Academic Research in Educational Sciences
ISSN: 2181-1385
DOI: 10.24412/2181-1385-2022-2-961-973
Volume 3 | Issue 2 | 2022
Cite-Factor: 0,89 | SIS: 1,12
SJIF: 5,7 | UIF: 6,1

# AUTHORITY CONTROL, NEW LIBRARY STANDARDS, AND THE SEMANTIC WEB

**Gulchiroy Erkinovna Ziyodullaeva**
Tashkent University of Information Technologies named after Mukhammad al-Khorazmi
E-mail: gulchiroy.ziyadullayeva.81@bk.ru

## ABSTRACT

Linked data principles benefit from graph theory concepts. Graph theory itself models the relationships between objects.

Traditional authority files managed by local or global communities ofpractice provide excellent input into the creation of entities described by the metadata representing library collections. Linked data theorists advocate the importance of making assertions about real world objects, whereas library authority files document the authorized form of a name or title for entry into a filing system or machine-readable index. An entity that provides the full context for a person is required for an entity of value on the web.

The paper proposes a tool that generates authority files to be integrated with Linked Data by means of learning rules.

**Keywords:** Authority control software, Linked data, Records exchange, Semantic Web, Interoperability, Cataloguing.

## INTRODUCTION

The need to improve interoperability within the World Wide Web gave rise to the development of the Semantic Web, which in turn led to the appearance of many new ways to control and standardize the description of documents, solve problems surrounding diverse indexing systems, and improve the interoperability of records (SKOS, SIOC, Dublin Core, FOAF, etc.). Authority control is a global problem, affecting not only libraries but organizations of all kinds. Publication of authority data on the Web in a heterogeneous or arbitrary way produces inefficiency in information retrieval and creates complications when attributing authority to a given work.

Libraries and organisms of international prestige such as the Library of Congress, the Bibliotheque Nationale de France and IFLA unite forces to share data and thereby contribute to authority

control. These bodies acknowledge the fact that the information exchange protocols on the Web are insufficient means of controlling authority in the catalogues and systems of library management, since not all countries and organizations can deploy the same level of technological or human resources in their cataloguing efforts, making cooperative cataloguing nearly impossible.

The OCLC, IFLA and the US Library of Congress have fueled initiatives for authority control by sharing the records of various cataloguing agencies. Fruit of this work is *the Virtual International Authority File* (VIAF), which has meant advances in the construction and generation of authority entries, though it has not reached all the major information institutions at the international level (Bourdon and Zillhardt, 1997).

## MATERIAL AND METHODS

The proposed tool combines the potential of Linked Data with the strengths of the protocols of the Semantic Web, to which we added working rules based on the experience of librarians in creating access points (Tillett, 2004). Two variables were used for evaluation: capacity to create records, and quality in record creation. The indicators assessed for each variable are defined in the evaluation section.

## AUTHORITY CONTROL: GENESIS

Authority control is a matter that has exacted the effort of generations of librarians and cataloguers. The need to uniformly record information on each author included in a catalogue is addressed in work and research stemming from several international organizations. Thanks to the efforts of the IFLA, LITA and ALA, among others, the community of cataloguers has come to adopt standards for generating catalogue entries in a homogeneous manner. The overall goal is to unify the criteria of diverse library consortia and create bibliographic records guided by universally described and accepted cataloguing standards. A brief outline of the development of authority control would include the following landmarks:

- The need for authority control is made explicit, and the Name Authority Cooperative (NACO) comes to light within the US Library of Congress. In Asia, the Hong Kong Chinese Authority Name (HKCAN) is established. This meant recognition of the issue in just two organisms worldwide — far, however, from the syndetic goals set forth by Charles Cutter in the 19th century (Cutter, 1986).
- Lubetzky (Lubetzky, 1969) improves the search and

retrieval of authored works in bibliographic records, eliminating the deficiencies that interfered with the retrieval and location of authors in a catalogue.

- Ritvars Bregzis (Bregzis, 1982) creates the ISADN (International Standard Authority Data Number) to overcome difficulties when retrieving bibliographic records with works relative to a given author and with works recorded under a uniform title. The ISADN became an important tool to connect bibliographic records of diverse authors and multiple levels of citation, using fixed numbers for each author or associated work. Soon, the ISADN was key to operations in US libraries.
- The Control Interest Group (ACIG), created under the ALA in 1984, can be seen as a quantum leap in the development of cataloguing activity. Its research into authority control in the United States focused on the use of catalogues in public and specialized libraries to make recommendations for the uniform treatment of authority.
- The guidelines known as *Functional Requirements for Bibliographic Records* (FRBR) and the AACR2 facilitate the use of several tags for information retrieval (Pino, 2004; Danskin, 1996; 1998), among them: Find: to locate an entity or entities through attributes and relations; Identify: confirms the correspondence between the records searched for and the ones actually located; Obtain: facilitates acquisition of an entity or item; and Navigate, which helps both the cataloguer and general public to navigate through materials related with their search in the document collection.
- MARC appears on the international scene, along with national formats such as IBERMARC and UKMARC, plus versions like UNIMARC and MARC 21. Thus, bases are established for worldwide cooperation among entities and the interchange of records.
- IFLA proposes the ISAN, International Standard Authority Number, and the codes of the ISO International Standard Text Code family (ISTC) to identify works and expressions, thereby facilitating information exchange within organizations on an international level.

Procedures and standards largely described in the 1960s and the 1990s therefore served as the starting point for the development of authority control in the digital realm.

**Internet, the Semantic Web and VIAF: union of procedures for authority control.**

The Semantic Web, introduced by Berners-Lee, Hendler and Lassila, paved the way for Web interoperability and the existence of formats for information exchange and ontologies

(Berners-Lee *et al.*, 2001), elements that bring authority control into an increasingly flexible arena. Below we highlight some elements that, in our view, must be taken into account when building tools for authority control on the Semantic Web.

**1.** The growth of digital libraries sets the stage for the Z39.50 protocol to enhance interoperability. Thanks to the strong points of this protocol, copy cataloguing is easier, as is the reutilization of bibliographic records.

**2.** The so-called FRANAR *(Functional Requirements and Numbering of Authority Records*) appear on the scene, integrating elements of access and processing in the realm of authority records (Bourdon, 2002). The remarkable aspects of FRANAR stem from its capacity to agglutinate, from a single record, elements specifying information about the author in all its dimensions. FRANAR heightens the potential for information retrieval from bibliographic records by means of the following options: **Search** (for authors or entities), **Identify** (authors or entities), **Control** (creating mechanisms for authority control) and **Relate** (showing resources relative to an item). Moreover, it manages records by: **Process, Sort** and **Display.** Both FRBR and FRANAR are standards that allow for creating authority records with diverse links, facilitating the establishment of different types of relations. In this paper, we use them as models to structure authority records.

Deserving special mention are RDA and VIAF, in view of their importance for the development of our application, AUTHORIS.

RDA (Resource Description and Access) appears in 2005 as a standard for data description and exchange (US RDA Test Coordinating Committee, 2011). It includes the *Functional Requirements for Bibliographic Records* (FRBR) and the *Functional Requirements for Authority Data* (FRAD), which filled the gap of a descriptive cataloguing standard allowing for:

- Quick insertion into the dynamic context of libraries and other producers and users of information.
- Flexible relations between or among entities.
- Greater use and data management in conjunction with digital media.
- More precise description (beyond the possibilities of existing formats) of printed monographs and serial publications.
- Greater ease when using cataloguing metadata in Linked Data operations, aside from ensuring data tagging to facilitate exchanges among bibliographic and non-bibliographic organizations.
- Flexibility, departing from the exclusive focus on Anglo-American rules, meaning metadata can be easily reutilized.

RDA is nourished by new IFLA principles and stands as a noteworthy cataloguing advance toward object-oriented databases. The Web environment called for improving aspects such as the recognition of bibliographic contents, the use of bibliographic data by search engines, and the inclusion of user needs in the processes of describing resources.

The development of VIAF is the result of what we explained earlier in section 3, the genesis of authority control in digital environments.

According to Boeris, VIAF has become the venue for a vast community of libraries and agencies to reconfigure their bibliographic data so as to better serve users of different languages (Boeris, 2011).

The VIAF initiative is organized by OCLC, in charge of revising and comparing records containing author names and their assignments, as well as the documents existing in national bibliographic records and in the WorldCat. Each VIAF record is generated with information drawn from the comparison of records, and it includes underlying data from authority catalogues and bibliographic catalogues.

Nonetheless, the VIAF objective cannot be achieved by all libraries worldwide due to the fact that:

1. It does not take into account the cognitive factors of the user, and at times describes matters in "librarian" terminology, opaque for the average Web user.
2. Librarians cannot modify or improve their catalogue entries.
3. Cataloguing work is not organized all over the world in such a way that would permit the formation of work groups to attend VIAF in all regions.
4. There is a need for software that would monitor errors, homogenizing Web searches and authority entries, aside from being able to recognize and group elements under a uniform title. Those existing at present respond to the demands of librarians, not those of other organisms where bibliographic information is commonly used.

In order to implement the VIAF postulates and authority control within the Semantic Web, or in Linked Data, a new dimension for processing information on the Web is needed (Greenberg and Robertson, 2002). To proceed in this direction, the bibliographic records of libraries must be correctly mixed with the protocols of the Semantic Web and similar organizations. In this sense, we can underline the efforts invested by the Library of Congress from 2005 onwards, with their contributions from MARC through XML, MODS (Metadata Object Description Schema) and MADS (Metadata Authority Description Schema).

A further essential element for adapting to the Semantic

Volume 3 | Issue 2 | 2022
Cite-Factor: 0,89 | SIS: 1,12
SJIF: 5,7 | UIF: 6,1

Web lies in the authority control by information organizations (Qiang, 2004; Taylor, 1999) and, more specifically, their conversion to RDF (Resource Description Framework) (Miles *et al.*, 2005). This is a complex task that sometimes entails huge programming costs, as organizations may have records in diverse bibliographic formats.

Such obstacles do not detract from the semantic wealth of these vocabularies, with potential for constructing linguistic structures that improve the ratio of recall and precision in the retrieval of information from the Web. The lexical structures involved enable diverse users to engage in communication, overcoming linguistic barriers when searching for or retrieving information.

Human and technological resources on the library horizon show great variation. Many libraries, for instance that of the University of North Carolina, devise systems with a high syndetic level; whereas others might not even have generated catalogues in a first level of bibliographic description. Such great differences across the board underscore the significance of even attempting a global system for authority control. Along with the phenomenon of diverse technological capacities, we encounter a proliferation of tools produced through the very development of the Web, by non-librarian organizations (Harper and Tillett, 2007). The appearance of exchange protocols on the Semantic Web (FOAF, SKOS, Dublin Core), their linkage and simplification, are making it possible to generate records manageable for any user or level of technology.

## AUTHORIS

### *2.1. Conception and underlying methodology*

AUTHORIS aspires to facilitate the processing of authority data in a standardized fashion, following the principles of Linked Data. Unlike systems such as Virtual Open Access in Agriculture and Aquaculture Repository (VOA3R), AUTHORIS can be used by all sorts of bibliographic agencies, publishing companies, associations or libraries. This software was produced by members of the Department of Information and Communication of the University of Granada, Spain, whose collaborative goal is to develop a platform for information exchange and authority encoding. The tool has taken into account the obsolescence of MARC and its substitution by FRBR and RDA.

The software features multiple functionality, to favor transformation of bibliographic records and to determine uniform entries for corporate authors as well as individual ones. Each

function is derived from the facilities of Linked Data, the potency of the algorithms for converting bibliographic data, and the precision of learning rules. AUTHORIS relies on Drupal, a CMS (Content Management System) that works with semantic information, and contains the protocols of Dublin Core, SIOC, SKOS and FOAF.

Drupal is a content management system created by Dries Buytaert in 1999 and developed under GNU license two years later. Creating a website in Drupal consists of combining several "blocks" in order to adapt the site functionality to specific needs. It furthermore provides a Content Management Framework (Byron, Berry and Bondt, 2012). Information is stored in a relational database (it works with MySQL, PostgreSQL, SQLite and others) using PHP programming language.

Drupal permits publication of data in RDF format, or alternative formats such as N-Triples, JSON, XML, RSS 1.0 and Turtle. It handles the URIs of the published RDF data, and accommodates an Endpoint SPARQL for consulting the data. The RDF fields and Namespaces can be personalized.

Hence, it is very flexible software, while featuring a robust security mechanism and sufficient online documentation. A major strength of this CMS resides in the possibility of adding modules. In this case, we opted to generate a new one that would surpass the capacity of the Biblio module, designed to process bibliographic references.

Although there are numerous methodologies for developing information services, in the case of AUTHORIS we opted for the approach devised by Garrido and Tramullas, which highlights the most important aspects for a proper design of digital information (Garrido and Tramullas, 2006). The actions carried out to create the service were:

1. Study the needs of potential users of the service.
2. Develop automata and program text-processing models.
3. Test-run the program.
4. Train staff and create the software documentation.
5. Trial stage of the software.
6. Publication and diffusion of the service.

Below we go over the main features of the system:

- **Authentication:** a user is assigned a role by the system administrator. Each user has specific tasks that include record conversion, automatic cataloguing and information searches.
- **Search for Information:** all users —even those not authenticated by the system— can consult the access points of

each record. To this end, they dispose of a system for information search and retrieval based on Semantic Web technologies capable of filtering the records of over 200 information entities (libraries, archives, virtual libraries, etc.). The search for information facilitates data retrieval by author, title, subject, and the combination of Boolean operators for more complex searches, in addition to having a system for document clustering. The search results can be obtained in XML, RDF, FOAF, MARCXML, RDA, FRAD or FRBR format, among others. It is also possible to visualize the information entities where a given record is located, obtain an image of the author's works, and view his/her main collaborators or the bookshops and publishing houses that commercialize the works. The user who searches for authority information can moreover find suggestions about which entries are more complete, or which is recommended for reference.

- **Conversion of files:** converting files is a real strong point of AUTHORIS. A recorded user can import and export files to diverse formats (XML, RDF, FOAF, MARCXML, RDA, FRAD and FRBR). The imported records may be from libraries, publishing companies or other organizations. Within this section one can create new authority entry, by means of learning rules used only in the event that one wishes to produce a uniform title or introduce a new authority. The user merely has to select the nationality of the author (if an individual; or the entry data in the case of a corporative author), and the system automatically assigns the correct entry using the learning rules declared for this purpose. If the system already has the authors introduced by another agency, it will suggest the best entry based on cases previously described and stored. Furthermore, the user can select the output format to export the authority record created. The authority rules are also used to group the uniform titles most used by agencies included in the project. File conversion makes it easier to obtain data in RDF generated by another 200 databases that are not associated with the AUTHORIS project, including DLBP, VIVO, Dbpedia, IEEE, PubChem and Chem2Bio2RDF.
- **Automatic cataloguing:** by means of a Z39.50 client, one can search for and retrieve catalogue records from other libraries in MARC, MARC 21 and MARC XML to complete the cataloguing data in question. This option is proposed for users who do not have RDA-based catalogues and whose national catalogues mainly use MARC 21.


Figure 1. Interface for automatic cataloguing by AUTHORIS

• **Editorial visualization:** thanks to this option, it is possible to extract information referring to a work and its author by means of online consultation of various publishers or libraries.

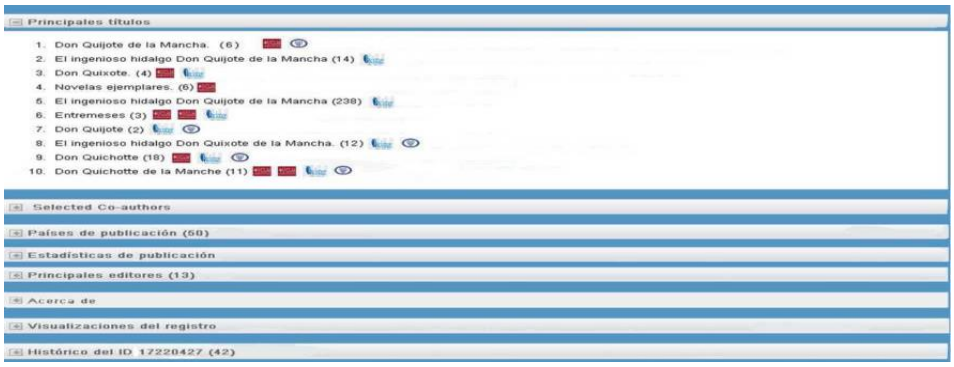


Figure 2. Data extracted from different publishing companies regarding works by Cervantes

The first action undertaken by the system is to access and navigate through authority records generated by bibliographic agencies, publishers and professional associations. Library records are created by means of author entry rules. Those of professional associations and publishers are generated using specific standards adopted by each organization. Once the stage of access and navigation has finished, the system uses rules to generate authorized access points by transforming records into bibliographic standards that may be exchanged with other agencies or organizations. Under AUTHORIS, access entails three elements:

• **Authorized heading:** this is the standard entry generated according to Anglo-American Cataloguing Rules.

• **Heading generated by a non-bibliographic entity:** entries made by other organisms (not libraries, archives or information agencies) that process entries for persons and

institutions.

- **Variant headings:** the different forms of subject headings that may be seen in particular cases.

To create an authority record, this software explores and locates records in diverse formats used in other entities, in view of:

- **Authority records:** they present all the information about corporate or individual authors that an information organization may possess. In the authority records, all author entries are standardized. They constitute the foundation of the system and serve to make comparisons between records, to determine which is the best or most complete.
- **Databases:** they contain information on authors and their publications, and include those of DBLP Computer Science Bibliography, Web of Science, Scopus, etc.
- **Virtual publishers and bookshops:** these hold a large portion of the international publishing output worldwide. They take in the titles of works as well as the names of authors.
- **Pages of organizations:** generally speaking, they provide information about persons associated with an organization.

*Authority rules*

The rules for authority conversion with which we work in this module were derived from a need for systems to be more scalable and professional (Miles *et al.*, 2005), and from daily practices of persons dedicated to information use and management. To transform the data into a bibliographic format, the system converts all sources into XML. This XML file then undergoes transformations into RDF, XML, MARC 21, MARC, UNIMARC or FRBR.

The authority rules were created bearing in mind AACR2 and RDA standards, and regularity in subject headings in libraries all over the world. To formulate these rules, more than 5000 cases (from different organizations) were used. In this way, part of the authority work was "intelligent", carried out automatically, gathering cases where specific conditions were set out for:

- Personal author.
- Corporate author.
- Non-governmental institutions.
- Government institutions.
- International institutions.
- Religious institutions.

Volume 3 | Issue 2 | 2022
Cite-Factor: 0,89 | SIS: 1,12
SJIF: 5,7 | UIF: 6,1

- Events.
- Subtitles.
- Parallel title.
- Alternative title.
- Statement of responsibility.
- Shared responsibility.
- Mixed responsibility.

The system of rules for authority control is in charge of standardizing the entries under uniform criteria, which are the same as those used by information organizations worldwide. The rules serve to instrumentalize a learning system that selects an entry by weighing its quality, in view of the following parameters:

Rules for the control of individual and corporate authors: they feed on the Anglo-American Cataloguing Rules and serve to standardize the entries. The system has 300 cases for each rule, and a deciding algorithm that chooses the most complete entry for each heading. Figure 4 illustrates one of the rules employed to process government authority entries.

**Rule 3 Heading for embassies**

• If the heading for an embassy or delegation has no country for which it is accredited, select as valid heading the one held by the country before which it is accredited. Cases:

- Mexico (embassy) Peru 5 correct
- Mexico Peru embassy
- Embassy. Mexico and Peru
- Selection:
- If the main entry contains the country, the embassy and the place, it is selected as the most complete entry.
- location 1 + embassy 2 + (place 2) make 5

Figure 3. Rules employed to process government authority entries

❖ **Rules for assigning a uniform title**: these rules function through a comparison of cases. An agent locates the title of the work to be catalogued, and selects cases that repeat a similar title, to then assign the most used title.

❖ **Rules to assign a standard number to each author:** each author name is standardized using the publisher number or the ISAN, if recorded by some library. Otherwise, the number identifying the author in the publishing company is used. If none of these options is feasible, a code is assigned to record the existence of the author in question.

❖ **Decision-making rules:** implying the three above rules, they are also referred to as "if, then" rules, in agreement with the weight given to each author record in each particular rule, selecting the most adequate procedure.

## CONCLUSIONS

The development of authority control faces new challenges in the Semantic Web. The need to facilitate interoperability and connection among non- bibliographic and bibliographic entities is one promising area to be implemented by the designers and developers of future cataloguing and authority control systems.

The tool presented in this paper is not meant to be a panacea in this sense. Indeed, the authors hold that such applications will not be needed in the library world of the future, given that integrated library systems will eventually adopt the technologies used by the Linked Data movement as a global reality, not just an extension toward RDA cataloguing.

In the meantime AUTHORIS paves a path toward terrain where future information systems might be oriented: the integration of data and contents. It is our understanding that all efforts should lead to promoting the creation and use of Open Data.

The possibilities lent by the Open Data movement are an aid in accessing remote as well as homogeneous and standardized data, which may be reutilized in other processes. For this reason it is important that libraries open their databases and contribute to the use of formats by non-bibliographic entities.

The model for authority control presented in AUTHORIS is flexible and inclusive. Although certain solutions for international authority control have been put forth previously, progress is slow and tends to be limited to National Libraries or very large ones. Non-library organizations (particularly publishing companies and bookstores) are left out in the cold, despite their capacity to generate vast volumes of authority entries that might be used not only for authority control, but also for information exchange.

AUTHORIS, our proposal in this direction, offers institutions the means of sharing data in a global manner with a high level of stability, moreover helping to detect records that are duplicated, while contributing to lexical disambiguation and data enrichment.

## REFERENCES

1. Berners-Lee, T., Hendler, J., Lassila, O. (2001), "The Semantic Web", *Scientific American*, Vol. 284, pp. 34-43.

2. Berners-Lee, T (2009), *Linked Data*, available at: http://www.w3.org/DesignIssues/LinkedData.html (accessed 7 December 2012).

3. Boeris, C. (2011), "Algunas reflexiones sobre el control de autoridades en Argentina", in *VII Encuentro Internacional y III Nacional de Catalogadores,* Biblioteca Nacional de la Republica, Buenos Aires, Argentina, 23-25 November 2011.
4. Bourdon, F. (2002), "Functional requirements and numbering of authority records (FRANAR): to what extent can authority control be supported by technical means?", *International Cataloguing and Bibliographic Control,* Vol. 31 No. 1, pp. 69.
5. Bourdon, F., Zillhardt, S. (1997), "AUTHOR: vers une base europeenne de notices d'autorite auteurs", *International Cataloguing and Bibliographic Control,* Vol. 26 No. pp. 34-37.
6. Bregzis, R. (1982), "The Syndetic Structure of the Catalog Authority Control: The Key to Tomorrow's Catalog", in *Proceedings of the 1979 Library and Information Technology Association Institutes*, Phoenix, AZ, pp. 24.
7. Byron, A., Berry, A. and Bondt, B. (2012), *Using Drupal: Choosing and Configuring Modules to Build Dynamic Websites.* 2nd ed., O'Reilly Media, Sebastopol.
8. Cutter, C. A. (1986), *Rules for a Printed Dictionary Catalogue,* Government Printing Office, Washington, D. C.
9. Danskin, A. (1998), "International Initiatives in Authority Control", *Library Review,* Vol. 47 No. 4, pp. 200-205.