

# Analysis of the methodologies created for libraries in the construction of authority control

**Anamarys Toscano Menocal**

Universidad Central “Marta Abreu” de Las Villas, Cuba.

**Email:** anamaryst@uclv.cu

**María del Rosario Fernández Rodríguez**

Universidad Central “Marta Abreu” de Las Villas, Cuba.

**Email:** mfrdriguez@uclv.cu

## ABSTRACT

Libraries are immersed in a process of structural change, where the implementation of new competency practices means bringing them into conformity with the needs and demands that arise in the information and knowledge society. The growth of technologies and information sources, which are scattered on the Internet, implies the task of obtaining information from a large number of sources available on the Web. The realization of information queries from heterogeneous and geographically dispersed sources requires the implementation of methodologies and regulations, as well as the construction of tools. The research aims to analyze the methodologies created so far in libraries, for the realization of authority control. It was necessary to use both theoretical and empirical methods in the collection of information. In the last 10 years, it is evident in the analyzed methodologies, the beginning of the use of Resource Description Framework (RDF) format for linking and interoperability of data. What used to be a long and time-consuming process, with the use of RDF, the speed and flexibility of the information expressed in this format can be verified.

**Keywords:** authority control, heterogeneous sources, methodologies, university libraries

**How to cite:** Menocal, A. T., & Rodríguez, M. del R. F. (2024). Analysis of the methodologies created for libraries in the construction of authority control. In M.J. Peralta González (Ed.), *Generation and Transfer of Knowledge for Digital Transformation, II International Symposium, SITIC 2023, Santa Clara, 15-17 November 2023, proceedings. Advanced Notes in Information Science, volume 6* (pp. 1-14). Pro-Metrics: Tallinn, Estonia. DOI: 10.47909/978-9916-9974-5-1.89.

**Copyright:** © 2024, The author(s). This is an open-access work distributed under the terms of the CC BY-NC 4.0 license, which permits copying and redistributing the material in any medium or format, adapting, transforming, and building upon the material as long as the license terms are followed.

## **1. INTRODUCTION**

---

In the context of evolving library and information center environments and the growing diversity of information sources on the network, the concept of authority control is emerging as a new trend. The advent of new technologies has led to a proliferation of automated bibliographies, which has contributed to the designation of the current era as the “Information and Knowledge Era.” The automation of objects on the Web is a challenging endeavor, primarily due to the sheer volume and complexity of the information they contain, which renders manual management impractical.

In the context of the Cuban web environment, there exists an exchange network utilized by academic institutions for the organization, representation, and dissemination of information within the academic domain. It serves as a gateway to authoritative databases, facilitating the retrieval of pertinent and current information. The University Cooperation Network 1, “Strengthening the role of information and communication technologies (ICTs) in Cuban universities for the development of society,” is a collaborative initiative comprising the University of Pinar del Río, the University of Informatics Sciences (UCI), the University “Oscar Lucero Moya” of Holguín, the University “Ignacio Agramonte y Loynaz” of Camagüey (UC), the University of Oriente (UO), and the Central University “Marta Abreu” of Las Villas (UCLV).

The network has an institutional repository (DSpace), a library management system (ABCD), and an educational

learning platform (Moodle). The primary objective is to unify all the authorities within the network, thereby facilitating efficient search and retrieval of information. In view of the necessity and significance of authority control in the network, the present study has been undertaken. The proposed methodology, which emerges from the project as a result, represents a significant step forward in terms of the organization and retrieval of information contained in the ICT Network. Furthermore, it will facilitate the identification of the most viable and effective elements for the advancement of both the academic and working environments.

The research conducted over the years has contributed to the international context by developing methodologies and methodological guidelines for the control of authorities, as well as for the work with authorities (Herrero & Díaz, 2009; Herrero Pascual, 1999; Hyland & Wood, 2011; Magliano, 2002; Martínez Saldaña, 2006; Vaca Palacios, 2018; Villazón et al., 2011; Yi-hang Pong, 2005).

In Cuba, research is conducted with greater emphasis on collaboration with authorities, with findings disseminated in theses and scientific articles. On an annual basis, each member university of the ICT Network conducts a comprehensive assessment, presenting a detailed report on the research activities undertaken within their respective faculties and study and research centers. The procedure fulfills an informational requirement for this specific purpose; however, it does not permit the retrieval of individual data regarding the authors of the scientific output of each study center precisely because it is not organized or structured through the control of authorities.

A further issue that arises in the context of scientific production is the variety of signatures that authors have

historically employed in the preparation of their research works. As a result, the evaluation of both external and internal research is an ongoing and challenging process for those responsible for this task, with the aim of contributing to decision-making and the development and promotion of the visibility and impact of scientific production in universities. The member universities of the ICT Network employ a considerable number of professors and researchers who disseminate their work at the national and international levels. These institutions possess a substantial corpus of published and unpublished research papers, which are currently not accessible through a unified system that catalogs and organizes the scientific output of their creators.

As time progresses, data assumes greater significance across a range of domains, given its potential to encompass information that may be inaccessible or challenging to obtain otherwise. The removal of duplicate records within a single database represents a crucial phase in the data cleansing process, as these can impact the outcomes of a search or the processing of information. As databases grow in size, the complexity of the matching process represents a significant challenge to the integration of heterogeneous data.

A number of authors have put forth their respective proposals for the methodology of authority control. Each of these approaches, in accordance with the specific attributes of their respective institutions, elucidates the nuances of their work, thereby facilitating the identification of pertinent references and variant forms of the heading. It is evident that none of these methodologies can be applied to the context of Cuban universities, as they remain merely a proposal and have not been implemented. Furthermore, each methodology aligns with the specific characteristics of the institutions where they were developed.

These methodologies are not aligned with the technological advancements that university libraries are currently experiencing. Additionally, they do not fully address the heterogeneity of data present in modern information systems. These norms are outdated and do not align with the technological advancements of the modern era. Additionally, they are not designed to accommodate the diverse and evolving nature of data. As a result, these methodologies cannot be applied in the Cuban national context.

## **2. METHODOLOGY**

The research is predominantly quantitative in nature. However, both qualitative and quantitative data are subjected to investigation, analysis, grouping, and description in order to provide a solution to the previously proposed objectives. The study is descriptive in nature, offering a comprehensive account of the terminology, control authorities, access points, and heterogeneous data sources, along with a detailed examination of their characteristics, development, and evolution on the Web. Furthermore, the research examines and describes the experiences of working with authority control at the national and international levels, as well as in academic settings.

A documentary analysis was conducted to identify materials pertinent to the research topic. The information was obtained through the documentation gathered from primary and secondary sources, including electronic journals, websites, and web articles, in order to gain insight into the existing literature on the subject and to understand the significance of authority control. Upon completion of the bibliographic review, the topic to be addressed was selected and subjected to analysis.

### **3. RESULTS AND DISCUSSION**

---

A number of methodological proposals have been put forth with regard to the implementation of a control of authorities. It is essential to consider the fundamental aspects in order to develop a proposal that is accurate and effective. It is crucial to consider the primary regulations that govern the preparation process. Many institutions, based on the documents to be standardized and their specific attributes, determine the regulations to be utilized. These regulations outline the manner in which the contents should be employed for the creation of the control of authorities.

It is crucial for cataloguers, in collaboration with their respective institutions, to ascertain that no existing authority record corresponds to the subject matter of the proposed new record. It is essential to establish the specific terms that will be used for the controlled vocabulary, as well as the structure of the heading and the relationships that these may have. It is incumbent upon the specialist engaged in this endeavor to document the sources consulted in the preparation of this record.

It is of the utmost importance that there is a rationale to justify the creation of the record and its references. It is not possible to create a name authority record in the absence of a bibliographic entity that serves as an access point. In the event of a potential conflict between similar or identical geographical names, additional sources of information will be consulted to resolve the issue. The various methodologies identified in the scientific literature are presented below. Each of the aforementioned sources is then subjected to a detailed analysis, with particular attention paid to the most significant aspects identified.

As illustrated in Figure 1, each author establishes a methodology that is informed by their place of origin and the linguistic characteristics of their language. In her initial proposal for the establishment of a control of authorities, author Cristina Herrero Pascual outlines four fundamental steps. The initial step delineates the nature of these access points, which pertain to authors, titles, and subjects. The second step addresses the issues that arise with regard to the names of institutions, including the use of upper and lower case letters, the language of the name, geographical names, and so forth. The registration of the authorities and their references in the specified format and software constitutes the third step of this methodology. In general terms, this step addresses the importance of using compatible formats when converting the data in the authority files to an international data interchange format. Finally, the maintenance of the list of authorities is responsible for updating the file and ensuring its consistency, as this will affect both the access to records by users and the online cataloging work that is done on a daily basis.

In her methodology, author Magliano (2002) proposes that ratings in the form of a word or phrase should be based on reference works. Furthermore, she demonstrates the validity of the choice of areas by subject despite the difficulties associated with the numerous cases of identical names. The qualification of authors with identical names is conducted through the use of dates, both for national and international authors. While the Chinese researcher Yi-hang Pong's (2005) methodology is oriented towards satisfying the needs of the bilingual community, improving and streamlining authority control operations, establishing standardization for name titles, and establishing principles for the selection of authority records.

In the context of the Hong Kong Chinese Authority Name (HKCAN) authority check, the presence of a word or combination of words within a name entry may be regarded as a distinctive element. However, this is not to say that such elements are inherently problematic. Indeed, they may be qualified in such a way as to minimize the likelihood of confusion. The distinctive element is typically defined as one of the following: “English compound words,” “non-dictionary English words,” or “unusual combinations of two or more letters as a key part.” In certain instances, commonplace terms employed in a manner that is deemed “distinctive” may also be regarded as distinctive elements.

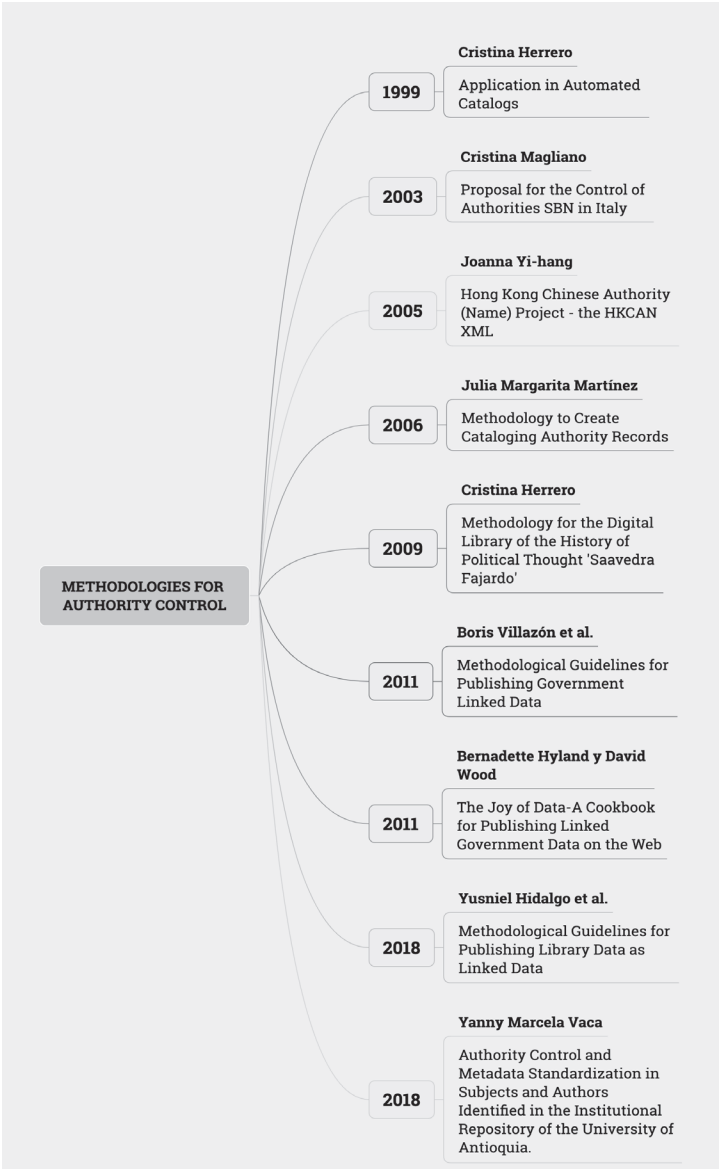
In her master’s thesis, author Martínez Saldaña (2006) puts forth an alternative proposal for the formulation of policies and rules governing the functioning of the work. Furthermore, the adoption of international standards will ensure the quality of the research process for new authorities and facilitate the standardization of name and subject registration. The final proposal, presented years later than the initial methodology proposed by Cristina Herrero Pascual, emphasized the importance of ensuring that homonymy is avoided. Historically, authors have adopted a variety of signatures in the preparation of their research works. Consequently, the designation of an author’s name is contingent upon the existence of a singular form, irrespective of the number of publishing options that the author may have pursued.

A group of researchers led by Villazón (2011) put forth a series of methodological steps for the publication of data from government administrations as linked data. The authors put forth the proposition that the process of publishing linked data should be subject to a defined life cycle,



which they themselves have divided into five fundamental stages. In the initial phase, designated “Specification,” an examination of government data sources is conducted, a Uniform Resource Identifier (URI) is devised, and the licenses associated with the published linked data are established. Phase 2, designated as “Modeling,” pertains to the construction and reuse of ontologies. These ontologies are expressed in Web Ontology Language (OWL) or Resource Description Framework (RDF) and comprise the cleaning of the data and the linking of the data with other previously published data collections, thereby enhancing the discovery and reuse of the data. Generation ensures the transformation of existing data in the data sources to RDF networks and the cleaning of said data. In the subsequent phase, the RDF graphs are published on the Web, accompanied by the metadata pertaining to these graphs, specifically utilizing the Vocabulary of Interlinked Datasets (VoID) vocabulary. The final phase of this methodological process is the development of tools that utilize the published linked data.

Hyland and Wood (2011) published an article entitled “A Cookbook for Publishing Linked Government Data on the Web,” in which they proposed a set of five steps for the publication of linked data. The initial step, entitled “Modeling the Data,” facilitates the reusing and merging of data sets. The proposal introduces a fundamental schema that streamlines the conventional methodology for generating linked data (Hidalgo et al., 2018). The second step is known as “Naming Things on the Web with URIs.” This step considers the URI naming strategy, which provides principles for choosing the appropriate domain for URI sets, path structure, change management, and machine- and user-readable formats. The third step is to



**Figure 1.** Methodologies for the control of authorities. Prepared by the authors.

reuse vocabularies whenever feasible. This entails the use of vocabularies to describe people, places, and other entities. Furthermore, it encompasses the utilization of Dublin Core terms, which pertain to the description of metadata pertaining to published works. The subsequent phase is the conversion of data to RDF, which is simply the source data in a representation or serialization of linked data. Validation is an essential process that helps avoid unnecessary errors when data are loaded into an RDF database. The publication of human- and machine-readable descriptions represents a form of public communication by an organization. Consequently, data policies must be verified in a Web environment.

In 2018, a collective of authors from various countries published the “Methodological Guidelines for Publishing Library Data as Linked Data.” The publication presents five methodological guidelines that provide guidance on the publication of library data as linked data, with a particular focus on ensuring the interoperability of such data. The initial guideline proposed by the authors is data extraction. They assert that their primary objective is to extract and store library data from heterogeneous data sources. The second proposed activity is data preprocessing, which ensures the cleaning and standardization of select metadata fields, including date, volume, and journal numbers, thereby markedly enhancing their quality. The third guideline is data modeling, which is the definition of an ontological model for the sharing and annotation of library data. This model determines the ontologies that will be used to model library data. In order to transform previously extracted, stored, and modeled library data into RDF triples, the fourth step, referred to as data publishing, provides the defined ontology model and the intermediate

database. The output is a set of one or more RDF graphs comprising the library data. This phase is subdivided into three distinct tasks: transformation, linking, and publishing. The final activity, data exploitation, enables the creation of practical applications that utilize the previously published linked data.

In her doctoral dissertation, Vaca Palacios (2018) asserts that a review of the policy manual enabled her to obtain the findings of her research. This constituted a fundamental step in the analysis of the information and in the clarification of the contents of the documents in the repository of the University of Antioquia. The methodology allows for the advancement of attributes in the controlled languages of the subjects, thereby facilitating access and recovery of information contained in the repository's databases. Concurrently, the descriptors are unified to ensure greater organization and validation with the various web-based tools currently available. The objective of this authority control is to verify the correct spelling of the names of the researchers and thus unify the authority records belonging to a given author.

The aforementioned methodologies share a number of similarities, including their foundation on fundamental international formats and guidelines for authority work. Herrero Pascual (1999) developed a methodology, "Application in Automated Catalogs," which aligned with GARE standards. This approach facilitated the standardization of catalog headings related to access points. In 2004, Magliano primarily draws upon extant models within international projects such as VIAF, LEAF, NACO, FRANAR, and IFLA. The standard utilized is the UNIMARC/Authorities: Universal format for authorities.

The proposed Chinese names authority is based on the MARC 21 format, which is designed for use with multi-lingual materials. The Anglo-American Cataloguing Rules, 2nd edition (AACR2) oversee it, as well as the Library of Congress Subject Headings (LCSH) thesaurus rules. The methodology proposed by the author, Martínez Saldaña, makes fundamental contributions to the processing of personal, corporate, meeting, and geographic name records, as detailed in the PCC/SACO program and NACO training manual. The proposed method for creating records is through the OCLC ConnexionClient interface. MARC tags are employed for the purposes of referencing, indicating related items, providing notes, and sourcing, among other functions. Herrero Pascual and Díaz (2009) demonstrate that it adheres to the authority data model established by the primary international standards and guidelines, including FRANAR, GARR, MARC 21, and AACR2.

#### **4. CONCLUSIONS**

In the field of library data mining, two fundamental technological issues emerge: the design of middleware and the mapping of schemas. One of the limitations of these methodologies, particularly those developed at the turn of the century, is their inability to handle heterogeneous data and their adherence to highly outdated regulatory frameworks.

#### **REFERENCES**

- Herrero, C., & Díaz, P. M. (2009). Gestión de un fichero de autoridades en el marco de la biblioteca digital de historia del pensamiento político "Saavedra Fajardo". *Ibersid*, 3, 209–215. <https://doi.org/10.54886/ibersid.v3i.3741>
- Herrero Pascual, C. (1999). El control de autoridades. *Anales de Documentación*, 2, 121–136.

- Hidalgo-Delgado, Y., Mariño-Molerio, A. J., Amoroso-Fernández, Y., & Leiva-Mederos, A. A. (2018). Some reflections on linked open data in Cuba. *Revista Cubana de Información en Ciencias de la Salud (ACIMED)*, 29(4), 1–9.
- Hyland, B., & Wood, D. (2011). *The joy of Data-A cookbook for publishing linked government data on the web*. Springer. [https://doi.org/10.1007/978-1-4614-1767-5\\_1](https://doi.org/10.1007/978-1-4614-1767-5_1)
- Magliano, C. (2002). Guidelines and methodology for the creation of the SNB authority file.
- Martínez Saldaña, J. M. (2006). *Control de autoridades. Metodología para crear Registros de Autoridad Catalográfica* [Tesis de maestría, Tecnológico de Monterrey-Universidad Virtual, México].
- Vaca Palacios, Y. M. (2018). *Control de autoridades y estandarización de metadatos en las materias y en los autores que se hayan identificado del repositorio institucional de la Universidad de Antioquia* [Tesis de grado, Universidad de Antioquia, Colombia].
- Villazón, B., Vilches, L. M., Corcho, O., & Gómez, A. (2011). *Methodological guidelines for publishing government linked data*. Facultad de Informática, Universidad Politécnica de Madrid. [https://doi.org/10.1007/978-1-4614-1767-5\\_2](https://doi.org/10.1007/978-1-4614-1767-5_2)
- Yi-hang Pong, J. (2005). *Hong Kong Chinese Authority (Name) project—The HKCAN XML*. <http://dspace.cityu.edu.hk/bitstream/2031/609/1/presentation.html>