JLIS.it

The future of bibliographic services in light of new concepts of authority control

Michele Casalini[(a)]

a) Casalini Libri, http://orcid.org/0000-0003-4643-8895

ABSTRACT

Over the last three decades, a number of major changes in the field of cataloguing have led to the definition of new forms of authority control. The introduction of FRBR and of IFLA LRM have been followed by continuing studies, including, more recently, the implementation of linked data in library catalogues, as well as improvements to data models in order to ensure the broadest possible interoperability among systems. A new approach to authority control and its connected services can be based on the combination of manual and automatic processes of data validation and enrichment, together with the use of knowledge bases as authoritative sources. This will also grant wider data interoperability, opening up a new level of cooperation among the international institutions and organisations concerned with the dissemination of knowledge.

KEYWORDS

Authority control; Bibliographic metadata; Cataloguing ecosystem; Linked Open Data (LOD).

JLIS.it

New era, new needs

The surprising technological innovations and significant changes in the field of cataloguing have opened the doors to new horizons that see machines play a proactive and effective role in the decoding and sharing of bibliographic metadata. It is thanks to these new advances in technology that it is possible to overcome linguistic barriers and venture beyond purely bibliographic fields.

In the new, digital, era, the fast growing quantity of – sometimes perishable – data, requires those who operate in the cultural heritage sector to carry out a task of fundamental importance: to react to the need for an authority control that "guarantees" the homogeneity, stability and formal quality of access entries as an integral operation within the cataloguing ecosystem. It is a technique influenced by the technology of its time as well as by the standards and cataloguing conventions in use in the contexts of linguistic, cultural and disciplinary specializations.

The concept of traditional authority records is also evolving, in order to comply with the new open philosophy of data sharing and reuse. The transition from the concept of record to that of entity, in the context of the semantic web has forced a rethinking not only of data, but also of the organization and management of authority control itself. Previous discussion on whether authority control should be based centrally or locally will be subject to transformation, as the focus shifts from a rigid conception to a more flexible notion of entity identification and relationships between entities. The direction in which this field is advancing has already been partially outlined in the enlightening profile within the international conference proceedings of the Authority Control in Organizing and Accessing Information, held in Florence in 2003.

In combination with the technological developments that support this cause, since then it has been necessary to radically rethink the conceptual models of data interpretation. The transition from the Functional Requirements for Bibliographic Records (FRBR) to the IFLA Library Reference Model (IFLA LRM) has propelled the international community towards a new modeling of bibliographic levels, linked together by primary relationships and accompanied by further relationships with entities and properties.

The Bibliographic Control function continues to be valid today but shifts the focus to a global level, supporting growing international cooperation, which is facilitated now by the interoperability of the data models. The contribution of a heterogeneous group of organizations concerned with the dissemination of knowledge also promotes cooperative authority control, with collaboration and mutual assistance among actors of various kinds; by comparing and integrating their data with those of others, the information they convey will be more complete and more reliable.

Organizations such as libraries, archives, museums, but also publishers and providers will engage with each other in the generation of new data and the discovery of new resources, crossing the boundaries of specific domains to create data enrichment opportunities that would previously have been unthinkable. The theme of facilitating the sharing of authoritative sources through persistent and reconciled resources for the benefit of a more precise and wider discoverability was also addressed from 2016 to 2018 by the Institute of Museum and Library Services (IMLS) funded National Strategy for Sharable Local Name Authority National Forum (SLNA-NF).

To implement the interoperability of metadata, it has become necessary to create a new conformation and structure, so that each entity can be identifiable by a single and unambiguous name or code that is used by all agencies creating bibliographic metadata: the Uniform Resource Identifier

JLIS.it

(URI) avoids the ambiguity of using natural language. Data structured according to the Resource Description Framework (RDF) data model, in contrast to the traditional record-based approach, focuses on individual metadata declarations represented by triples of data in the subject-predicate-object form.

These triples can become quads, containing the provenance information necessary to take advantage of data enriched through authoritative sources, while maintaining local preferences for the labeling and display of data through customizable application profiles.

Statements can be combined and matched from many different sources to link different standards and models as well, such as Resource Description and Access (RDA) and the Bibliographic Framework Initiative (BIBFRAME). The schemas expressed in the RDF linked data structure allow other communities to reuse the data in their own environments.

New models into practice: the Share Family

Following the path that was initiated, developed and progressively applied by the Library of Congress with BIBFRAME, encouraged by the vision of the Linked Data for Production (LD4P) projects promoted by Stanford University, and in light of the extensive and exciting possibilities offered by new technologies and data models, in 2016 a community-driven initiative, the Share Virtual Discovery Environment (Share-VDE or SVDE), emerged, with the aim of putting the new developments into practice and applying them to an entity based discovery environment for the benefit of libraries and their users.

As one of the founding organizations of the Share-VDE initiative, and in its role as a bibliographic agency, Casalini Libri has been, and continues to be active in testing linked data technologies for libraries together with its technological arm and sister company @Cult.

Building on the experience of all involved parties and drawing from it, one of the aims of SVDE has been to develop innovative approaches for the authority control of bibliographic records and for the creation and improvement of authority control procedures, providing new authority services to libraries and supporting their transition to linked data.

The starting point for this evolution, like the initiative itself, stems from the real and emerging needs of the library community, more specifically the need for libraries to receive constantly updated information on their bibliographic and authority records from authoritative sources, both in MAchine Readable Cataloguing (MARC) format and in the BIBFRAME linked data structure. The services designed and the underlying technological infrastructure are the result of the development of new Linked Open Data (LOD) technologies influenced by the direct input of the various Share Family collaborative environments involving national and research libraries. These processes facilitated the experimentation in the creation and handling of linked data entities, but also provide direct interaction with operational library systems that will coexist for a long time in both MARC and RDF.

The overall goals include the enrichment of MARC records with identifiers from external sources (e.g. ISNI, VIAF); the reconciliation and clusterization of entities and the publication of the Cluster Knowledge Base (CKB); the conversion from MARC to RDF using the BIBFRAME vocabulary together with other ontologies; batch/automated authority services, data updating and data

JLIS.it

dissemination procedures; a manual entity management tool (J.Cricket); the publication of data on an entity-oriented user interface (www.svde.org).

An active role in determining directions and priorities is played by the Share-VDE Advisory Council and by the various Working Groups, one of which is dedicated specifically to the Authority/Identifier Management Services (AIMS).

Flexibility in handling and in profiling the integration of data from external sources is a crucial aspect for the processes involved, as each institution may have a different list of priorities.
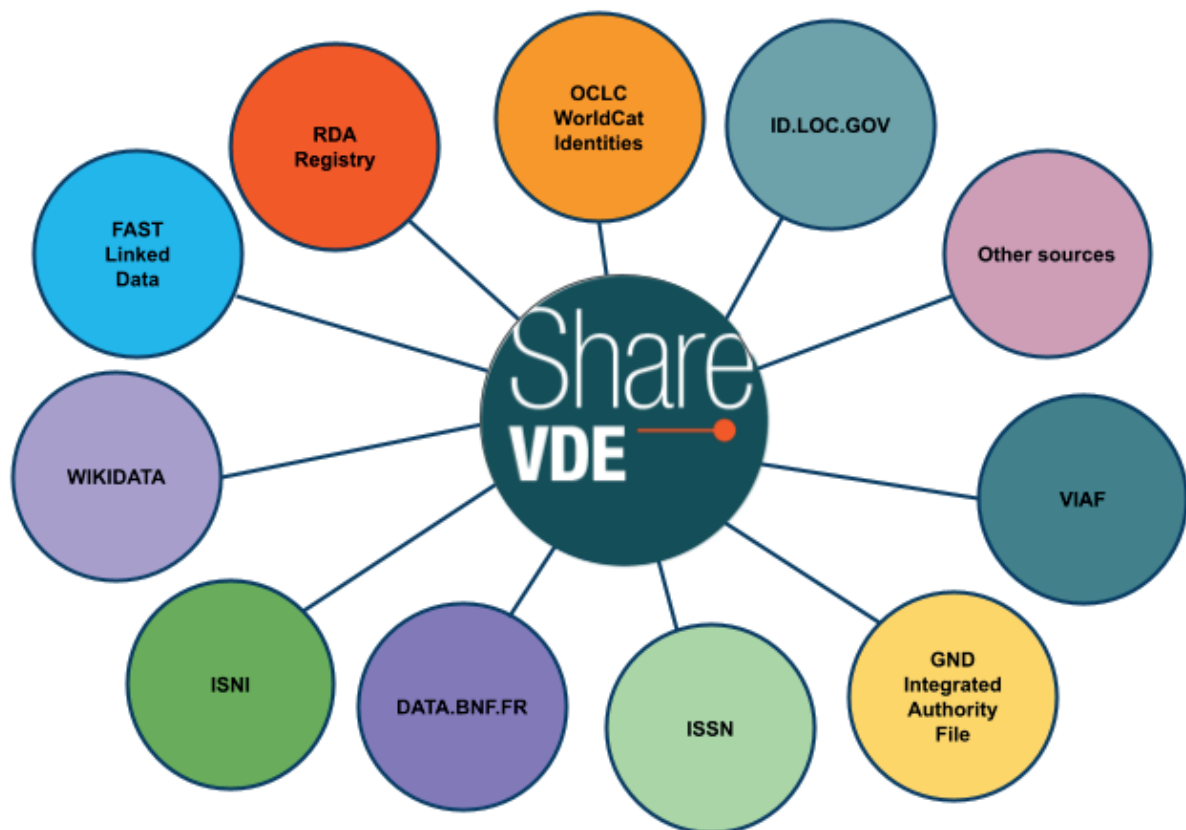

Fig. 1. Integration of data from external sources

Wikidata is an example of interaction that allows for sources to be searched and for SVDE data to be enriched with Wikidata entity information – and vice versa – as SVDE has a property in Wikidata for the author ID.

From a technological viewpoint, Application Programming Interfaces (APIs) architecture simplifies interconnections, reusability, sustainability and scalability, opening the window to an open world.

JLIS.it

The challenge of data models interoperability

The challenge of data interoperability among systems, which is indispensable in order to bring into practice implementations at a wide scale, however, requires comparisons among data models and mapping that maintain the granularity of information. With this aim, on June 10th 2020 the SVDE Entity Identification Working Group approved the SVDE Opus class, also a BIBFRAME Work, as the SVDE Work is too.
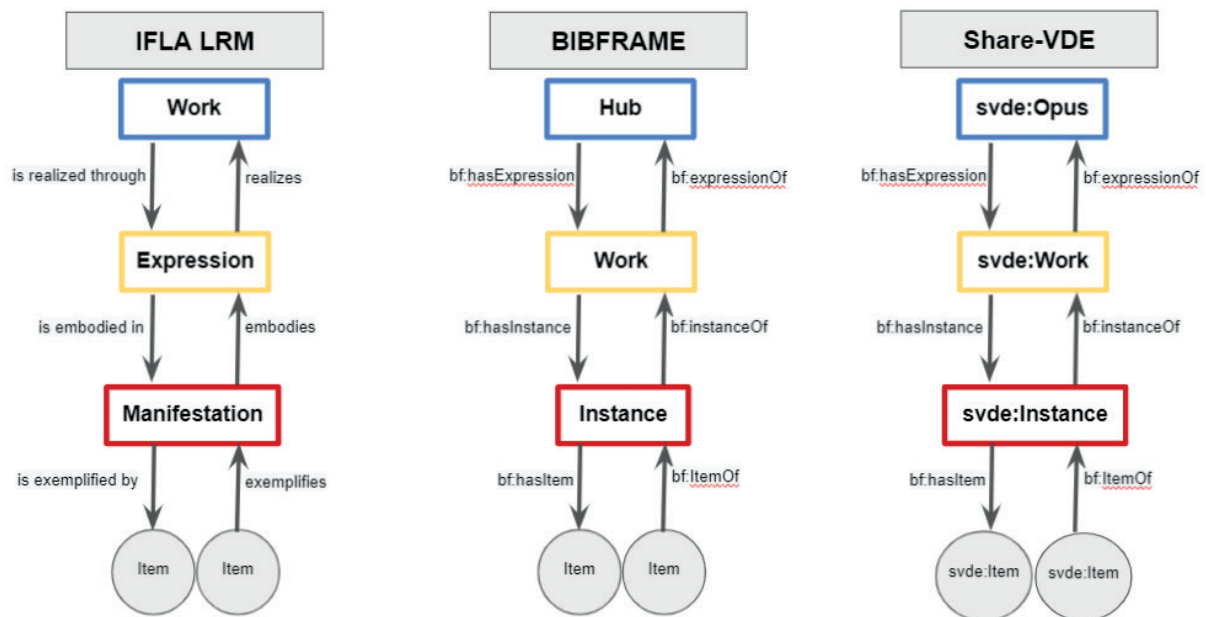

Fig. 2. IFLA LRM, BIBFRAME, Share-VDE Data model comparison. Further details on this structure can be found on https://wiki.svde.org

It is important to highlight that SVDE is trying to practically reconcile an approach to entity modelling that is "North-American oriented" (BIBFRAME) with a "Europe-centric" approach (IFLA LRM). This reconciliation aims to create a flexible crossover between different cataloguing practices, thus allowing it to adapt to different data modelling contexts that cannot be confined in restricted geographic, linguistic, cultural borders. Such trait d'union has been facilitated by the entry in SVDE of European libraries such as the National Library of Norway, the National Library of Finland and the British Library.

We have now mentioned several of the pillars that relate to one another and create the broader ecosystem with the Share-VDE Cluster Knowledge Base, named Sapientia, in the center.
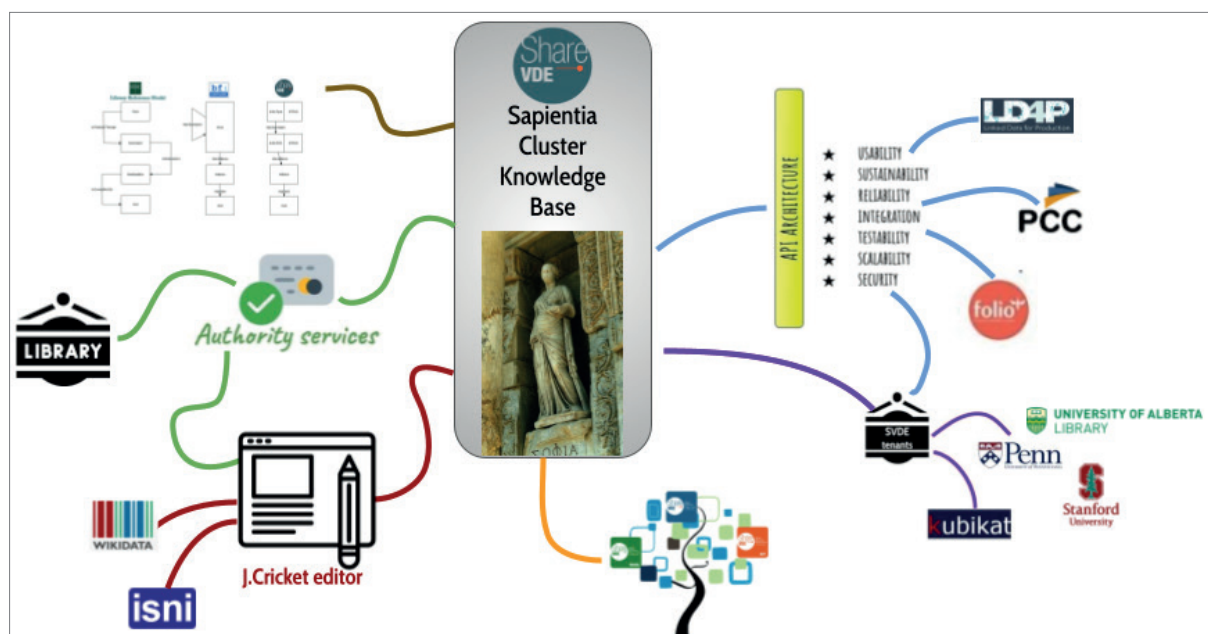
Fig. 3. Sapientia Cluster Knowledge Base ecosystem

Around it, in a clockwise direction, are the APIs layer (back-end to interact with other environments), the skin and tenant architecture, the authority flows handling both automatic and manual processes, traditional data flows in MARC or entity oriented ones, and factors regarding interoperability with other data models.

The CKB, Sapientia, represents the ambition to build an authoritative knowledge base with the tools to improve it, with 1) mechanical algorithms but also 2) an editor, J.Cricket, which allows a transversal community including data producers (e.g. librarians) to collaborate jointly in order to raise the level of quality of the data offered. The combination between the CKB, produced through automated processes that aggregate data from many different sources, and its editor could represent a sort of ideal union of the Virtual International Authority File (VIAF) world and the Wikidata community; that is, on the one hand the quality and authority proposed by the VIAF model, and on the other hand the open and cross-domain approach of the Wikidata model, carrying its vision of a collaborative tool open to a wider community of users. Therefore Sapientia, with its control and editing tool, J.Cricket, opens up to a vision that combines authority with the ability to interact: openness and control.

Authority processes in the new bibliographic ecosystem

In the new context, highlighted above, the aim of authority control services is to facilitate the control and standardization of bibliographic data. This is achieved through a combination of automatic and manual processes that make it possible for local cataloguing practices to be integrated within a global, participatory dimension.

Automatic authority control operations allow a high level of productivity, while manual operations

JLIS.it

guarantee a higher level of quality: for each context, therefore, the best balance of the two must be identified.

The automated processes are divided according to whether the library is handling a record-based ILS, an RDF-based system, or a hybrid one. These are some of the processes involved. Variables may be the frequency of the dataflows and whether the library also holds the authority file locally.

- The MARC record validator, the MARC corrections for errors and obsolete forms, and the matching/enrichment with profile sources compose the record-based scenario.
- Access point enrichment (including Series and Subjects), matching, import and interaction with the Sapientia Cluster Knowledge Base are necessary for interaction with the RDF-based systems.

In both cases the processes are enabled through Representational State Transfer (RESTful) modules of the LOD Platform, which provide bibliographic, authority and full text search services with entity detection and identification including relator terms capabilities.

The manual processes are divided into two groups: the operational tasks of the original cataloguing processes to validate and enrich metadata elements, and the editing processes to enhance the common Cluster Knowledge Base.

The first set of manual processes can be characterized by the following operations:

- Authority control of the access points of bibliographic records for similar matches and non matches, including the checking, validation and reconciliation of imported URIs.
- Manual enrichment of entity Work and Agents (including Publishers).
- The creation of original authority records; Casalini Libri already does this for the International Standard Name Identifier (ISNI), in compliance with its role as an ISNI Registration Agency (Personal Names, Corporate Names, including publishers, Meeting Names and Uniform Title) and sends reports to ISNI in the case of duplicate records existing for a single entity or of relationships with incorrect titles.

These operations are enabled through the dedicated URI Registration Platform.

The second set of manual processes employs the CKB editor, known as J.Cricket, as the instrument for the direct management of entities represented in RDF. The new application, dedicated to the editing of SVDE community data, is a collaborative tool that not only makes it possible to validate automatic matches that the clustering procedure identifies as uncertain, but also allows library professionals to merge, split or create new clusters autonomously. Conceived as a collaborative editing environment, the application foresees different levels of access and interaction with the data, enabling users to manually create, modify and reconcile clusters of the entities saved in the CKB.

The entities present in Sapientia and managed by J.Cricket are based – conceptually – on the SVDE four labelled entity model (Opus, Work, Instance, Item). The clusters to be modified, automatically and manually, are: Opus, Works and Agents. The next achievement will be to treat the Instance as an entity.

These two examples show how the interaction between J.Cricket and Wikidata IDs is envisioned, from both perspectives.

The scope and capacity of the CKB editor will be extended over time to include the management of authority services for libraries, with quality control procedures for data. With this twofold purpose, not only will J.Cricket facilitate the creation and handling of linked data entities within SVDE, but it will also provide direct interaction with library systems both in MARC and RDF formats.

JLIS.it

Interconnections both with the Sinopia linked data cataloguing module of the LD4P initiative and with the data from the Program for Cooperative Cataloguing (PCC) will be the primary testbed for J.Cricket to prove its ability to act as a pivotal tool between traditional MARC-based cataloguing workflows and innovative linked data processes.

Linked data for the future

The linked data paradigm is laying the groundwork for new level of cooperation among international organizations to create new bridges across the library, archives and museums domain, which serve to increase discoverability for students, scholars and the wider community, to reveal data that would be otherwise remain hidden, to contribute to promoting a culture of openness towards knowledge, and to foster – on the one hand – the preservation of existing knowledge and – on the other – the progress undertaken by younger generations.

Initiatives such as LD4P, Share-VDE and others with each of the institutions involved, the leading role of many national libraries, of cooperative programs such as the PCC, and of other players in the information chain are crucial not only for bringing these developments into practice, but for reaching the critical mass of implementation across cultural heritage collections.

In conclusion, the present challenge for the organizations that have bibliographic control at heart is not only to facilitate libraries in handling constantly updated information on their records or datasets from authoritative sources, but also to improve the level of collaboration between actors of differing nature, thanks to data interoperability, in a future vision of authority control which is more open and cooperative on a global scale.

Acknowledgements

JLIS.it

References

Taylor, Arlene G, and Barbara B. Tillett, eds. 2004. *Authority Control in Organizing and Accessing Information: Definition and International Experience*. New York: The Haworth Information Press. https://doi.org/10.4324/9780203051092

Bibliographic Framework Initiative (BIBFRAME). Accessed June 3, 2021. https://www.loc.gov/bibframe

IFLA Study Group on Functional Requirements for Bibliographic Records. Functional Requirements for Bibliographic Records (FRBR): Final Report. 1998. Munchen: K.G. Saur. https://www.ifla.org/publications/functional-requirements-for-bibliographic-records

IFLA Library Reference Model (LRM). 2017. Den Haag: International Federation of Library Associations and Institutions. https://www.ifla.org/publications/node/11412

International Standard Name Identifier (ISNI). Accessed June 3, 2021. https://isni.org

Linked Data for Production (LD4P). Accessed June 3, 2021. https://wiki.lyrasis.org/display/ld4lGW

MARC 21 Format for Authority data. Accessed June 3, 2021. https://www.loc.gov/marc/authority

Casalini, Michele, Chiat Naun Chew, Chad Cluff, Michelle Durocher, Steven Folsom, Paul Frank, Janifer Gatenby, Jean Godby, Jason Kovari, Nancy Lorimer, Clifford Lynch, Peter Murray, Jeremy Myntti, Anna Neatrour, Cory Nimer, Suzanne Pilsk, Daniel Pitti, Isabel Quintana, Jing Wang, and Simeon Warner. National Strategy for Shareable Local Name Authorities National Forum [SLNA-NF]: White Paper. 2018. https://hdl.handle.net/1813/56343

Program for Cooperative Cataloging (PCC). Accessed June 3, 2021. https://www.loc.gov/aba/pcc/

Resource Description and Access (RDA). Accessed June 3, 2021. https://www.rdatoolkit.org

Share Virtual Discovery Environment (Share-VDE or SVDE). Accessed June 3, 2021. https://wiki.svde.org

International Federation of Library Associations and Institutions. *UNIMARC manual: authorities format*. 2009. 3rd ed. Munchen: K.G. Saur. https://www.ifla.org/publications/ifla-series-on-bibliographic-control-38

Virtual International Authority File (VIAF). Accessed June 3, 2021. https://viaf.org