commerciale» nella convinzione che in questo modo si possano raggiungere gli obiettivi di «incrementare la visibilità dei dati della BnF, attraverso una migliore pubblicazione sul web e rendere i dati interoperabili, all'interno e all'esterno dei cataloghi, contribuire alla cooperazione e allo scambio dei metadati creando collegamenti tra risorse strutturate e affidabili, facilitare il riuso dei metadati (sotto licenze aperte) da parte di terzi». ⁵³ Il GND, l'authority file integrato prodotto in modo cooperativo dalla comunità bibliotecaria di lingua tedesca, è accessibile con licenza Creative Commons CC0 in vari formati, come per esempio, in formato MARC 21, MARC21-XML, RDF/XML, RDF (JSON-LD) e RDF (Turtle). ⁵⁴

I dati bibliografici e di autorità di SBN non sono disponibili, al momento, in nessun formato. Per chi scrive, è stato però sorprendente scoprire che i dati statistici su SBN sono disponibili addirittura a pagamento, sia per scopi di ricerca sia quando a richiederli sono i singoli poli SBN. ⁵⁵ Il contrasto con le politiche di apertura messe in atto dalle altre agenzie bibliografiche nazionali europee e dall'ICCD, che fa capo alla stessa struttura ministeriale, non potrebbe essere più

⁵³ <<https://data.bnf.fr/en/about>>.

⁵⁴ <https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html>.

⁵⁵ <<https://www.iccu.sbn.it/it/SBN/monitoraggio-e-statistiche-indice-sbn/>>.

stridente (ICCD 2021).

Ci si augura che non sia questa la ragione dell'indisponibilità, fino a oggi, di dati bibliografici e di autorità prodotti con lo sforzo – professionale e finanziario – collettivo di bibliotecari di tutta Italia nell'arco di decenni. Si spera anche che non sia questa la risposta a quanto si chiedevano i membri del gruppo di lavoro sulle biblioteche digitali dell'AIB nel 2018: «ci domandiamo perché non procedere subito anche alla pubblicazione sul web come Linked data/RDF dei dati gestiti centralmente dalla cooperazione SBN, a maggior ragione sapendo che “l'arricchimento semantico” – ovvero la maggiore usabilità/interconnessione delle informazioni bibliografiche nel web dei dati oltre che nel web dei documenti – sarà il frutto di un impegno graduale e progressivo che conviene mettere in opera prima possibile» (AIB. Gruppo di lavoro sulle biblioteche digitali 2018).

Il nodo centrale da sciogliere per la pubblicazione di tutti i dati di SBN in modo allineato con le altre maggiori agenzie bibliografiche nazionali è quale prospettiva di sviluppo si vuole assumere in un contesto che è profondamente cambiato dalla nascita di SBN.

5.3. Interoperabilità dei dati di autorità

In un contesto di cooperazione internazionale ci si aspetta che ciascuno contribuisca con quanto ha di disponibile e facendo interamente la propria parte, nel presupposto irrinunciabile che il Controllo Bibliografico Universale sia raggiungibile esclusivamente come sforzo collettivo e cooperativo da parte di tutti e in un'ottica di continuo miglioramento reciproco e in una prospettiva permanente di non finito.

L'interoperabilità tramite il collegamento agli identificatori di fonti come VIAF, ISNI e Wikidata dovrebbe essere ottenuta con il loro inserimento «nelle voci di autorità maggiormente controllate [...] ma anche in quelli con un livello di controllo inferiore. Accade, infatti, che nomi presenti in SBN con un livello di autorità non elevato siano presenti anche in Wikidata ma senza quel reciproco collegamento

che aiuterebbe proprio nella definizione della forma corretta» (AIB. Gruppo di lavoro sulle biblioteche digitali 2019).

Nell'ottica del web semantico, che si fonda sull'interoperabilità tra fonti di dati di agenzie diverse, l'invisibilità della maggior parte delle voci di autorità di SBN rappresenta un ostacolo virtualmente invalicabile alla possibilità per altri siti di porsi in relazione con tali voci, con i dati che contengono e con le registrazioni bibliografiche alle quali sono connesse.

Non si tratta tanto di riconoscere, in modo totalmente utilitaristico, che da tali link in ingresso (es. da Wikidata) SBN potrebbe arricchire le proprie voci di autorità sia completando quelle povere di dati sia individuando – senza oneri economici – errori ed omissioni e, così, migliorare progressivamente la loro qualità. Si tratta piuttosto di comprendere fino in fondo la portata di uno dei pilastri del web semantico (Berners-Lee 2006), ovvero che ciò che nel web semantico non è esposto e collegato, è invisibile e, soprattutto, è destinato a rimanere tale.

5.4 Possibili prospettive

La mancata corrispondenza nelle voci di autorità tra le aspettative in termini di qualità legate al loro livello di autorità e la qualità effettiva delle voci stesse ha posto in evidenza problemi che necessitano di una soluzione. Si dovrebbe prendere atto che la qualità non dipende dal livello come è inteso, assegnato e registrato attualmente, e agire di conseguenza, sia a livello pratico sia teorico.

All'atto pratico, si potrebbe procedere a un'analisi accurata della qualità reale delle voci di autorità di livello inferiore a 90 (che non sono pubblicamente disponibili); dall'altro, si potrebbe riclassificare il livello delle voci di autorità in base a criteri diversi da quello del catalogatore che le ha create, come per esempio la loro effettiva completezza e interoperabilità (misurabili anche in modo automatico).

Quanto alla riclassificazione delle voci di autorità, sarebbe quantomeno opportuno procedere in modo generalizzato e automatico

anche a quella di tutte le voci di autorità minimali provenienti dal polo CUB, portandole a un livello al quale possano essere bonificate dall'intera comunità dei bibliotecari di SBN (per esempio a livello 51 o 05) e sanando una situazione che sarebbe stato probabilmente più opportuno non creare nel 1997-1998.

Più in generale, invece, la qualità bassa di un'importante parte del dataset di voci di autorità attualmente pubblicato (POST) suggerisce l'adozione di decisioni che riguardano la politica di apertura di SBN: pubblicare – nell'OPAC SBN e nel VIAF – l'intero dataset delle voci di autorità sia per adeguare quantitativamente SBN ai dataset delle altre agenzie nazionali presenti nel VIAF sia per poter ottenere tutti i vantaggi che offre l'interoperabilità dei dati, in particolare per le voci di autorità di qualità bassa e media. La decisione deve essere supportata dal riconoscimento che voci di autorità che sono di fatto di livello 51 sono già pubblicate e disponibili nel VIAF da anni.

Un'altra coraggiosa decisione di più ampia portata e implicazione sarebbe quella di smantellare il sistema gerarchico che ostacola una reale partecipazione al processo di catalogazione e di orientarsi all'individuazione di forme di collaborazione più ampie e libere. Non si tratta tanto di riconoscere l'evidenza che il sistema piramidale non ha dato gli effetti desiderati e ha contribuito con tutta probabilità a rallentare il processo di produzione di voci di qualità, e nemmeno di prendere atto del paradosso che oggi i catalogatori di SBN possono contribuire a migliorare Alphabetica attraverso l'arricchimento degli item di Wikidata ma, a causa del sistema dei livelli, non possono intervenire direttamente su una consistente parte delle voci di autorità di SBN. Si tratta di prendere esempio da collaborazioni che – coinvolgendo la comunità degli stakeholders – hanno attivato un circuito virtuoso di arricchimento (in termini di identificazione e/o caratterizzazione) della qualità delle voci di autorità, come la partnership tra GND e Wikidata⁵⁶ o, per rimanere in ambito nazionale, l'uso di

⁵⁶ Dal 2005 la comunità degli utenti di Wikipedia in tedesco e, dal 2012, di Wikidata raccoglie nelle sottopagine di <<https://de.wikipedia.org/wiki/Wikipedia:>

OLAF da parte dell'ICCD, da parte del CoBiS o della CEI.⁵⁷

Alphabetica rappresenta un importante passo in avanti per le voci di autorità di SBN per la loro integrazione con quelle di EDIT16 e di Manus Online e per l'interessante e innovativo riuso di LOD da Wikidata. Il riuso di dati di fonti esterne – che è parte della logica dei Linked Open Data – adottato da Alphabetica è un importante esempio della doppia funzione che deve avere oggi un archivio di autorità e della prospettiva più aperta e ampia che si deve adottare nell'individuare le soluzioni catalografiche che si pongono in essere.

Infatti, i dati presenti nella voce di autorità di SBN devono essere visti come strumenti utili per l'*identificazione* dell'entità rappresentata nell'intero dominio dell'universo bibliografico, e non soltanto per la sua *individuazione* all'interno delle voci di autorità locali di SBN. Soltanto in questo modo, infatti, è possibile stabilire che ciò che SBN descrive e 'individua' con, per esempio, "IT\ICCU\CUBV\160180" è la stessa entità che Wikidata descrive e 'individua' con Q59528533, cioè è possibile procedere al processo logico dell'*identificazione* nel web semantico.⁵⁸ La semplice individuazione delle entità rappresentate dalle voci di autorità basata su un criterio di disambiguazione valido solo in presenza di omonimia con altre voci di autorità locali non è più sostenibile in una prospettiva ampia e internazionale – come quella che dovrebbe assumere SBN. La capacità identificativa delle voci di autorità di SBN si deve confrontare con un ecosistema biblio-

GND/Fehlermeldung> segnalazioni di errori, duplicazioni e possibili migliorie alle voci di autorità del GND e ai record bibliografici del catalogo <<https://portal.dnb.de/opac.htm>> della Deutsche Nationalbibliothek (DNB); tali segnalazioni vengono progressivamente vagliate e in massima parte accolte dallo staff della DNB.

⁵⁷ Per una presentazione su OLAF in ICCD, si veda <<https://github.com/synapta/OLAF-ICCD>>; il servizio è disponibile a questo indirizzo: <<https://olaf.beniculturali.it/get/arco/login>>; per l'uso con il CoBiS, si veda <<https://synapta.it/blog/dai-linked-open-data-agli-strumenti-wiki/>>; per OLAF con l'authority file della CEI si veda D'Agnelli - Guerrieri - Rizzo 2021.

⁵⁸ Sulla distinzione tra individuazione e identificazione si veda Serrai 1983, pp. 114-116.

grafico molto più ampio, universale, e per questo c'è bisogno di nuove regole orientate a una reale identificazione.

Una volta che l'entità è stata correttamente identificata, i dati più ricchi che caratterizzano l'entità in altri dataset – come Wikidata per Alphabetica – possono essere riutilizzati per una migliore caratterizzazione dell'entità descritta da SBN: «Dopo avere identificato e collegato l'entità, il processo di creazione dei dati è concluso. Ciò che oggi è definito descrizione, domani [...] sarà il risultato della visualizzazione di un insieme di attributi e relazioni» (Bianchini - Guerrini 2014, p. 17) relativi all'entità. L'entità viene quindi caratterizzata oltre e a prescindere dalla sua funzione 'autorale', che rappresenta solo una delle tante faccette di un 'protagonista', proprio come si vede in Alphabetica. Il riuso in Alphabetica è anche un esempio molto chiaro di ciò che si intende quando si parla della necessità di passare «dal record management al data management, ovvero da un concetto di redazione di un record a quello di identificazione e descrizione di ogni singolo dato» (Guerrini - Bianchini 2016, p. 17) e dimostra come sia ormai necessario prevedere «una netta linea di demarcazione tra linee guida e istruzioni per la registrazione e linee guida e istruzioni per la presentazione dei dati» (Guerrini - Bianchini 2016, p. 18).

La direzione da prendere è quella segnata da Alphabetica. Riconoscere che la comunità di interesse è più ampia della sola comunità bibliotecaria è un passaggio fondamentale per il successo del lavoro di autorità stesso: «è importante comprendere che lo scopo dell'authority control non è il controllo delle funzioni bibliografiche (autore, dedicatario, titolo, soggetto, etc.) bensì di quelle che possiamo definire entità con rilevanza bibliografica, bibliotecaria, museografica, archivistica e delle forme che le designano» (Guerrini - Sardo, p. 11).

Per concludere, il mutamento del contesto esterno, che avviene quando SBN si confronta con il VIAF o, come nel portale Alphabetica, con il web semantico, nel quale i dati possono essere riutilizzati, implica necessariamente un cambiamento di prospettiva che investe anche l'organizzazione stessa di SBN, dalla creazione delle voci alla

loro pubblicazione. SBN deve scegliere tra il rimanere arroccato in una fortezza che sta diventando la sua stessa prigione o tentare di mettere in moto un potenziale umano – interno ed esterno a SBN stesso – che è interessato al suo sviluppo perché lo riconosce come una risorsa irrinunciabile per tutta la comunità bibliotecaria italiana e internazionale.

APPENDICE

Tabella 1. Tipo e numero di campi UNIMARC presenti nelle voci di autorità

campi	note campi in OPAC SBN	PRE			POST			differenze	
		N. totale	N. deduplicati	% su N. duplicati	N. totale	N. deduplicati	% su N. duplicati	diff N. deduplicati	diff % N. deduplicati
001	Identificativo SBN	185132	185132	100,0%	326849	326849	100,0%	141717	0,0%
005	Data di aggiornamento	185132	185132	100,0%	326849	326849	100,0%	141717	0,0%
010	ISNI	74650	74650	40,3%	81644	81644	25,0%	6994	-15,3%
100	<i>General Processing Data</i>	185132	185132	100,0%	326849	326849	100,0%	141717	0,0%
101	Lingua	33902	33902	18,3%	49715	49715	15,2%	15813	-3,1%
102	<i>Nationality of the Entity</i>	82279	82279	44,4%	108305	108305	33,1%	26026	-11,3%
152	Regole di catalogazione	70374	70374	38,0%	118818	118818	36,3%	48444	-1,7%
200	<i>Heading - Personal Name</i>	185132	185132	100,0%	326849	326849	100,0%	141717	0,0%

300	Nota informativa	100436	100436	54,2%	158925	158925	48,6%	58489	-5,6%
400	Forme varianti	41779	23032	12,4%	68782	40657	12,4%	27003	0,0%
410	Forme varianti [sic]	17	15	0,0%	35	31	0,0%	18	0,0%
510	Vedi anche	288	140	0,1%	501	252	0,1%	213	0,0%
801	<i>Originating Source</i>	185132	185132	100,0%	326849	326849	100,0%	141717	0,0%
810	Fonti	140037	75448	40,7%	163538	91844	28,1%	23501	-12,6%
815	<i>Source Data Not Found</i>	44216	22558	12,2%	46680	23678	7,2%	2464	-4,9%
830	Altre fonti	47668	47668	25,7%	63794	63794	19,5%	16126	-6,2%
999	[?]	0	0	0,0%	20110	18053	5,5%	20110	5,5%

Tabella 2. Provenienza delle voci di autorità in base al polo SBN

polo	PRE	POST	diff	polo	PRE	POST	diff	polo	PRE	POST	diff
CFI	45514	107479	61963	BRI	194	277	83	CAM	11	27	16
CUB	74130	75631	1466	UMC	172	270	98	IST	21	26	5
UFI	1179	20220	19041	CSA	176	268	92	ISU	14	22	8
BVE	9638	19217	9577	RMR	159	236	77	RL1	8	21	13
PAL	2100	16923	14822	PIS	165	229	64	UDA	12	20	8
RAV	8275	11463	3187	LZ1	27	225	198	UBS	9	16	7
IEI	1213	8886	7673	PA1	72	223	151	SNT	8	14	6
SBL	5697	8295	2596	ANA	171	217	46	LEK	4	13	9
MIL	4726	7006	2280	UFE	115	200	85	LI3	4	13	9
LO1	4834	6582	1746	RT1	75	174	99	PCM	2	11	9
TO0	4203	6224	2020	UTO	69	159	90	CER	5	10	5
UBO	3698	5355	1657	PAV	66	154	88	VBA	5	10	5
SBN	2258	3357	1098	AUF	145	149	4	ATM	7	8	1
PUV	1700	2614	912	LI2	14	138	124	TES	6	8	2
RML	991	2131	1140	PBE	59	134	75	EVE	2	7	5
VEA	1337	2126	788	BCT	37	130	93	MES	2	7	5
NAP	1228	1789	561	SBT	71	126	55	CBR	1	6	5
TSA	1162	1664	502	RCA	66	124	58	ABR	1	4	3

VIA	1027	1500	473	RLZ	64	121	57	ART	3	4	1
CAG	610	1493	883	RMB	60	112	52	BZM	2	4	2
MOD	931	1345	414	FOG	60	101	41	PTA	1	4	3
RMS	863	1322	459	TER	58	92	34	TA1	2	4	2
CNC	207	1209	1002	RER	61	89	28	ARE	0	3	3
MUS	635	1100	462	RMC	54	84	30	PIA	2	3	1
DDS	346	960	614	BIA	56	78	22	RE2	0	3	3
RMG	490	839	349	LUA	44	78	34	TPS	1	3	2
UMI	351	543	192	MOL	66	78	12	AGR	0	2	2
LIA	405	509	104	MO1	58	75	17	BAT	1	2	1
REA	335	495	160	PMI	41	72	31	GMP	0	2	2
USM	301	487	186	BAS	35	55	20	MSE	1	2	1
GEA	327	474	147	TO1	15	47	32	UBG	0	2	2
URB	302	472	170	BMT	24	46	22	USS	0	2	2
PAR	278	417	139	SCM	38	43	5	BMC	0	1	1
FER	311	404	93	UAN	32	43	11	FRI	0	1	1
AQ1	245	330	85	SIP	23	41	18	PBI	0	1	1
INT	155	323	168	SGE	14	37	23	RNT	1	1	0
BA1	107	294	187	UPG	28	37	9	TOTA- LI	185132	326849	141664
LIG	183	288	105	UR1	20	34	14				

Tabella 3. Voci di autorità minimali

Ordine cronologico					Ordine decre- scente (POST)			
anno di aggiorn.	PRE	POST	diff		polo	PRE	POST	diff
1985	0	1598	1598		CUB	61757	63051	1294
1986	1	3889	3888		CFI	4110	49093	44983
1987	2	6965	6963		UFI	26	16928	16902
1988	2	5809	5807		IEI	39	6276	6237
1989	4	6832	6828		BVE	513	2830	2317
1990	0	9498	9498		RAV	161	807	646
1991	0	7767	7767		CNC	0	627	627
1992	7	5323	5316		MIL	134	397	263
1993	43	7814	7771		SBL	186	356	170
1994	123	5277	5154		TO0	133	351	218
1995	80	3491	3411		UBO	189	312	123
1996	34	1195	1161		LO1	112	287	175
1997	37516	40010	2494		LZ1	0	176	176
1998	22962	24097	1135		RML	26	166	140
1999	1340	2063	723		NAP	57	154	97
2000	703	1354	651		PUV	69	149	80
2001	1037	1394	357		VIA	18	93	75

2002	373	968	595	BA1	1	93	92
2003	410	961	551	VEA	17	91	74
2004	411	722	311	RMG	46	69	23
2005	491	802	311	PAL	20	65	45
2006	372	694	322	LIA	37	57	20
2007	268	758	490	CAG	25	54	29
2008	138	375	237	TSA	23	44	21
2009	649	743	94	REA	21	43	22
2010	68	156	88	RMS	21	36	15
2011	167	196	29	GEA	18	34	16
2012	171	204	33	BCT	0	32	32
2013	33	108	75	MOD	17	31	14
2014	10	45	35	FER	13	23	10
2015	9	47	38	AQ1	6	21	15
2016	79	254	175	URB	14	16	2
2017	85	325	240	UM1	9	16	7
2018	50	191	141	CSA	13	16	3
2019	59	288	229	PIS	7	13	6
2020	153	460	307	UMC	5	12	7
2021	48	263	215	<i>altri</i>	55	117	62
TOTALI	67898	142936	75038	TOTALI	67898	142936	75038

Tabella 4. Aggiornamenti e creazioni per anno

anno	PRE			POST			diff		
	PRE-agg.	PRE-creati	PRE-mai agg.	POST-agg.	POST-creati	POST-mai agg.	diff agg.	diff creati	diff mai agg.
2021	73348	181	181	82990	1333	1333	9642	1152	1152
2020	4467	546	427	9343	1560	1176	4876	1014	749
2019	1746	445	254	5912	2085	1662	4166	1640	1408
2018	1275	2676	267	7088	4099	1420	5813	1423	1153
2017	2245	3013	879	7763	4111	1627	5518	1098	748
2016	1398	3655	616	5596	4692	1289	4198	1037	673
2015	2475	2930	816	5951	4650	1420	3476	1720	604
2014	1753	2878	1085	3958	3983	1414	2205	1105	329
2013	2427	4193	1731	4163	5362	2019	1736	1169	288
2012	4905	7068	4117	7065	8231	4451	2160	1163	334
2011	4072	6383	3214	6614	7756	3866	2542	1373	652
2010	4002	7224	2881	6667	8718	3447	2665	1494	566
2009	2985	5454	2090	4248	6862	2494	1263	1408	404
2008	2271	6089	446	4133	7723	1058	1862	1634	612

2007	2474	7914	665	3813	9594	1302	1339	1680	637
2006	2433	7778	682	3660	8930	1031	1227	1152	349
2005	2263	6860	648	3405	8094	1048	1142	1234	400
2004	969	3815	381	1691	4901	746	722	1086	365
2003	887	2727	321	2142	4489	981	1255	1762	660
2002	1181	3077	453	2242	4636	1123	1061	1559	670
2001	1771	3480	559	3304	5667	1393	1533	2187	834
2000	871	2779	634	2828	5665	2021	1957	2886	1387
1999	1583	3488	1450	3069	5918	2473	1486	2430	1023
1998	23197	28718	23090	25035	31261	24354	1838	2543	1264
1997	37713	46117	37660	41293	50910	40572	3580	4793	2912
1996	41	1160	27	2140	4303	1517	2099	3143	1490
1995	98	1458	80	4747	7681	4043	4649	6223	3963
1994	215	4017	139	6836	14764	6134	6621	10747	5995
1993	47	1782	34	8392	13334	8138	8345	11552	8104
1992	7	1823	7	5753	11888	5647	5746	10065	5640
1991	0	1081	0	8257	12032	8257	8257	10951	8257
1990	0	871	0	9959	13246	9959	9959	12375	9959
1989	4	991	4	7222	10363	7222	7218	9372	7218
1988	3	892	3	6194	9000	6194	6191	8108	6191
1987	4	875	4	7469	10570	7469	7465	9695	7465

1986	1	476	1	4175	5940	4175	4174	5464	4174
1985	1	204	1	1732	2478	1732	1731	2274	1731
1979 ⁵⁹	0	1	0	0	1	0	0	0	0
1978	0	1	0	0	2	0	0	1	0
1976	0	2	0	0	2	0	0	0	0
1901	0	10	0	0	15	0	0	5	0
TOTALI	185132	185132	85847	326849	326849	176207	141717	141717	90360
			46.4%			53.9%			

⁵⁹ L'anno 1901 è un evidente errore; probabilmente lo sono anche gli anni 197x.

Tabella 4bis. Aggiornamenti e creazioni per anno nel polo CUB

	PRE			POST			diff		
anno	PRE-agg.	PRE-creati	PRE-mai agg.	POST-agg.	POST-creati	POST-mai agg.	diff agg.	diff creati	diff mai agg.
2021	10012	0	0	10074	0	0	62	0	0
2020	305	0	0	303	0	0	-2	0	0
2019	235	0	0	243	0	0	8	0	0
2018	174	118	1	183	120	4	9	2	3
2017	185	114	0	187	115	3	2	1	3
2016	150	96	1	153	96	0	3	0	-1
2015	118	42	0	121	42	0	3	0	0
2014	147	55	0	147	55	0	0	0	0
2013	71	54	0	71	54	0	0	0	0
2012	57	42	0	61	42	1	4	0	1
2011	52	42	0	58	42	0	6	0	0
2010	79	41	1	83	41	1	4	0	0
2009	102	48	0	107	48	0	5	0	0
2008	137	101	0	255	100	2	118	-1	2
2007	122	147	0	123	146	0	1	-1	0
2006	214	231	5	218	230	7	4	-1	2

2005	256	193	3	258	193	3	2	0	0
2004	63	74	1	64	74	1	1	0	0
2003	54	57	0	55	57	0	1	0	0
2002	101	124	8	101	124	13	0	0	5
2001	83	46	0	87	46	4	4	0	4
2000	243	296	239	251	304	247	8	8	8
1999	952	1146	952	982	1182	982	30	36	30
1998	22658	26468	22658	23111	26953	23111	453	485	453
1997	37495	44454	37495	38269	45421	38269	774	967	774
1996	0	3	0	0	3	0	0	0	0
1995	2	4	2	2	4	2	0	0	0
1994	41	91	41	43	94	43	2	3	2
1993	3	6	3	2	5	2	-1	-1	-1
1992	6	12	6	6	13	6	0	1	0
1991	0	0	0	0	0	0	0	0	0
1990	0	2	0	0	2	0	0	0	0
1989	4	6	4	3	6	3	-1	0	-1
1988	3	6	3	3	7	3	0	1	0
1987	4	5	4	4	5	4	0	0	0
1986	1	2	1	1	2	1	0	0	0
1985	1	4	1	2	5	2	1	1	1
TOTALI	74130	74130	61429	75631	75631	62714	1501	1501	1285

Tabella 5. Presenza e distribuzione di date nei campi UNIMARC 200\$f e 300

	PRE	% sul totale (185132)	% sui 300	POST	% sul totale (326849)	% sui 300
campo 200\$f (tutti)	29884	16,1%		50590	15,5%	
di cui campo 200\$f senza campo 300	2662	1,4%		6514	2,0%	
di cui campo 200\$f ma senza date nel campo 300 prima di //	3728	2,0%	3,7%	12186	3,7%	7,7%
di cui campo 200\$f e date nel campo 300 prima di //	23494	12,7%	23,4%	31890	9,8%	20,1%
campo 300 (tutti)	100436	54,2%		158925	48,6%	
di cui campo 300 con date prima di //	86694 ⁶⁰	46,8%	86,3%	125383 ⁶¹	38,4%	78,9%

⁶⁰ Di queste, 1091 voci di autorità hanno il campo 300 con date prima di // e, dopo, nessun'altra informazione.

⁶¹ Di queste, 6243 voci di autorità hanno il campo 300 con date prima di // e, dopo, nessun'altra informazione.

Bibliografia

- Agenjo-Bullón - Hernández-Carrascal 2018 = Xavier Agenjo-Bullón, Francisca Hernández-Carrascal, *Registros de autoridades, enriquecimiento semántico y Wikidata*, «Anuario ThinkEPI», 12 (2018), p. 361-372.
- AIB. Gruppo di lavoro sulle biblioteche digitali 2018 = AIB. Gruppo di lavoro sulle biblioteche digitali, Osservazioni preliminari al Progetto SRI, AIB-WEB, <<https://www.aib.it/struttura/commissioni-e-gruppi/gruppo-di-lavoro-biblioteche-digitali/2018/82796-osservazioni-preliminari-al-progetto-sri/>> (28.1.2022).
- AIB. Gruppo di lavoro sulle biblioteche digitali 2019 = AIB. Gruppo di lavoro sulle biblioteche digitali, Cosa SBN dovrebbe fare, AIB-WEB, <<https://www.aib.it/attivita/2019/82799-cosa-sbn-dovrebbe-fare/>> (28.1.2022).
- Allison-Cassin - Scott 2018 = Stacy Allison-Cassin, Dan Scott, *Wikidata: a platform for your library's linked open data*, «Code4Lib Journal», 40 (2018), <<https://journal.code4lib.org/articles/13424>> (31.1.2022)
- Association of Research Libraries 2019 = Association of Research Libraries, *ARL White Paper on Wikidata. Opportunities and Recommendations*, 2019.
- Bargioni - Bianchini - Pellizzari 2022 = Stefano Bargioni, Carlo Bianchini, Camillo Carlo Pellizzari di San Girolamo, *Dati e script di «Le voci di autorità dei nomi di persona in SBN e Alphabetica : problemi e prospettive»*, «Zenodo.org», (2022), DOI: 10.5281/zenodo.5893399.
- Berners-Lee 2006 = Tim Berners-Lee, *Linked Data - Design Issues*, 2006, <<http://www.w3.org/DesignIssues/LinkedData.html>> (31.1.2022)
- Bianchini - Bargioni - Pellizzari 2021 = Carlo Bianchini, Stefano Bargioni, Camillo Carlo Pellizzari di San Girolamo, *Beyond VIAF. Wikidata as a complementary tool for authority control in libraries*, «Information Technology and Libraries», 40 (2021), 2, p. 1-31, <<https://ejournals.bc.edu/index.php/ital/article/view/12959>> (31.1.2022).
- Bianchini - Guerrini 2014 = Carlo Bianchini, Mauro Guerrini, *Introduzione*

- a RDA. Linee guida per rappresentare e scoprire le risorse*, Milano, Editrice Bibliografica, 2014.
- Bianchini - Sardo 2022 = Carlo Bianchini, Lucia Sardo, *Wikidata : a new perspective towards universal bibliographic control*, «JLIS.it», 13 (2022), 1, p. 291-311.
- Biblioteca nazionale centrale 1968 = Biblioteca nazionale centrale, *CUBI: bibliografia nazionale italiana 1886-1957 : catalogo cumulativo del Bollettino delle pubblicazioni ricevute per diritto di stampa dalla Biblioteca nazionale centrale di Firenze*, Nendeln (Liechtenstein), Kraus Reprint, 1968.
- Biblioteca nazionale centrale 2002 = Biblioteca nazionale centrale, *CUBI: bibliografia nazionale italiana 1886-1957 : catalogo cumulativo del Bollettino delle pubblicazioni ricevute per diritto di stampa dalla Biblioteca nazionale centrale di Firenze*. Milano, Informazioni editoriali, 2002.
- British Library 2022 = British Library, Free dataset downloads, The British Library, <<https://www.bl.uk/collection-metadata/downloads>> (30.1.2022).
- Caffo 2013 = Rossella Caffo, *SBN tra presente e futuro*, «Biblioteche oggi», 31 (2013), 7, p. 8-13.
- Coladangelo - Ransom 2021 = L. P. Coladangelo, Lynn Ransom, *Semantic Enrichment of the Schoenberg Database of Manuscripts Name Authority through Wikidata*, <<https://www.researchgate.net/publication/358246309>> (30.1.2022)
- Commissione permanente per la revisione delle regole italiane di catalogazione 2009 = Commissione permanente per la revisione delle regole italiane di catalogazione, *Regole italiane di catalogazione*. REICAT, Roma, ICCU, 2009.
- D'Agnelli - Guerrieri - Rizzo 2021 = Francesca Maria D'Agnelli, Claudia Guerrieri, Maria Teresa Rizzo, [et al.], *L'autorità work nel sistema dei beni culturali ecclesiastici*, «DigItalia», 2021 (2021), 2, p. 91-107.
- Deana 2021 = Danilo Deana, *Forse non tutti sanno che ... Un'analisi a tutto campo in occasione dei 35 anni di SBN*, «Biblioteche oggi», 39 (2021), 1, p. 29-40.

- Dostál 2021 = Vojtěch Dostál, Bringing Czech authority files into 21st century: Integration with Wikidata. 2021. <<https://www.youtube.com/watch?v=JKqZTVisHC4>> (31.1.2022).
- Dunsire - Willer 2021 = Gordon Dunsire, Mirna Willer, *The local in the global: universal bibliographic control from the bottom up*, in IFLA WLIC 2014, Lyon, France, 2014, <<http://library.ifla.org/817/>> (31.1.2022).
- Gorman 2003 = Michael Gorman, *L'autorità di controllo bibliografico in ambiente elettronico*, in *Authority Control. Definizione ed esperienze internazionali. Atti del convegno internazionale, Firenze, 10-12 febbraio 2003*, Firenze, Firenze University Press ; Associazione italiana biblioteche, 2003, p. 17-26.
- Guerrini 2003 = *Authority control, definizione ed esperienze internazionali: atti del Convegno internazionale, Firenze, 10-12 febbraio 2003*. A cura di Mauro Guerrini, Firenze : Roma , Firenze University Press ; Associazione italiana biblioteche, 2003.
- Guerrini 2020 = Mauro Guerrini, *Dalla catalogazione alla metadattazione: tracce di un percorso*, Roma, AIB, 2020.
- Guerrini - Bianchini 2016 = Mauro Guerrini, Carlo Bianchini, *Manuale RDA: lo standard di metadattazione per l'era digitale*, Milano, Editrice Bibliografica, 2016.
- Guerrini - Sardo 2003 = Mauro Guerrini, Lucia Sardo, *Authority control*, Roma, Associazione italiana biblioteche, 2003.
- ICCD 2021 = ICCD, Dati aperti, <<http://www.iccd.beniculturali.it/it/per-condividere/dati-aperti>> (31.1.2022).
- ICCU 2017 = ICCU, Norme per il trattamento di informazioni e dati comuni a tutte le tipologie di materiale, <<https://norme.iccu.sbn.it/>>, <https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Codici/Informazioni_di_servizio/Livelli_di_autorit%C3%A0> (10.4.2021).
- ICCU 2019 = ICCU, *Linee guida per la compilazione delle registrazioni di autorità dei nomi di persona in SBN*, Roma, ICCU, 2019.
- ICCU 2020a = ICCU, Guida alla catalogazione in SBN – Musica. Appendice XV – Livelli di catalogazione, <<https://norme.iccu.sbn.it/>>, <https://norme.iccu.sbn.it/index.php?title=Guida_musica/Appendici/Appendi

ce_XV> (27.1.2022).

ICCU 2020b = ICCU, Norme per il trattamento di informazioni e dati comuni a tutte le tipologie di materiale. 1.2 Qualificazioni, <<https://norme.iccu.sbn.it/>>, <https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file/Nomi/Registrazione_di_authority/Qualificazioni> (31.1.2022).

ICCU 2020c = ICCU, Norme per il trattamento di informazioni e dati comuni a tutte le tipologie di materiale. 1.3 Datazioni, <<https://norme.iccu.sbn.it/>>, <https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file/Nomi/Registrazione_di_authority/Qualificazioni> (31.1.2022).

ICCU 2020d = ICCU, Norme per il trattamento di informazioni e dati comuni a tutte le tipologie di materiale. 1.4 Nota informativa, <<https://norme.iccu.sbn.it/>>, <https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file/Nomi/Registrazione_di_authority/Fonti> (31.1.2022).

ICCU 2020e = ICCU, Norme per il trattamento di informazioni e dati comuni a tutte le tipologie di materiale. 1.5 Fonti, <<https://norme.iccu.sbn.it/>>, <https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file/Nomi/Registrazione_di_authority/Fonti> (31.1.2022).

ICCU 2021 = ICCU, Alphabetica e il nuovo ecosistema digitale dei servizi bibliografici nazionali, <<https://www.iccu.sbn.it/it/eventi-novita/novita/Alphabetica-e-il-nuovo-ecosistema-digitale-dei-servizi-bibliografici-nazionali/>> (28.1.2022).

IFLA Cataloguing Section e IFLA Meetings of Experts on an International Cataloguing Code 2017 = IFLA Cataloguing Section e IFLA Meetings of Experts on an International Cataloguing Code, *Dichiarazione di principi internazionali di catalogazione (ICP)*. Edizione 2016 con piccole correzioni, 2017, IFLA, 2017.

Illien - Bourdon 2014 = Gildas Illien, Françoise Bourdon, *A la recherche du temps perdu, retour vers le futur: CBU 2.0*, in *IFLA WLIC 2014*, Lyon, France, 2014, <<http://library.ifla.org/956/>> (31.1.2022).

Linked Data for Production 2020 = Linked Data for Production, Wikidata

as a hub for identifiers, Google Docs, <https://docs.google.com/presentation/d/1jWz3_nCf5rdd-7ejETGlfv99UV2PnD1v/edit?usp=embed_facebook> (27.8.2020).

Moi 2020 = Alessandra Moi, *When Linked Data is (not) enough : cataloguing Tools between Obsolescence and Innovation*, «JLIS : Italian Journal of Library, Archives and Information Science = Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione», 11 (2020), 2, p. 1-19.

Permanent UNIMARC Committee 2022 = Permanent UNIMARC Committee, UNIMARC Authorities (3rd ed.) Updates, <<https://www.ifla.org/>>, <<https://www.ifla.org/unimarc-updates/unimarc-authorities-3rd-edition-with-updates/>> (25.1.2022).

Petrucciani - Turbanti 2021 = Alberto Petrucciani, Simona Turbanti, *Manuale pratico di catalogazione: principi, casi e problemi*, Milano ; Roma, Editrice Bibliografica : Associazione italiana biblioteche, 2021.

Riemer 2020 = John Riemer, *The Program for Cooperative Cataloging & a Wikidata Pilot*. 2020.

Serrai 1983 = Alfredo Serrai, *Ricerche di biblioteconomia e bibliografia*, Firenze, La Nuova Italia, 1983.

Tillett 2006 = A Virtual International Authority File (VIAF), in *IFLA Cataloguing Principles: Steps towards an International Cataloguing Code*, 3, Walter de Gruyter – K. G. Saur, 2006, p. 90-116.

Wikimedia Foundation 2021 = Wikimedia Foundation, Terms of Use, <https://foundation.wikimedia.org/wiki/Terms_of_Use/en> (31.1.2022).

Willer - Dunsire 2014 = Mirna Willer, Gordon Dunsire, *Bibliographic information organization in the semantic web*, Oxford, Chandos Pub, 2013.

Abstract

L'articolo presenta i risultati dell'analisi qualitativa e quantitativa delle voci di autorità di nomi di persona di SBN, confrontando due dataset, uno basato sui dati inviati al VIAF nel 2021 (PRE) e l'altro sui dati prelevati dall'OPAC SBN poco dopo la pubblicazione del portale Alphabetica (POST). Lo scopo dello studio era quello di stabilire se la qualità delle voci di autorità corrisponda effettivamente ai requisiti indicati da SBN per le voci di livello 90 e superiore. Inoltre, lo studio si è proposto di verificare se il sistema di produzione e pubblicazione delle voci di autorità, basato sui livelli di autorità delle voci e dei catalogatori, sia ancora attuale o meno, anche alla luce di Alphabetica. L'analisi evidenzia che una percentuale rilevante delle voci di autorità (PRE e POST) non risponde ai requisiti, che la scelta di pubblicazione delle voci in base al livello di autorità non è adeguata agli standard e alle scelte di altre agenzie bibliografiche nazionali e che le innovazioni introdotte in Alphabetica suggeriscono l'opportunità di un cambio di prospettiva nell'organizzazione del lavoro di autorità di SBN.

ICCU; SBN; Authority Control; nomi di persona; Wikidata; VIAF; UBC; Universal Bibliographic Control

The article presents the results of the qualitative and quantitative analysis of SBN authority records for personal names, by the comparison of two datasets, the former based on data sent to VIAF in 2021 (PRE) and the latter on data taken from the SBN OPAC shortly after the publication of Alphabetica, the new ICCU portal (POST). The aim of the study was to establish whether the quality of the authority records actually corresponds to the requirements specified by SBN for the entries of level 90 and above. In addition, the study aimed to verify whether the system of production and publication of authority records, based on the authority levels of entries and of cataloguers, is still current or not, also with

respect to the novelties introduced by Alfabetica. The analysis shows that a relevant percentage of authority records (both in PRE and in POST) do not meet the requirements, that the choice of publication of records according to the levels of authority does not match to the standards and to the choices of other national bibliographic agencies, and that the innovations introduced in Alfabetica suggest a change of perspective in the organisation of SBN authority work.

ICCU; SBN; Authority Control; personal names; Wikidata; VIAF;
UBC; Universal Bibliographic Control