548

Knowl. Org. 46(2019)No.7
R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies
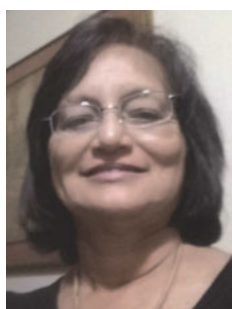
The Representation of Archival Information
in Controlled Vocabularies:
The Context of the Archival Institutions in Rio de Janeiro

Rosale de Mattos Souza

Federal University of Rio de Janeiro State, Brazil,

<rosale.m.souza@unirio.br >

Rosale de Mattos Souza is Associate Professor in archivology at the Federal University of Rio de Janeiro State-UNIRIO; Department Boss and Deputy Director of the School of Archivology of UNIRIO; Specialist in Arrangement, Description and Research in Public Archives-UNIRIO, Specialist in Brazilian History-UFF. She holds a masters degree and a doctorate in information science from the UFRJ ECO IBICT agreement.

Abstract: We aim to trace the scenario of the use of controlled vocabularies as tools of research and work in the scope of representation and retrieval of information in institutions that have archival collections, in order to highlight the need for greater emphasis on the subject of representation of archival information in the academic field, increasing the visibility for the study and analysis of the collections in question and their contents, relevant to the information society. We investigate the current scenario of the use of controlled vocabularies in the archival collections of Rio de Janeiro, the theoretical-methodological changes arising from the impacts of information technologies on analysis, representation, such as classification and indexing, content retrieval, information needs in the contemporary world. The representation of information is associated with classification and retrieval of information to the organization of knowledge in information science. There is a gap in the archival area regarding the expression representation of information from the description of the nineteenth century. As for the theoretical-methodological aspect, there was a theoretical survey of the representation of information in publications in the interdisciplinary areas; as to the operational methodology, questionnaires were applied to information agencies on the use of controlled vocabularies, in relation to the treatment of information in archival collections. We conclude by demonstrating the importance of adopting the concept of information representation in archives, using controlled vocabularies associated with new information technologies and informational ecology, consolidating the area as a scientific and interdisciplinary field for information science.

1.0 Introduction

From the research project "Documentary Languages, Semantics and Representation of Archival Information," we have been working since 2014 in the research area of a federal public university in Rio de Janeiro. In this way, we aim to survey the situation of information agencies and archival institutions that contain in their work programs the existence of controlled vocabularies as a tool that can assist in the elaboration of classification plans, temporality tables of documents, and as an instrument of control of terms used in archival information systems. It is noteworthy that in archivology there are gaps in the use of classification theory linked to the organization of knowledge, concept theory, information representation, investigations related to controlled vocabulary and indexing polic.

Pombo (1998) deals with the relation of classification as an old object of analysis by several theoreticians and area of knowledge, such as in Aristotle in philosophy, Piaget in education, Augusto Comte and many other authors and thinkers. In the tree of Porphyry, which established a hierarchy of terms and concepts, it was inspired by Aristotelian thought, in which a thought is made by opposition. Pombo also gives as an example that in literature is also in Jorge Luis Borges' inspiration for Michel Foucault, in the work "Words and Things" to classify the world, breaking with the Eastern classification of the Western.

Already in a study about the organization of knowledge and its meaning, Mazzocchi (2018, 54) highlights the theory and some tools used in the representation of information, such as:

Knowl. Org. 46(2019)No.7
R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies

549

Knowledge organization system (KOS) is a generic term used for referring to a wide range of items (e.g. subject headings, thesauri, classification schemes and ontologies), which have been conceived with respect to different purposes, in distinct historical moments. They are characterized by different specific structures and functions, varied ways of relating to technology, and used in a plurality of contexts by diverse communities. However, what they all have in common is that they have been designed to support the organization of knowledge and information in order to make their management and retrieval easier.

According to Mazzocchi, there is a pluralism in the perspectives of understanding of KOSs, discussed in the theoretical plane from the issues related to mental and cultural patterns, notions of what is knowledge in a KOS, to the issues of semantic tools and typologies such as classification and categorization, term lists, metadata models, term relationship models, terminology and comparison issues. By extension, it raises Dahlberg's, Svenonius's, Hjørland's, and other related approaches, if there is an unbiased ontology, and analyzes KOSs past and future prospects (55):

> In the view of many KO scholars, such as Broughton et al. (2005) and Hjørland (2008), there are two main items that characterize knowledge organization (KO): 1) knowledge organization processes (KOPs), such as abstracting, indexing, cataloging, subject analysis, classifying, and, 2) knowledge organization systems (KOSs), i.e. tools designed for the general purposes described above, which will be analyzed here.
> An important issue to be underscored is that, while their basic scope has remained unvaried over time, the environment in which KOSs have to operate has

instead drastically changed, and it will continue to change: from the world of physical libraries, for whose purpose grand classification schemes were created, to databases, the digital environment, and the internet.

Gnoli (2018) observed KOSs from the aspects of mentalism, which leads to individual and sociological knowledge, in a "pluralistic epistemology, where individual minds and social communities are only two successive levels of knowledge production and use ... this can, in turn, be analyzed in artifacts and mindfacts." Therefore, just as both Mazzocchi (2017) and Gnoli (2018) point out, there is a mentalist and cultural perspective promoted in society.

For Hjørland (1995; 2002; 2019) the social dimension of domain analysis is verified, involving discursive community, domain and language—that is, a domain of specialized or specific knowledge by a given community with analysis and interpretation of the domain data with the social conception of the use of information. Domain analysis can have the following concept (Hjørland and Albrechtsen 1995, 23):

> It is a theoretical approach to Information Science (IC), which states that the best way to understand information in information science is to study the areas of knowledge as "discourse communities," which are part of the division of labor society. Knowledge organization, structure, patterns of cooperation, language and forms of communication, information systems and relevance criteria are reflections of the work objects of these communities and their role in society. Psychology, knowledge, the need for information and subjective criteria of relevance should be viewed from this perspective.

Categories of KOSs	General features of the categories	Specific types of KOSs
LISTS	Linear and less structured systems; emphasis on the lists of terms (frequently provided with definitions)	Authority files Glossaries Dictionaries Gazetteers
CLASSIFICATIONS AND CATEGORIES	Hierarchically structured systems; emphasis on the creation of subject sets	Subject headings Classification schemes Taxonomies Categorization schemes (the last three terms are frequently used interchangeably)
RELATIONSHIP LISTS	Complex and highly structured systems; emphasis on the connections between terms and concepts	Thesauri Semantic networks Ontologies

Table 1. Classification of KOSs (Hodge 2000).

550

Knowl. Org. 46(2019)No.7
R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies

It is important to consolidate the theoretical questions in order to contextualize cultural aspects that influence the languages forged by discursive communities with the domain of knowledge, and the operational tools in society such as lists, thesaurus, taxonomies, ontologies and controlled vocabularies.

A controlled vocabulary is an artificial language, which results from choices of terms in the language code. Smit and Kobashi (2003, 14) emphasize the importance of a "strict classification," that is, a classification plan that can be considered appropriate in relation to the institution. It is important to analyze the end-activities of institutions. Thus, in order to make the representation of the documentary universe, it is necessary to incorporate a more specialized language. The ordering, hierarchization, structuring or categorization of terms generate the architecture of functions and activities of the institution, attributing meaning to the whole. A controlled vocabulary should hierarchize the functions and activities, contemplate the control of vocabulary, especially with regard to synonymy, homonymy and almost synonymy, relating the terms present in different categories, in addition to conceptualizing terms and conditions of use.

According to Sousa (2002, 15), the archival classification function of documents, considered a matrix for archival work, is the explication of functions or structures within a context of documentary production. It is the representation of the mapping of this productive context and that subsidizes the other stages of document management. The documentary item is classified thinking about the whole and the existing relations, in the role played by the information within the organism. Unlike the book for the library, an information unit that is independent of other units, the document, or archival information, thought at the level of information unit, can only be signified if related to other informational units that represent an activity, subfunction or specific function. As they are being accumulated, the information establishes relationships with each other. They are united by a bond created at the moment in which they are produced and received, called organicity, determined by the reason for its elaboration and which is necessary to the very existence and the capacity to fulfill its objective. They are an indivisible set of intellectual relations.

According to Sousa there is no instrumentalization that is not supported by theoretical and/or methodological bases. For the author, one must be aware that the classification is not an intuitive process but carried out as a technical activity that demands methodological and scientific procedures, proposing that one must perceive an interdisciplinary triangulation between the archival principles, classification theory and the methodology of data collection.

In archival classification, it is necessary to take into account the mission, functions, activities, typologies, species and documentary types present in all archival documentary ages from the current archive to the permanent archive, facilitating the elaboration of classification plans and of temporality tables of documents and, consequently, document management. The objective aspect, given by both diplomatic and documentary species as well as documentary organic features, facilitates the conceptual determination of the terms of their documentary series, while the thematic representation that leads to subjectivity can be observed with greater relevance in permanent archives. With respect to the hierarchical and classificatory principles pertinent to a file, the following principles must be taken into account: function categories (middle and end), functional áreas, function, subfunction, activity, subactivity, task or act.

The concept of information representation corresponds in works of reference of the area of documentation and information with the establishment of concepts and symbolic notations in analog and digital information systems. There are different approaches to information representation and retrieval related to information science and librarianship, such as those focused on ontology standards, thesaurus, controlled vocabularies and taxonomies, exemplified in studies and practices in the legal and medical fields; as well as the lines of research focused on semantic web, folksonomias or cognitive ecologies, faceted classification, domain analysis, information and semiotics, related to language, artificial language and cybernetics, etc.

The approach of the representation of archival information still presents itself as a gap, since there are currents of thought of this area originating from nineteenth-century historiography, which privilege the term description to the detriment of the term or activity of representation. Information representation is more associated with the postmodern notion of archival practice; dynamic, fluid, independent of archival documentary ages that guide the management of documents and information, since the archival area consists of a field in construction, which is consolidating over time; an environment of constant changes, due to its interdisciplinarity with other areas, such as information science, which justifies its changing nature, the way it relates to other areas, such as administration, library and information technology, and does not remain merely as a simple autonomous auxiliary field of history.

The methodology proposed for the development of this study consists of a theoretical and empirical survey of the field of KO, including the archival area, the relevant scientific material on the subject of information representation, presenting lines of research on information representation, contained in publications in the interdisciplinary areas mentioned above, as well as the application of questionnaires in archival, museological, library and documen-

Knowl. Org. 46(2019)No.7
R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies

551

tation centers about the use of controlled vocabularies used by these institutions in their daily practices regarding the treatment of information in archival collections.

As a result of this research, we hope to ratify the importance of absorbing and adopting the concept of representation of archival information through the use of controlled vocabularies for the area in question, as well as proving that archival theory does not remain static through the years; that its flexible and adaptable nature to the new technologies and informational needs of users, without losing the parameters of its archival principles, such as the principle of respect to the fonds, origin and original order, consists in the condition of its own maintenance and existence as scientific and interdisciplinary field to information science. Therefore, the archivology dialogue is important with other areas such as librarianship, information science and communication to develop the theory of concept, principles, theory and archiving in the knowledge society.

2.0 Context of major research around the organization of knowledge and representation of information

2.1 In the studies on currents of thought and lines of research on information representation

Araújo (2018, 54-59) analyzes the research possibilities of representation, information organization and information retrieval. Thus, with the advent of digital technologies, information and communication, the new perspectives of information organization in a collaborative and social way, classification, organization and description of information have given greater visibility to information representation and retrieval research.

With regard to digital technologies, new patterns of description and codification, controlled vocabularies have emerged, promoting ontologies and new taxonomies. In the trajectory of the representation of information associated with digital technologies we must also remember the semantic web and internet 2.0, associated with open data and access to information. Today, we have seen techniques used in the processing of information related to gender studies and the social role of archives and archivists in the construction of citizenship and democracy. Still on the terrain of contemporary information representation studies in digital environments and relational databases, we cannot fail to remember faceted classification, inspired by Ranganathan's theory. Thus, in the structures of information retrieval systems (SRI) are factors involving hierarchies, facets and levels of reality.

Another aspect is the representation of information or classification from the cultural context of a society or domain analysis, which relates to the cultural and scientific domain of certain areas of knowledge and their terminological specificities, as well as the cultural aspects imbricated with the discursive communities. Araújo states (2018, 57):

> The fundamental contribution of the domain analysis perspective is the understanding that it is not an isolated subject that has needs, ways to seek and use information. "Need for information" is something that arises collectively. It is a group of people that develops certain patterns of what kind of situation or activity generates need for information, or what kind it should be needed in each context, and so on. It is, in short, a collective (and, in the case of this approach, a domain, that is, a certain area of scientific knowledge with its own object, own methods) that has a way of dealing with information. It is up to the researcher who develops domain analysis to detect these collective ways, which are public because they are socially shared and experienced by members—that is, because the meanings they attach to processes are public. And the professional of the field of information, it is necessary to develop services and systems of information adapted to these collective standards.

In the aspect of the interactions between citizens and information system professionals, a new current of thought is highlighted—folksonomy, also called social indexing or cognitive ecology, that represents a new social form in the web of processes of consolidation of signs or terms with new signifiers and meanings, expanding information resources in an open environment and sharing collective construction of information retrieval systems (SRI). According to Araújo (2018, 55), we have seen this expression and modes of operation as follows:

> The term folksonomia was created in 2005 by Vander Val to designate the labeling of WEB resources in a social environment made by the users themselves. It arose with web 2.0 and its proposal of an architecture of participation and, in the field of information science, articulated to a decentralized dynamic of the actions of information representation. An alternative designation is the expression social indexing, which refers to the dynamics through which the users themselves describe the same resource, resulting in an intersubjective description, performed through.

In a theoretical approach, the perspectives of the language sciences have been rescued, associating semiotics with information and knowledge, with applications in the instru-

552

Knowl. Org. 46(2019)No.7
R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies

ments of research and studies on discourse and power, discourses and narratives, ideologies related to the choices of funds and documentary series to be treated to the detriment of others. In a change of paradigm in archivology in the twentieth and twenty-first centuries, the influence of these areas of language analysis on archival functions, such as classification, description, diffusion and access to information, was considered.

According to Araújo (2018, 58), there was a revaluation of the computer systems through the systems of information retrieval from hypertext, automatic indexing, conceptual maps and information architecture. Another dimension or line of research on information representation and information retrieval is in cybernetics, relating man-machine, usability and artificial intelligence.

In this way, we have seen in the contemporary period new forms of looking at the representation of information that not only the simple automatic retrieval of information in information retrieval systems (SRIs) but also those that transform information into knowledge and that have the power to transform and develop individuals and society as a whole.

3.0 Vocabulary controls and vocabulary control in archives and archivology

3.1 Vocabulary controls and vocabulary control

Smit (2018) deals with controlled vocabularies and vocabulary control, taking into account the management of documents and the perception of the advantages of using this technique from the current files to the permanent files, giving the idea of continuity of the information system. According to Smit, permanent file documents are quieter and more visible, and documents in current archives are more dynamic and document-oriented. Smit (citing Menne-Haritz 2004) emphasizes that archives and archivists still give more prominence to documents as physical objects and before this points out (47-48):

> Archiving still places great emphasis on the physical paradigm, which is particularly clear in the bibliography dealing with permanent archives, almost entirely composed for the time of "physical" documents. The importance attributed to the document is fully justified when it is understood as evidence of institutional transactions and its custody means the guarantee of maintaining institutional stability, by recording both the activities and the forms of organization of decision-making processes

Smit presents questions about aspects such as the physical paradigm, the document or "information as thing" (Buck-

land 1991) from the 1970s, and the cognitive paradigm from the 1980s to the 1990s, which made subjects active in the process, transforming information into knowledge. Finally the third paradigm is the social one. Smit states (2010, 3):

> From the 1990s, there is a movement of synthesis, called the social paradigm, which continues to presuppose the existence of stocks of registered information, but inserts the user of information into society, that is, the user is no longer isolated, since is a participant and protagonist of life in society. To return to something I affirmed a few years ago, "the physical paradigm prioritized the object, the cognitive paradigm emphasized the subject, and the social paradigm sees the subject, contextualized, in its relation to the information object.

Smit (2018, 52) is concerned with questions of standardization of information derived from the new demands of information technologies, promoting since the 1990s the development and creation of International Standards for Description, such as the *International Standard for Archival Description (General)* (or *ISAD-G*), and the *International Standard Archival Authority Record for Corporate Bodies, Persons and Familie*s (*ISAAR-CPF*), with the standardization and entry of terms and names in the archival databases through the access points. For Smit, we have seen the advantages of new technologies with the study of information networks, in which there may be relational sources of documentary funds within the entity that generated them or that keeps them, and between documentary funds existing in archival institutions.

Smit also analyzes the theoretical-methodological principles of terminology, aiming that for each term in the controlled vocabulary there must be a process of terminological univocity in order to avoid multiplicity of meanings, avoiding ambiguity and seeking to establish a meaning for each signifier. Another aspect is the clarity and the accompaniment of the grammatical rules, identifying synonyms, almost synonyms and antonyms. Two issues are imposed, ISO no. 2788, dated 1986, which deals with standardized nouns in the masculine and singular form, and, in the case of archival collections, the designation of functions and activities, through verbs. The *ISAD (G)* (2000), *ISAAR (CPF)* (2004) and the *International Standard for Describing* Functions (*ISDF*) (2007) standards initially propose the standardization of names and entities.

The norms of description constitute a milestone for the area in the search for descriptive standardization from the late 1990s and early 2000s on the possibility of greater communication among information agencies, not only public and private archives but also between documenta-

Knowl. Org. 46(2019)No.7
R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies

553

tion centers, libraries and museums that have collections and archival documentary fonds.

3.2 Controlled vocabularies as auxiliary research tools associated with description in archival and information representation

The norms of description in archives, such as the *ISAD (G)* and the Brazilain national standard *Norma Brasileira de Descrição Arquivística e os Documentos Audiovisuais* (*NOBRADE*), among other norms of description, inaugurated a new moment in the form of understanding description and recovery of archives, taking into account the technologies and new needs of information systems. We cannot fail to reflect on the main characteristics of these standards and the multilevel description, such as: description of the general for the particular, information relevant to the level of description, relationship between descriptions, not repetition of information.

There are seven areas of information in *ISAD (G)*:

1) identification área;
2) area of contextualization;
3) area of content and structure;
4) area of conditions of access and use;
5) area of related sources;
6) notes area; and,
7) description control area.

There are eight information areas in *NOBRADE*, which included the area of access points and subject indexing, which we will not detail in this article. In the *NOBRADE* glossary, there is only one conceptualization of what is coordinated indexing and pre-coordinated indexing, however, there is no greater explanation about indexing and indexing policies. In addition, the term "subject heading" is not part of the archival terminology, requiring an interdisciplinary knowledge of the documentary process, as we can see (Arquivo Nacional 2005): "Post-coordinate indexing: Indexing for terms that must be combined at the time of search for filtering of the desired information, usually called descriptors 'and' Pre-coordinate indexing: Indexing by previously combined terms, usually identified with subject headings."

There was a concern in this particular project with the eighth *NOBRADE* information area, regarding access points and indexing of subjects. It should be mentioned that archivists need more knowledge and training regarding the representation of information, particularly regarding the indexing, treatment and dissemination of information, since there is no clear understanding of KO, a subarea of information science, which is associated with archival theory. According to Campos (2006, 17-31) it analyzes the descriptive and thematic representation of information in archives:

> We can cite as a concern of this nature *ISAAR* itself—*International Standard for Archival Authority Registration* for collective entities, individuals and families. In addition, *ISAD (G)* itself uses ISO standards for documentary description for standardization of sources. However, regarding the access points related to thematic indexing, this same rule only concerns itself with indicating that the vocabulary should be controlled, but does not address methodological aspects related to the procedures of analysis and even the elaboration of controlled vocabularies.

There must be more consistent interdisciplinary studies on the subject, such as descriptive representation and thematic representation of information. In addition, we emphasize that controlled vocabularies are tools and can be used as a methodology to aid in classification plans, temporality tables of documents and in the elaboration of research instruments.

In this sense, we come to the controlled vocabulary: an instrument that names the terms corresponding to the documentary language used in the archives of a certain collection in order to organize and retrieve documents and archival information with consistency and accuracy. To do so, it seeks to standardize all language to be used by the institution, by naming a single term for a particular purpose, thus avoiding multiple entries, ensuring complete retrieval of information. Smit and Kobashi state (2003, 16-17):

> Let's imagine that car orders are named in five different ways: transportation order; vehicle requirement; vehicle requisition; transportation request and vehicle request ... If, for example, problems occur when scheduling a car, to locate its document and check what happened, it will be necessary to remember how the request was named at the time of the request. In order to avoid such discomforts, the most rational solution is to control vocabulary: at the entrance of the system, with the adoption of a single form of designation to name documents that are generated by the same activity; and the output of the system (in the search), informing how each activity is named by the system.

From these considerations it is verified that controlled vocabulary is a method that can be used both in the organization and in the retrieval of documents, serving all routine archival processes, since it consists of a common reference system shared by producers and researchers of archival information from the same custodial entity, making

554

Knowl. Org. 46(2019)No.7
R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies

possible the optimization of classification plans and tables of temporality of document allocation.

According to Smit and Kobashi (2003), there are two different procedures in the construction and use of controlled vocabularies: a micro- and a macro-procedure. The first seeks to introduce control in the expressions or terms and between them, encompassing aspects related to linguistics (such as synonymy, homonymy and quasi-synonymy control and grammatical normalization), while the second focuses on ordering the terms; the instrument becomes a long list of controlled expressions for which the timing and purpose of use is not known for certain, and it does so through the establishment of a hierarchy, ordering, structuring or categorization of terms. The use of controlled vocabularies facilitates the daily life of archives at all documentary ages, when there is concern about the terms appropriate to the document type already in its current file as well as when it is used to reveal appropriate entries at the time of access search in a permanent archives.

4.0 Methodology: application of questionnaire in an informal field

The panorama of the use of controlled vocabularies in Rio de Janeiro, Brazil is consolidated from the application of questionnaires to the information agencies that have archival collections, regarding the existence or not of the presence of controlled vocabulary as a working tool for information retrieval, enabling future policies and indexing term manuals. Therefore, we did a survey of the institutions that have archival collections in the city of Rio de Janeiro, in order to quantitatively and qualitatively construct the current scenario of the treatment of archival information, considering its representation. Questionnaires were supplied in 2016 to twenty-six institutions: eight documentation centers, ten archives, three libraries, and five museums. However, only sixteen institutions returned answered questionnaires or addresses to questions presented: of five museums, four responded; of three libraries, two answered; of eight documentation centers, only four answered; out of ten archives, six responded.

We found that for documentation centers, museums and libraries, at least 50% work with these tools, whereas only one documentation center consulted has this concern more advanced—the Casa de Oswaldo Cruz (COC), the Fundação Oswaldo Cruz (FIOCRUZ), which has an interdisciplinary working group on an indexing policy and has promoted the elaboration of a term-indexing manual. The National Archives had experience with terminology and controlled vocabulary, but it did not continue to function in the diverse sectors of the institution or its diverse forms of search. The Latin American and Caribbean Center on Health Sciences

Information tool BIREME (https://www.paho.org/bireme/index.php?lang=en) is a powerful, available and specialized language tool, associated with a discursive healthcare community, and a reference for FIOCRUZ COC controlled vocabulary indexing and writing, taking into account the descriptors—DECs, in health sciences.

The Latin American and Caribbean Center for Health Sciences, also known by its original name Regional Library of Medicine, is a specialized center of the Pan American Health Organization / World Health Organization (PAHO / WHO) oriented to technical cooperation in scientific health information. BIREME's headquarters are located in Brazil, on the central campus of the Federal University of São Paulo (UNIFESP) since its inception in 1967, in accordance with the agreement between PAHO and the Government of Brazil. We attribute this result to the two factors studied in these areas: treatment focused on information content of the collection, which requires more attention to the issue of information representation and which is the biggest concern for the user. It is evident that the consideration of both is constantly present in discussions of librarianship and museology, which, in turn, affect the documentation centers, while archives are still considered very recent issues and little studied. It can be seen that there is a greater number of occurrences of archival institutions that do not work with the concept and techniques of information representation, a tool used in large scale in other documentation and information institutions, such as libraries, museums and documentation and information centers.

While the results clearly illustrate the need for investment in this issue, it is important to note that quantity is not quality. That is why we have also brought the result of the practices of using this tool. It is fair to argue that even if the archives do not use it, their information retrieval systems work, though not in their entirety. However, this finding does not invalidate the discussion, especially when we realize that the return of system efficiency is not yet absolute in entities that work exclusively with controlled vocabulary (such as museums and documentation centers). Questionnaires were supplied to institutions (information agencies) in the year 2018, with archival collections of about eight questioned institutions, after two years, including Casa de Oswaldo Cruz, Public Archive of the State of Rio, Archive of the City of Rio de Janeiro, among other institutions such as libraries and documentation centers, obtaining the following results:

– The user does not interact with the information system: 100% of the interviewees;
– Only archival collection: 37% yes and 63% no;
– It has a bibliographical collection: 72% yes, 14% no, 14% no response;

Knowl. Org. 46(2019)No.7
R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies

555

– It has a museum collection: 67% yes, 16% no, 17% no response;
– It has a description policy: 37% yes, 50% no, 13% no response;
– Uses some standard description:
   ISAD (G): 10% yes
   ISAAR (CPF): 10% yes
   NOBRADE: 60% yes
   Other: 20% yes;
– Works with information representation: 12% yes, 50% no, 38% no response;
– Uses controlled vocabulary: 12% yes, 63% no, 25% no response;
– There was resistance to implementation of controlled vocabulary in the institution:
37% no, 63% no response;
– There was difficulty during the implementation of the information retrieval system by the users: 37% no, 63% no response;
– Works with indexing: 50% yes, 50% no;
– It has an indexing policy: 25% yes, 50% no, 25% no response;
– The adopted Work Methodology is adequate to the information retrieval systems:
25% yes, 50% no, 25% no response;
– The terms used are easily understood by users: 37% yes, 63% no response;
– After the implementation of the controlled vocabulary there was improvement in the information retrieval system: 25% yes, 75% no response; and,
– The controlled vocabulary is available for consultation: 25% yes,12% no, 63% no response.

Most of the institutions consulted that have archival collections but are not archives, predominantly use controlled vocabularies for bibliographic collections, do not use international norms of description but use the national standard *NOBRADE*, work with indexing, but do not have indexing policies, and find that terms are not well understood by users.

5.0 Results

The importance of controlled vocabularies as tools for the representation of archival information in the context of documentary production, peculiar to the institutions producing archival documents, is confirmed, taking into account the context of production, which is peculiar to archives in relation to other information units, such as libraries, museums and documentation centers. To emphasize the adoption of term representation of archival information in the area in question, since, even to the detriment of some archival currents that are eminently historical, it

is necessary to adopt such theoretical and practical reference, in keeping with the changing profile of the archival area vis-à-vis new information needs that trigger emerging user behaviors. This practice is characterized by a paradigm shift in the cultural area, which migrated from the documentary to the informational paradigm, the latter influenced by information and communication technologies and their undeniable impact on the so-called information and knowledge society.

From the 2016 results, we began the recognition of institutions that have and use controlled vocabularies as a working and research tool: of four museums, all have the tool; of two libraries one has and another does not have the tool; of four documentation centers all have the tool; of six files, none of them have the vocabulary tool tracked. However, FIOCRUZ's National Archives and the Casa de Oswaldo Cruz (COC) have been developing a process to have controlled vocabularies, and the latter institution has been studying a policy of indexing and vocabulary control. Of six archives interviewed, only two work with information representation.

The main results from 2018 questionnaire marked the following points as follows:

– Most of the institutions interviewed have bibliographic collections, followed by museological collections and finally archival collections.
– Most of them have no description control, however they use *NOBRADE*—60%.
– Information institutions or agencies that work with information representation: 50% do not use, 38% do not answer.
– Uses controlled vocabulary: 12% yes, 63% no, 25% no response, indicating little use of controlled vocabularies.
– Works with indexing: 50% yes, 50% no, marking half of institutions that use indexing, half not using this tool.
– It has an indexing policy half of the interviews said no.
– The adopted work methodology is adequate for the information retrieval systems: half said no.
– The terms used are easily understood by users.
– After the implementation of the controlled vocabulary there was improvement in the information retrieval system.
– The controlled vocabulary is available for consultation: most respondents did not respond.

There was a demonstration in two years, during different moments of interviews, of the little use of controlled vocabularies in the interviewed information agencies, pointing out that much has to be advanced in the use of indexing of these knowledge organization tools, in particular in Archiv-

556

Knowl. Org. 46(2019)No.7
R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies

ology. To advance in the area of archivology through inter-disciplinary relations with librarianship, information science and the organization of knowledge, it is necessary to expand the theoretical and practical investigations of the use of documentary languages and thematic representation of information in information systems in institutions that contain archival assets.

6.0 Conclusions

Knowledge organization (KO) literature continues to observe the pluralism of views on representation of knowledge and information in a mentalistic way and by the social dimension (Mazzocchi 2017). Gnoli (2018) points out that there is a mentalist and cultural perspective promoted in society. In Hjørland (2002), we also observe social and psychological aspects of the social actors involved, with the domain of knowledge, the discursive communities and the specialized languages, used in practice in their use in the tools that validate the terms and in information retrieval systems. On the other hand, KOSs seek to analyze processes and systems of knowledge representation, such as subject lists, taxonomies, thesauri, ontologies, controlled vocabularies and other tools adapted to the new contexts and impacts of information and communication Technologies on social networks, in folksonomic processes and validated terminologies in web 2.0 and 3.0.

With this research we intend to contribute to and consolidate the use of artificial languages (controlled vocabularies) in the archives as representational tools, with a view to the construction of better archival classification instruments, descriptions and more efficient information retrieval systems, in addition to excellence in service to the users. Another objective is the stimulus to reflect on the concept of information representation in the archival field, since this concept permeates other areas of documentation-information with perspectives common to informational treatment, aiming at the optimization of the construction of recovery systems of information in organizations, both public and private.

The impact of the representation of archival information, which is notorious despite some conceptual challenges in the archival area itself, can be reflected in the optimization of better and more efficient information retrieval systems and information management, promoting greater service to users, according to their research profiles. The acceptance and application of the concept of representation in archives brings to the fore the discussion about such activity beyond what is known as archival description.

It was observed that the archives and the archivists, when familiarizing themselves with the perspectives of currents of thought of the representation of the information, will be able to widen their range of searches in organization of knowledge, allowing reflexes in archival functions, as in production, classification, description and diffusion. By extension, the researchers will be able to reach aspects of language, semantic web, cognitive, social—through folksonomia—of cybernetics and etc.

Archival description standards, such as *ISAD (G), NOBRADE, ISAAR (CPF)* and *ISDF* at the end of the twentieth century and the beginning of the twenty-first century were fundamental instruments for the consolidation of descriptive standards, for the documentary ages in the archives and for both current archives and permanent archives. In particular, the standards of description promoted the consistency of information produced in the producing bodies, facilitating continuity in information systems, allowing related sources, both in documentary fonds within the same institution and among the fonds of different institutions or information agencies. In addition, we highlight the eighth information area of *NOBRADE*, which allowed the concern with access points and indexing of subjects that can be used in computer environments on the internet.

It has been found that most institutions consulted with archival collections do not use controlled vocabularies, but libraries and documentation centers use these and other tools, such as taxonomies, ontologies, thesauri and controlled vocabularies. The experiences found in Casa de Oswaldo Cruz, as a documentation center, at FIOCRUZ, which allowed the elaboration of a manual and an indexing policy, and the National Archives pointed out the existence of projects in this sense, although the latter did not proceed with the job. These experiences may present domains of knowledge, languages from specialized knowledge areas such as Casa de Oswaldo Cruz in health, as well as the *BIREME*, and the National Archives of Brazil with the federal government agencies and fonds that were collected, featuring discursive communities with specialized knowledge.

It is necessary to stimulate archival thinking and doing with greater intellectual production on reflections that involve works in congresses, articles and books on information representation and contributions to the archival field. The archivist must always be in constant adaptation to new market demands in relation to professional, theoretical and practical conduct due to the changing times we are experiencing—fluid and changeable information, modernization of technological devices, dynamics between information professionals and users, whether actors in libraries, archives, documentation centers or museums, of which we must undoubtedly argue to justify the existence of archives in the domain of information science.

References

Araújo, Carlos Alberto Ávila. 2018. "Representation of Information." In *What is Information Science*. KMA: 54-9

Knowl. Org. 46(2019)No.7

R. de Mattos Souza. The Representation of Archival Information in Controlled Vocabularies

557

Buckland, Michael. 1991. "Information as Thing." *Journal of the American Society for Information Science* 42: 351-60.

Campos, Maria Luíza de Almeida. 2006. "Indexing and Description in Archives: The Question of Representation and Retrieval of Information." *File & Administration*, 5, no. 1: 17-31.

Conselho Nacional de Arquivos-CONARQ. 2006. *NOBRADE-Norma Brasileira de Descrição Arquivística e os Documentos Audiovisuais*. Rio de Janeiro: CONARQ.

Gnoli, Claudio. 2018. "Mentefacts As a Missing Level in Theory of Information Science." *Journal of Documentation* 74: 1226-42. doi:10.1108/JD-04-2018-0054

Hjørland, Birger. 2002. "Domain Analysis in Information Science: Eleven Approaches: Traditional as Well as Innovative." *Journal of Documentation*. 58: 422-62.

Hjørland, Birger and Hanne Albrechtsen. 1995. "Toward a New Horizon in Information Science: Domain-Analysis." *Journal of the American Society for Information Science*. 46: 400-25.

Hjørland, Birger. 2008. "Deliberate Bias in Knowledge Organization?" In *Culture and Identity in Knowledge Organization: Proceedings of the Tenth International ISKO Conference 5-8 August 2008 Montréal, Canada*, ed. Clément Arsenault and Joseph T. Tennis. Advances in Knowledge Organization 11. Würzburg: Ergon, 256-61.

Hodges, Gail. 2000. "Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files." http://www.clir.org/pubs/reports/pub91/contents.html

Mazzocchi, Fulvio. 2017. "Relations in KOS: Is it Possible to Couple a Common Nature with Different Roles?" *Journal of Documentation* 73: 368-83.

Mazzocchi, Fulvio. 2018. "Knowledge Organization System (KOS)." *Knowledge Organization* 45: 54-78.

Menne-Haritz, Angelika. 2004. *Business Processes: An Archival Science Approach to Collaborative Decision Making, Records, and Knowledge Management. Dordrecht*: Kluwer Academic.

Pombo, Olga. 1988. "From the Classification of Beings to the Classification of Knowledge Readings." *Journal of the National Library of Lisbon* 2: 19-33.

Smit, Johanna Wilhelmina. 2018. "Controlled Vocabulary and Vocabulary Control in Archives." *Acervo* 31, no. 3: 46-56.

Smit, Johanna Wilhelmina and Nair Yumiko Kobashi. 2003. *Como elaborar vocabulário controlado para aplicação em arquivos.* São Paulo: Arquivo Público do Estado de São Paulo.

Sousa, Renato Tarciso Barbosa. 2002. As bases do processo classificatório em Arquivística: Um debate metodológico. São Paulo: Associação dos Arquivistas de São Paulo.