

**RESEARCH ARTICLE****Subject representation in institutional repositories of university libraries: the perception of managers and cataloguer-indexer librarians from USP, UNESP and UNICAMP**

Roberta Cristina Dal'Evedove Tartarotti ¹  <https://orcid.org/0000-0002-0246-4971>

¹ Universidade Estadual de Campinas, Campinas, SP, Brasil / e-mail: roberta_tartarotti@yahoo.com.br

ABSTRACT

In present times, the institutional repositories enable the gathering, storage, treatment, preservation and visibility of informational resources produced in the academic field. The objective of the research was to investigate the thematic representation of documentary information in institutional repositories in the context of academic libraries, by means of the application of organizational diagnostic interviews as a qualitative methodology with the managers of the institutional repositories and cataloguers-indexers librarians from USP, UNESP and UNICAMP, with documentation collection. The systematize of the comparative analysis of the universe of institutional repositories investigated revealed the need for institutional repositories adequacy in the thematic representation performed by the cataloguer-indexer librarian. As closing remarks, the importance of the quality of thematic representation of informational resources in institutional repositories is emphasized, enabling the appropriate retrieval by subjects by the academic community.

KEYWORDS

Subject representation. Institutional repositories. Librarian's role. Semi-structured interview. University libraries.

Representação temática em repositórios institucionais de bibliotecas universitárias: a percepção de gestores e bibliotecários catalogadores-indexadores da USP, UNESP e UNICAMP**RESUMO**

Na contemporaneidade, os repositórios institucionais possibilitam a reunião, armazenamento, tratamento, preservação e visibilidade dos recursos informacionais produzidos no âmbito acadêmico. O objetivo da pesquisa foi investigar a representação temática da informação documental em repositórios institucionais no contexto de bibliotecas universitárias, por meio de entrevista semiestruturada de diagnóstico organizacional com gestores e bibliotecários catalogadores-indexadores dos repositórios institucionais da USP, UNESP e UNICAMP, com coleta de documentação. A sistematização da análise comparada no universo de repositórios institucionais investigado revela a necessidade de adequação destes sistemas de recuperação da informação quanto à representação temática realizada pelo bibliotecário catalogador-indexador. Como considerações finais, salienta-se a importância da qualidade da representação temática dos recursos informacionais em repositórios institucionais, possibilitando a adequada recuperação por assuntos pela comunidade acadêmica.

PALAVRAS-CHAVE

Representação temática. Repositórios institucionais. Atuação do bibliotecário. Entrevista semiestruturada. Bibliotecas universitárias.



JITA: DD. Academic libraries

1 INTRODUCTION

University libraries are inserted in an information system that is part of a broader system called the academic information system, in which the generation of knowledge is the object of university life (FUJITA, 2005, p. 98). Given the inevitable technological changes in the form of access to information in the most accessible format and at the lowest possible cost, and in the curricula and research interests of universities, university libraries play a strategic role in the acquisition, organization, preservation and dissemination of informational resources to the academic community of the university to which they are linked, in line with the mission of the institution.

As Thomas Friedman notes in his book "The world is flat: a brief history of the 21st century", which deals with issues surrounding globalization, "never have so many people been able to find so much information about so many things themselves and about other people. Although these aspects have always influenced university libraries throughout their history, their main role remains the same: that of supplying the information needs of their users.

As a nuclear investigative field of Information Science, Knowledge Organization encompasses the theoretical and practical foundations in order to solve problems that surround representation and retrieval in information systems. In the applied dimension, its main theoretical approaches are represented by the triad of Information Treatment - Cataloguing, Indexing and Classification - and the so-called Knowledge Organization Systems (SOCs) (lists of subject headings and terminologies, thesaurus, taxonomies, classification schemes, ontologies, etc.), tools that allow categorizing informational resources according to an organization scheme with the aim of facilitating their later retrieval (MORALES DEL CASTILLO, 2011, p. 89). Such activities are challenging and complex, both in epistemological (theoretical) and pragmatic (practical) perspective, especially in systems used by a large and heterogeneous group of users and dealing with large amounts of information, like university libraries and the web environment (MAI, 2000, p. 270; MAI, 2011, p. 116).

In the face of a scenario of significant changes in which knowledge is acquired, represented, managed and exploited, questions about knowledge in the digital world and the thematic content of informational resources have been included as the object of study of the Knowledge Organization (DAVID, 2014, p. 329; GNOLI, 2014, p. 329). The major challenges in the field of Knowledge Organization focus on "[...] avoiding information garbage, largely due to the technological context that allows, in the shortest possible time, to manage and identify large informational lots" and on the interoperability of systems, for a fast, effective and culturally significant information retrieval (GUIMARÃES, 2017, p. 92).

Contemporary themes such as institutional repositories, research data repositories, metadata aggregation and richer content raise debates around the role of university libraries (HERNON; MATTHEWS, 2013, p. 23). In this sense, there is a noticeable trend and increasing relevance of institutional repositories as systems of information retrieval both in university libraries and in the academic context in which they are inserted, also becoming instruments of university management by means of scientific indicators.

Inserted in a broader context, which is the scientific communication system, institutional repositories provide greater relevance to universities and their respective university libraries and can serve as tangible indicators of university quality through the scientific, social and economic relevance of their research activities, increasing the visibility, status and public value of the institution (CROW, 2002).

For institutional repositories in the context of university libraries to adequately fulfill their objectives, it is necessary that managers have the dimension of theoretical, practical and contextual implications that involve their planning, implementation and operation in the academic context. From this perspective, "the construction of open access institutional repositories requires efforts that precede and go far beyond the simple installation and configuration of software" (LEITE, 2009, p. 98). Institutional repositories must be linked to and managed by university libraries "because libraries are the university segment par excellence responsible for storing, organizing, managing and disseminating the knowledge produced in organizations" (TARGINO; GARCIA; PAIVA, 2014, p. 131).

In practice, the visibility of the informational resources produced by the academic community is made possible in the institutional repositories by the processes of search and retrieval of records of the informational resources, allowing to know the titles, authors or subjects that make up the collection and bringing together all the informational resources of an author or about a particular subject, and other points of access. The institutional repositories are different from the online catalogs, which, in general, allow the retrieval of records of physical or digital informational resources and indicate their location in the physical collection of the university library, whose collection includes several types of informational resources, produced or not by the academic community.

Among the activities carried out in university libraries, the subject indexing stands out, a theoretical approach of the Knowledge Organization whose purpose is to determine the thematic content of the informational resources and express it in terms of indexing, favoring the retrieval by the academic community of subjects in institutional repositories. Of a pragmatic and technical nature, "it brings together a set of norms and instruments of its own, whose operational process runs through subjective issues" (CUNHA, 1989), a characteristic that "imprints to this activity a peculiarity of its own in the face of other processes aimed at the thematic representation of information" (TARTAROTTI; DAL'EVEDOVE; FUJITA, 2017, p. 111).

The subject indexing allows for the thematic representation of informational resources aiming at the retrieval by subjects in institutional repositories. In this intellectual process of thematic content assignment carried out by the cataloger-indexer librarian, the adequate use of documental language is fundamental for the quality of the representation of the issues of the informational resources in the institutional repository, favoring the adequate correspondence between the representation and the retrieval of the information by issues by the academic community in a certain specialized scientific area. The subject indexing (process) and the institutional repository (product) perform a mediation between the university's informational resources and the informational needs of the academic community.

In these information retrieval systems, the thematic metadata of the informational resources refer to the fields for describing the subjects, which can be keywords (natural language) or terms/descriptors (controlled language), for example, thesaurus or taxonomies of specific domains, provided by the author himself, cataloger/indexer librarian, or coming from a social indexing process characterized by folksonomies. However, as with other information retrieval systems, when there is no compatibility between the language of the institutional repository and the user's search language, "the credibility of the system is shaken, caused by a representativeness that does not match the investigative needs of these users" (BOCCATO, 2009, p. 71).

Based on the relevance of the adequate subject retrieval by users both in the context of university libraries and in the academic context, it can be observed that research on thematic representation in institutional repositories is still incipient. In this sense, the objective of the research was to perform a diagnostic study of the thematic representation of documentary

information in institutional repositories in the context of university libraries by the perception of managers and cataloger-indexer librarians who work in these information retrieval systems, in order to contribute to the improvement of the representation and retrieval by subjects in these informational environments.

2 METHODOLOGICAL PROCEDURES

The research approach adopted is the exploratory-descriptive one, which focuses on the delineation of information and formulation of hypotheses on the subject and identification and analysis of the characteristics that relate to the phenomenon in question.

The research universe were institutional repositories of the University of São Paulo (USP), Paulista State University (UNESP) and Campinas State University (UNICAMP), which has been consolidating itself as a reference in the management and dissemination of informational resources produced by their respective academic communities in the Brazilian context, meeting the demand of the Foundation for the Support of Research of the State of São Paulo (FAPESP) with the Council of Rectors of State Universities of São Paulo (CRUESP). They use DSpace as a software platform, whose source code is made available in order to form an open source community in the addition of resources, improvement of different functions and adaptation of requirements to meet the needs of institutions (GONÇALVES, 2011, p. 727-728).

The Digital Library of Intellectual Production of USP (BDPI)¹ was created on October 22, 2012 with the aim of bringing together the intellectual production (scientific, academic, artistic and technical) of the university, in accordance with the Information Policy of USP defined in Resolution No. 6,444 of October 2013, which aims to ensure the collection, treatment, dissemination, accessibility and preservation of the university's intellectual production (USP, 2019).

UNESP²'s Institutional Repository aims to "store, preserve, disseminate and enable open access, as a global public good, to the scientific, academic, artistic, technical and administrative production of the university" (UNESP, 2019). It is guided by the Internal Regulations of the UNESP Institutional Repository.

The Repository of Scientific and Intellectual Production of UNICAMP³, formalized by Resolution GR-013/2015 and officially launched on November 2, 2015, is the official information retrieval system for the collection, organization, dissemination and preservation of the academic community's production. Its main objective is "to provide open and public access to the scientific and intellectual production of the university, providing increased visibility, accessibility and dissemination" (SYSTEM ..., 2019).

Created from the institutional repositories of the three universities, the Repository of Scientific Production of CRUESP⁴, launched on October 6, 2013 during the opening session of the Portuguese-Brazilian Conference on Open Access (CONFOA), aims to "bring together, preserve and provide open, public and integrated access to the scientific production of teachers, researchers, students and servants. It also uses DSpace, adopting international standards and

¹ Available at: <http://repositorio.usp.br>.

² Available at: <http://repositorio.unesp.br>.

³ Available at: <http://repositorio.unicamp.br>

⁴ Available at: <http://cruesp.aguia.usp.br>.

norms of standardization and interoperability. The integration between the three repositories is done by the Primo metasearch, which allows the search and discovery through a single interface (CRUESP, 2019).

Methodologically, the interview is adopted in order to obtain qualitative, descriptive and detailed data of an individual and when the nature of the data is complex, favoring, unlike the questionnaire, a higher level of interaction between the researcher and the participant. It has been widely adopted in research around information in libraries (PICKARD, 2013, p. 195-196; 323). Each type of interview depends on the objective of the research, the nature of the subject addressed, the experience of the researcher, access to participants, the time available for data collection and the type of data to be collected, with the aim of raising answers to the research question (PICKARD, 2013, p. 205; 198).

The structured interview is based on a formal interview schedule (usually a questionnaire administered by the researcher), which specifies the formulation and ordering of questions to the participant and is subdivided into an open interview (free answers) and a closed interview (answers based on a set of alternative answers) (PICKARD, 2013, p. 323; 199). The unstructured interview is based on open-ended questions, which allow the participant greater freedom to discuss the subject and influence the direction of the interview, since there is no predetermined plan on the specific information to be collected from the participants (PICKARD, 2013, p. 323). In this sense, the survey of the perception of managers and cataloger-indexing librarians who work in institutional repositories with regard to thematic representation was based on a semi-structured interview, outlined below.

3.1 | Procedures for applying the semi-structured interview

The participants were defined in two categories: managers and cataloger-indexer librarians. In order to identify the activities developed in each stage, procedures were outlined before, during and after the application of the semi-structured interview, based on the qualitative methodology of the Verbal Protocol (PV):

3.1.1 Procedures prior to the application of the interview

- a) Definition of the survey universe:** the organizational contexts of USP, UNESP and UNICAMP were listed in order to collect data on institutional repositories through semi-structured interviews;
- b) Selection of participants:** the selection of participants for the interview resulted in the choice of a total of nine participants, namely: three managers of the institutional repositories; three librarian cataloger-indexers, being one participant from each category, from each university. An informal conversation was held with each of the participants by phone or e-mail, resulting in the acceptance and definition of the interview dates;
- c) Definition of the interview:** prior to the formal application of the interview, a pre-test (pilot) interview was conducted with a manager and a cataloger-indexer librarian from the Institutional Repository of the Federal Technological University of Paraná (UTFPR), in order to verify whether the questions listed were pertinent to the interview proposal of this study. According to Almeida (2005, p. 60), "the application of the pre-test or pilot test is a preliminary activity to the implementation of the project and aims to test the design and methodology of the project," ensuring "the suitability

of the instruments or processes to be used, as well as measuring the time to be spent on evaluation;

d) Collection of documentation: according to Almeida (2005, p. 58), other essential sources for organizational diagnosis can be extracted from reports, regulations, work plans, etc. Documentation on institutional repositories was surveyed at the USP, UNESP and UNICAMP Library Systems Portals, as well as requested from cataloger-indexer managers and librarians at the institutional repositories;

e) Selection of the base text: an interview script containing a list of open questions (Appendix) on the subject was used, divided into two parts, according to each participant's profile in order to: a) characterize the institutional repositories in the context of the participating universities in terms of administrative structure - applied to the managers of the institutional repositories; b) observe the subject indexing in the institutional repositories - applied to cataloger-indexer librarians. The interview script was based on an organizational diagnosis proposed by Almeida (2005, pp. 53-55); on a questionnaire applied by Gomes (2015) that identified institutional repositories of Brazilian federal universities in the standardization of metadata in the representation of information; and on Dal'Evedove (2014), whose application characterized the informational context of cataloger-indexer librarians regarding the indexing policy;

f) Informal conversation with the participants: the objectives of the research were mentioned and each manager and cataloger-indexer librarian was previously given the Term of Free and Informed Consent (TCLE), requested the signature of the document and formal acceptance of participation in the research, leaving a copy for them. It was emphasized that the identity of each of the participants would remain anonymous, with the purpose of not compromising the data and leaving them at ease during the interview in the most natural way possible.

| 7

3.1.2 Procedures during the interview

a) Interview and recording of participants' speeches: Skype software was used for the interviews at USP and UNESP, while at UNICAMP they were done in person. The recording was done through a cell phone voice recorder application. According to Pickard (2013, p. 205; 205), there is still no consensus in the literature about the use of online interviews. If on the one hand some visual, verbal and behavioral cues of the participants may be lost during the interview, the ease of communication and savings of resources and physical displacement are still advantageous. While the synchronous interview is done a posteriori, either by using email or other resources, sent by the interviewee to the researcher in a certain designated period, the asynchronous interview, as adopted in this research, is done in real time between the researcher and the interviewee. The intervention was made so that the list of questions could be covered during the data collection. Both the informal conversation and the interview itself were applied individually with each participant, respecting their own individuality and needs.

3.1.3 Procedures after the interview was applied

- a) literal transcription of the participants' speech recordings:** literal transcriptions were made, preserving the participants' identity by means of specific acronyms according to each category (manager or cataloger-indexer librarian) and the institution to which they belong. In order to allow a better visualization of the processes externalized by the participants, specific notations were adopted in order to analyze the data from the applicability of the Verbal Protocol (PV) methodology, widely adopted in Information Science, adapted by Tartarotti (2019, p. 359);
- b) Detailed reading of the data:** a detailed reading of the transcripts was made in order to search for significant phenomena in order to elaborate units and categories of analysis;
- c) Elaboration of units and categories of analysis:** elaborated based on the theoretical reference of the literature and the statements of the participants;
- d) Reading of data:** transcripts were re-read in order to remove excerpts that best exemplify each category of analysis, with a synthesis of the main aspects observed.

The results of the application of the methodological procedures with managers and cataloger-indexer librarians are presented below.

3 RESULTS AND DISCUSSION

The analysis of the data collected in 2019 was based on units and categories of analysis, with presentation of excerpts from the participants' statements on representation and retrieval by subjects in institutional repositories. Specific acronyms were adopted for the participants according to the academic category and the institution to which they belong, being G-Manager and I-Librarian cataloger-indexer. Based on the objective of the research, the theoretical framework and the interviews, seven categories of analysis were outlined, distributed in three analysis units: 1) Context-related aspects; 2) Aspects related to the cataloger-indexer librarian and 3) Process-related aspects.

4.1 Aspects related to the context of institutional repositories

It refers to the decisive issues of the functioning of the institutional repository, with the categories: 1) Implementation of the institutional repository; 2) Policy of operation and documentation of the institutional repository; 3) The user of the institutional repository; and 4) Actions to improve the institutional repository.

4.1.1 Implementation of the institutional repository

Historical path of implementation of institutional repositories. The main actions implemented in order to institutionalize the institutional repositories of USP, UNESP and UNICAMP were the choice of software, methodology design and collection and standardization of metadata of the respective universities' productions, in order to meet the deadline stipulated by FAPESP. Among the challenges of automatic collection, observed by the manager of UNESP's repository, the lack of standardization in the name of the university and in the authors' names in the sources of origin: Curriculum Lattes, PubMed, SciELO, Scopus and

Web of Science stands out. Librarians were hired to work specifically in the institutional repository for a certain period of time.

4.1.2 Institutional repository operation and documentation policy

General guidelines established for the functioning of the institutional repository, documentation and active team.

USP, UNESP and UNICAMP documentation are available on the pages of the respective institutional repositories. At USP, the documentation is updated by the management team in order to meet the guidelines of Ordinance 01/2019 of FAPESP⁵, especially regarding the availability of the PDF of the informational resources, because in the institutional repository of USP the access to them is made by links. At UNESP, there is no forecast of documentation revision. At UNICAMP, after the formalization of the institutional repository other actions were outlined in order to increase the number of informational resources, however, at a lower level of detail in the treatment of metadata. There is a need for resolution of copyright issues with the Rector.

As for the current team, at USP the management of the institutional repository is done by the Technical Department of the USP's Integrated System of Libraries (SIBi) (centralized), while the insertion of data, metadata and digital objects is done by the Libraries (decentralized). It is formed by three librarians, being one manager; two librarian-catalogers, responsible for answering questions about the informational resources to be inserted in the institutional repository, checking consistency, data quality and copyright. Centralized in the institutional e-mail or telephone service, the communication system receives questions, suggestions or demands. The team also has a librarian-analyst, who works in information technology.

At UNESP, the team has a manager and two librarians, who perform the chat service, due to the large volume of questions generated, especially after the implementation of self-archiving for graduate students. Recently, there has been a decentralization to the Libraries, which are responsible for completing in the institutional repository the thesis and dissertation records inserted by graduate students. At UNICAMP, the team consists of a manager and two librarian-catalogers, one of whom is a supervisor who works specifically in the institutional repository and the other who also contributes with other activities as required, in addition to a systems analyst, scholarship holders and intern who collaborate in the insertion of the records. The performance of the professionals is more focused on descriptive representation, not directly with thematic representation.

At USP, the types of content inserted in the institutional repository are journal articles, book chapters, reviews, in print or post-print versions, as well as theses and dissertations already inserted in BDTD. At UNESP, the main contents are journal articles inserted at the beginning of the institutional repository implementation and theses and dissertations. There is no provision for inserting other types of materials or retrospective informational resources. At UNICAMP, the main contents are journal articles and patents, including other types of material, besides theses and dissertations.

⁵ Establishes the "Policy for open access to publications resulting from FAPESP aid and grants", which applies to open access publications in repositories of any type of scientific production that contains results originated from research financed, partially or fully by the institution.

4.1.3 The user of the institutional repository

Perspective of the professionals who work in the institutional repository on the academic community that uses it, especially regarding the retrieval of informational resources. As the manager observes, between 2014 and 2016 a study was carried out at USP around user searches in the institutional repository, with identification of retrieval problems.

With regard to searches made in DSpace, USP's manager and UNICAMP's cataloger-indexing librarian agree on the existence of problems in the retrieval by users, due to the tool's specific characteristics regarding the level of accuracy and revocation of the system. At USP there is further refinement, enabling specific searches in the "Title" and "Authors" fields, while at UNICAMP there is no differentiation between the search options "Title", "Author" or "Publication date". However, the use of the graphical quotation mark ("") allows a higher level of precision. On the other hand, in both institutional repositories there is no possibility of refining searches in the "Subject" field, a situation shared by the Dublin University's institutional repository, taken as a model for the institutional repository to be followed by the three universities of São Paulo, in the conception of FAPESP.

4.1.4 Actions to improve the institutional repository

Actions not implemented in the institutional repositories, but that would contribute to the improvement of descriptive representation and thematic representation.

In general, the USP manager observes the need for continuous improvement in the institutional repository and few adjustments regarding the assignment of subjects in this information retrieval system, for example, insertion of subjects of controlled vocabulary in real time. Although the VOCAUSP Management Group is active, it assures that the aspects related to the exhaustive controlled vocabulary are due to the heterogeneity of the university context, which deals with diverse and specific areas of knowledge. According to UNESP's manager, international repositories are already dealing with research data repositories and in the Brazilian context there are still several improvements to be implemented in institutional repositories. In terms of thematic representation, she considers it relevant that the subjects used in informational resources (controlled language of the specialists) are compatible with the subjects used by users (natural language) during searches in the institutional repository, providing a more effective retrieval.

So, for example ...~~ more and more searches need to have subjects, the subjects need to be, is ... (...) intertwined, like the vocabulary of the person, right? (...) because it interferes a lot in the question of the search. (...) right ... the question of metadata we need to improve more and more. (...) we always think twice in all fields that we use ... and that we work with. (...) Regarding [the] subject (...) ... to have more and more a greater (...) integration, link between language, try at least to get closer to, the natural language and ... and that of the specialist, which is what we have controlled (G-UNESP)

In UNICAMP's view, regardless of the software used, there is a need to improve the interoperability of the institutional repository with the Acervus online catalog, making it possible to use the Issues Table, seeking greater standardization in the records of the informational resources inserted in the institutional repository.

[...] what we need is to improve this integration together with SophiA or to remain only with SophiA for this, so that we can use the own tables already available by

SophiA to standardize all this, to make all this control of... standardization of the data itself, nay, linked to these metadata (G-UNICAMP)

4.2 Aspects related to the cataloger-indexer librarian

Role of the cataloger-indexer librarian in the institutional repository, in the category:
5) Performance of the cataloger-indexer librarian in the institutional repository.

4.2.1 Work of the cataloger-indexer librarian in the institutional repository

Courses and training sessions held by cataloger-indexer librarians for professional updating in institutional repositories.

The three universities are concerned with the continuous training of professionals working in the institutional repositories. At USP, the courses and training are directed to cataloger-indexer librarians who work in the SIBi Libraries and are given in partnership with the management team. At UNESP, it is important to highlight the realization of on-demand training or the need to change procedures, both for the professionals of the Libraries and for those who work in the postgraduate secretariats, in order to integrate the two sectors in the flow of theses and dissertations at the university. Considering the decentralization of the activity since 2018, at UNICAMP the courses and training are directed to librarians and technicians of the Libraries responsible for the standardization of informational resources in the institutional repository.

| 11

4.3 Aspects related to the documental process in institutional repositories

Questions about the process of cataloguing/indexing matters in the institutional repository, with the categories: 6) Descriptive treatment of information in the institutional repository and 7) Thematic treatment of information in the institutional repository.

4.3.1 Descriptive treatment of information in the institutional repository

Aspects on the descriptive representation in the institutional repository: insertion of materials, self-archiving, manual of descriptive representation and standardization of authors' names.

At USP, two information retrieval systems are used to register informational resources: The Digital Library of Theses and Dissertations (BDTD), which contains theses and dissertations produced at the university, and the Digital Library of Intellectual Production (BDPD), with the other records. There is no link between the two databases, being presented in the Integrated Search interface, which allows the visualization of metadata. Theses and dissertations are inserted into the BDTD by self-archiving graduate students and the records are exported to the online catalog DEDALUS and validated by librarians of the Libraries.

(Always from DEDALUS to the repository, never the other way around, right?)
Correct, never the other way around (I-USP)

However, by the perception of the cataloger-indexing librarian, in practice they are being inserted into the BDTD by the Libraries themselves. According to the manager, for the other types of informational resources there has been an attempt to implement self-archiving, both by the author himself and with validation by the librarian. However, the lack of metadata control opened space for the lack of metadata standardization, compromising quality and causing problems around the legal issues involved. With the recent orientation of FAPESP that attributes the responsibility for uploading production to the Libraries, the possibility of self-archiving by the authors becomes less likely. In the perception of the cataloger-indexing librarian at USP, in practice, the self-archiving of other types of informational resources by the authors does not work, because there is some resistance in assuming this activity.

[...] we would like to implement self-archiving, we have some Units that demand it from us ... but ... in a way ... holistic way, so ... I think it will take a long time. Unfortunately. (...) It would be by the author himself the proposal. But the Libraries have the profile to ... deposit for the authors. (...) ... In the past we did both ways. ... Yes ... we did the self-deposit with the. ... the declaratory, from the author himself, without any kind of validation by the librarians and we did it with validation, (...) but if one day we come to ... fully implement the self-archiving ... we will probably need a librarian's validation (G-USP)

[...]in the Library we register the material in the database, in DEDALUS. (...) From the production, all the material we do in DEDALUS, and from DEDALUS goes to the repository, which is relatively recent... (...) So when we think about the treatment, we don't think about the repository, we think about doing it to DEDALUS, which is our database, right? (...) Because the submission process would actually have to be the student to do it and then the ... the postgraduate would do the revision and we would do the cataloguing, but this does not happen (I-USP)

At UNESP, the most prominent type of informational resource shared both in the Athena online catalog and in the institutional repository are theses and dissertations. There are also general e-books and those edited by the university's publishing house, which are inserted into the Athena online catalog and exported to the institutional repository, remaining with the same subjects already assigned by cataloger-indexer librarians. The metadata of theses and dissertations are inserted by graduate students in the institutional repository by self-archiving, on the occasion of thesis or dissertation defense. Afterwards, they are exported to the Athena online catalog, where the process of validation of the metadata by librarian cataloger-indexers of the Libraries occurs. In this sense, there is no professional who performs the cataloging of the informational resources specifically in the institutional repository.

[...] actually the dissertations are in the repository, but before that, the Libraries catalogued in the catalog ... and we sent these records to the repository. (...) Today it is the opposite, as I work with self-archiving the procedure starts in the repository, the students do the self-archiving ... and from there we send the already validated and complete record to the Aleph (G-UNESP)

Then the person does the self-deposit ... the theses get here ... to us ... (...) I open them, I work with the MARC, what do I do? I see... I see the thesis, the document on paper, sometimes it happens to have some difference in the title for some reason so we have to have both open in order to define... the valid ... is the paper. (...) So something needs to be changed it will be changed, with the proper ... permissions on the electronic document. (...) So the paper is with me, the electronic document is open ... and the work is done ... in the MARC, in the Aleph bibliography (...) It was a need of the academic community in general (...) to establish this policy ... of theses and dissertations to become more visible and to populate more also the ... the repository. (...) É ... Even we gain a little time because before we described it from scratch and now the author does one already ... he already describes everything although not everything has the quality of the standard but ... it simplified a little (I-UNESP)

As the librarian cataloger-indexer observes, the self-archiving, implemented about six years ago by a need of the academic community, facilitates the process of cataloguing as to the typing of metadata, since graduate students perform a pre-cataloguing of theses and dissertations.

At UNICAMP, theses and dissertations are catalogued in the Acervus online catalog by librarian-catalogers and exported to the institutional repository, maintaining the descriptive and thematic metadata. More careful attention has recently been paid to standardizing the metadata in the institutional repository. The implementation of the self-archiving process depends on the university's approval of the institutional repository guidelines. The proposal is that a work be carried out with the retrospective production collection Libraries and that the authors themselves, the Libraries or postgraduate secretariats have autonomy in the self-archiving of the production in the institutional repository, by means of a specific form based on the current system of cataloguing records. The metadata will be validated by cataloger-indexer librarians in order to obtain higher quality records of informational resources.

[...] there is an application profile that we follow in relation to the metadata, now the standardization itself of the data itself... must be done, ok? We have started this work of standardization... such as authorship as subject matter (G-UNICAMP)

[...] today they talk, SophiA talks to DSpace. So nothing, nothing that deals with thesis and dissertation we can do on DSpace, nothing. Everything we do on SophiA. Any dot, even a comma that we change in SophiA it changes ... on DSpace. (...) It's automatic (I-UNICAMP)

The three universities use the Dublin Core metadata scheme in their institutional repository. The insertion of records of other types of informational resources in the institutional repository is done by automatic collection, with import of records from PubMed, SciELO, Scopus and Web of Science databases and validated by librarians for the aspects described before insertion into their respective institutional repositories. At UNESP, the Curriculum Lattes is also used as data source. While at USP the records of its users' production are distributed to the Libraries, at UNESP and UNICAMP the validation is done by the management team of the institutional repository. For both cases, it is observed the lack of standardization in the records of the informational resources originated from the databases, both referring to the descriptive (affiliation, authors' names, etc.) and thematic aspects (subjects).

At USP there is a cataloguing manual in the online catalog DEDALUS, which contains a tutorial for inserting informational resources and procedures for requesting the opening of new terms in the thematic representation. UNESP and UNICAMP also use a cataloguing manual. However, in the three universities there is no specific descriptive representation manual for professional performance in institutional repositories.

The standardization of authors' names at UNESP is done only in the Athena online catalog, when the cataloger-indexer librarian uses the Table of Authors of the University itself. At UNICAMP, the standardization in the Acervus online catalog is made in accordance with the author's preference, adopting remissive for other entries. The implementation of ORCID in the three universities is in progress.

4.3.2 Thematic treatment of information in the institutional repository

Discusses the aspects of the representation of subjects in the institutional repository: cataloguing/indexing subjects of theses and dissertations and other materials, indexing policy

and indexing manual, documental language and its updating, number of assigned terms, evaluation of the subject indexing and suggestions for improvements pointed out by professionals.

At USP, the cataloging/indexing of theses and dissertations subjects in the institutional repository is done directly in the Digital Library of Theses and Dissertations (BDTD) by self-archiving, being used free terms or keywords assigned by the graduate students themselves. Later, in the Library to which they are linked, records and respective subjects are imported and validated in the online catalog DEDALUS, aiming at making natural language compatible with controlled language (VOCAUSP). The free terms remain in field 952 of MARC, while the terms already validated by cataloger-indexer librarians remain in field 650. Both natural language and controlled language are recoverable. The elaboration of the cataloguing sheet occurs in an online request system, where VOCAUSP is available for consultation by the user. However, in the perception of the cataloger-indexer librarian, the controlled vocabulary is not consulted by graduate students, preferring natural language.

The issue with the subjects is like this, we have a controlled vocabulary, ok, institutional. (...) ... So in this validation made by librarians they already check the subjects that are placed and already do if it is necessary ...the translation of the subjects of controlled vocabulary (G-USP)

[...] the indexing is actually done in DEDALUS but in BDTD, as it has the keywords, these keywords come to DEDALUS and stay in a field of words that we afterwards in the vocabulary do an analysis also to ... include these keywords in the vocabulary (...) And the form is for them to do online and in this form there is a link on the side for them to consult the USP vocabulary for terms. (...) But what I understand is that it is not consulted. (...) Usually it's not consulted, usually they put the terms of the keywords themselves. (And ends up entering these terms of the keywords in the digital library of theses). ... Yes. (And you check it). ... That's it. (Along with the USP vocabulary, right)? Yeah, that's it. (So in the Digital Library of Theses and Dissertations the terms are free, let's say, that they put, but in DEDALUS they have this validation from the librarian, they are the terms) ... There are both the terms that they put and the vocabulary terms, what we do we don't stop having the field of keywords, we take those terms out of that field of keywords, we leave them in our subject field, and ... as we use the MARC, our subject field is 650 and the keyword field is 952, which is the keywords suggested by them. So the words in the vocabulary are in the field 650 and the words that are not in the field show in the ... 952 (I-USP)

| 14

Because they are different databases, it is possible that inconsistencies may occur between the subjects assigned in the BDTD and the BDPI, especially regarding the qualifiers or sub-headings of the subjects. In searches for the same register in the online catalog DEDALUS and in the institutional repository (BDPI), it is observed that in the first information retrieval system the subjects are more complete.

[...] because they are different bases and this metasearch that we put ... the vocabulary part ... that is in the BDPI respects the criteria of the controlled vocabulary and that of the BDPD not necessarily. The repository uses it. (...) The BDPD does not. (...) That I would say ... 20% of the records can happen. (...) No, it's just that it's more... let's say, it is more controlled ((RI)) (G-USP)

And as they also sometimes put together, they throw a word that we put ... (...) ... that we put as a complement there and they use as a separate term. What happens is that we put all this together. It is the qualifier. (...) There is the qualifier and the place, so there is a ... a subject ... which englobes more, one thing only when sometimes it is separated in the keywords. But when that happens ... I take it out from the keywords, because in a certain way I used it in field 650. (...) ... if he is not from DEDALUS. There are the terms that I put as well as the terms that the student put (I-USP)

At UNESP, considering that the informational records of theses and dissertations are inserted in the institutional repository by self-archiving, the subjects are assigned by the graduate students themselves and automatically inserted in this information retrieval system. Afterwards, the records are imported into the Athena online catalog in field 650 of MARC, when the subjects undergo validation by a librarian cataloger-indexer of the Libraries, adopting UNESP Thesaurus as controlled language. If it exists in controlled language, the subject is maintained in the Athena online catalog. Otherwise, the possibility of using a corresponding subject is verified or, when pertinent, its insertion in the UNESP Language Group is suggested. According to the perception of the cataloger-indexer librarian, in the majority of the subjects originated from self-archiving there is a need for alteration. Requests for catalographic records are made in an online system. The professional believes that in some cases, the definition of keywords based on the Thesaurus UNESP is a challenge to graduate students, who end up opting for freer keywords.

[...] the candidate to master or doctor he does not have this care, in fact he does not have the training, it is not always easy to define the words because of the subjects, our previous knowledge. (...) So for him it's much easier to simply put those words that he ... understands that are the best to close his subject (...) And enters that thing a little more intellectual that is the question of defining the field 082, which is the subject (...) and the fields ... 600, which are ... entity, the personal name or the topic subject itself. ok. (...) (So you use classification for theses?) ... Yes, here we do, yes, here in Marília they are classified (...) I even notice that sometimes we have a difficulty, so we look for some thesis with a subject, with some similar word to see if we find a ... a ... a loose thread for us to take advantage of that work of cataloguing the definition of the subject (...) ... we work to ... determine the subject of the thesis or dissertation (...) then I check if that term ... it exists, right, in ... in the bases of ... in the authorities of terms there at UNESP and if it is ... convenient the, ... description of the material. Sometimes it may even be from the Area but suddenly the author ... misjudges, so we do this verification, so are there are two verifications, one of their existence, if it exists, ... and it is pertinent it is maintained. (...) If it ... does not exist I need to find ... some synonym or some other, or other words, sometimes more than one that helps to ... to compose the subject, that finishes ... the term ... let's say (...) and ... I fill in this way. I mean, checking the consistency of the ... of the term and whether it is pertinent to ... to work. (...) ... the majority, the great majority, we have to ... let's say, redo, because it's like you said, it's natural language, not only natural language but ... is ... depending on the area, so, or ... and the specificity of the work has some words that are very ... particular and those are not attended, sometimes almost a neologism. (...) So we end up having to work always, it's very rare to have a ... a job that we ... don't touch (I-UNESP)

In the manager's view, the validation of the subjects according to the controlled vocabulary and the extension of the terms suggested by the graduate students for dissertations and thesis in the Athena online catalog are important points to be considered. However, the same subjects remain in the institutional repository with free terms in natural language, without a vocabulary control.

Look, I think they should take more or less what comes, right? The students should insert there as ... recommendation ... (...) I think they should, use it, or include more (...). I would include more, I would take and enjoy theirs ... and go for vocabulary too (G-UNESP)

At UNICAMP, the cataloging/indexing of thesis and dissertation subjects is done by cataloger-indexer librarians without their own unique vocabulary, and several controlled languages are adopted, according to each area of knowledge. As for the other informational resources inserted in the institutional repository, the proposal is that it contemplates both controlled language and natural language (free terms or keywords listed by the authors

themselves). In this case, without the adoption of "see" and "see also" type remissive, adopted in controlled language and the subjects will be validated by a librarian, and then incorporate the other terms of controlled language. As the manager observes, such measure will enable improvements in the retrieval of informational resources in the institutional repository. The implementation of self-archiving for the authors is conditioned to the validation of the subjects by professionals, both in the descriptive and thematic representation. An online system is also used to request cataloguing sheets.

[...] we would do it in the form of theses and dissertations, that is, every subject, every subject linked to ... to a document it would be standardized, all treatment before entering the repository, which is that validation I commented just now (...) There is a metadata that ... that allows us to insert the natural language, it would be the free terms, but natural language itself, even those loose keywords that are linked to documents, would remain only themselves. They would stay there only to ... as a plus of retrieval, but what would validate even with remissive and everything else would be the descriptors formed there on the subject, now the ... the free terms that we will adopt is what comes in the document many times ... (...) Those subjects that the base brought I am calling those not treated subjects, free terms. (...) And then from the moment it is treated and everything else becomes a descriptor, with the subjects controlled even (...) little we have worked with subject in the repository, so. We (...) usually keep the subjects that already come with our own documents. (...) We don't do a job even with, with translation into documentary language, no, this is a later stage yet, we don't do this treatment ... fine of the subjects. (...) ... with the help of ... of the Libraries, of the other professionals of the other Libraries we have made this ... this standardization of subjects, this very entry, this search and treatment of the subjects in the repository, but not today, it does not really reflect, it is ... the refined treatment of the subjects (...) [...] we will make a validation as is done today for theses and dissertations (...) ... So as it is today in the thesis we will, it is ... will do for ... for repository documents (G-UNICAMP)

As for the indexing policy, at USP we observe its formalization in an indexing manual, which at the moment undergoes reformulation and is available to professionals in the technical area. The indexing manual is directed to the professional performance in the online catalog DEDALUS, a retrieval system where the description of metadata is made. There is no specific indexing manual for the institutional repository. As perceived by the cataloger-indexing librarian at USP, the indexing policy and the controlled vocabulary in force meet the university's needs. However, she points out the need for greater collaboration from the Libraries in requesting new terms.

We have an indexing manual yes. How to request the term, and such... (And in this case this indexing policy is general, there's no specific repository, right?) No, we didn't think of it as a repository, the repository here is now that it's starting ... to create a body (...) So... it was never thought of as a repository, it was always thought for the DEDALUS bank (...) (Do you think that this vocabulary satisfactorily meets the indexing ...) Look, I think it meets, but ... it all depends a lot on each ... each Unit to request the terms that it does not have, right? (I-USP)

At UNESP, a policy of indexing is observed in the Athena online catalog, formalized in an indexing manual that contains examples in order to guide the practice of subject indexing by professionals. As perceived by the cataloger-indexer librarian, the current indexing policy satisfactorily meets the standardization of informational resource subjects. One point to be highlighted is that, in the view of the manager, such guidelines established in the Athena online catalog can be adopted in the institutional repository.

We thus have a ... an apostille, which is based on the criteria defined by a commission (...) So this apostille comes with, with many examples, but to close, try to close the

possibilities there (...)... the documentation (...) is very good and meets a lot of yes to the ... the needs (I-UNESP)

So politics as far as I know is for ... is for the catalog but I believe ... it ... I think it can help in ... in the repository. It's because ... it's UNESP's production (G-UNESP)

At UNICAMP there is no indexing policy formalized in an indexing manual, no partnerships with professors specialized in knowledge areas or the institutionalization of a working group that can establish the indexing policy at the university. There are internal guidelines that are followed by professionals, learned in professional practice or in training. However, in the perception of the cataloger-indexing librarian, the subject indexing language adopted for the standardization of subjects satisfactorily meets the needs of the institutional repository. With the interoperability between the Acervus online catalog and the institutional repository, it will be necessary to update the procedures by using the Catalogue's own Tables of Authority.

[...] now everyone who starts working in the repository will have to follow. (...) This is ... manual is inside Excel, (...) if it really passes to SophiA it will ... change because then it will be in MARC (...), which will make our life much easier. (...) ... Then it will ... it's not that he dies, it will be, it's the same information but it will make it easier because we have the tables (...) At least that's what we hope, that he doesn't have it anymore, it's all inside SophiA (I-UNICAMP)

Regarding the documental language adopted, VOCAUSP is the controlled language developed by USP to standardize the subjects. Created in 2000, it was based both on controlled languages adopted by the Libraries and on the consultation of expert teachers. At UNESP, the most adopted controlled language is Thesaurus UNESP. For new terms, the Librarian cataloger-indexer's area of expertise is the Library of Congress Subject Headings (LCSH), the National Library Subject Terminology, or the term is sent to the UNESP Language Group for analysis and approval. At UNICAMP, there is no single vocabulary of its own, and several sources are consulted, the main ones being: Library of Congress Subject Headings (LCSH), National Library Subject Terminology, Health Sciences Descriptors (DeCS), FSTA Thesaurus and Ei Compendex.

We have our USP vocabulary, right. (...) It's what we use, the USP vocabulary it's been since 2000 (...) that we use and ... and it's updated, it's dynamic, we have a management group of which I'm also part ... and it is updated as people notice new subjects or subjects that do not have in the vocabulary they are inserted ... within a hierarchy and it's above this vocabulary that we index, it's with it, we work with it (I-USP)

[...] if I don't, there is a term, that term is not valid, or I look for another similar one, or others who do that. (...) Or it is possible... é ... sometimes we look in the ... in the National Library, if that term exists, (...) the one that in our base still doesn't ... does not exist in our base, it may be that it exists in the National Library, in the LC, (...), in the Library of Congress or we may request the Committee ... (...) (I-UNESP)

Ah, the theses what we do is take from Pesq, BN, OCLC and has a base from ... FEA ... of food that they take. And some take thesaurus or we ... is ... make a subject, putting definitions and ... positive references, right? (...) The same thing is happening to the article. That's why it was passed to the Units, because what they are doing is the same policy (I-UNICAMP)

As for updating the language, at USP the subjects not contemplated in the controlled vocabulary are forwarded to the VOCAUSP Management Group, formed by librarians representing the three major areas of knowledge and expert teachers, for analysis and approval and, later, dissemination of the new terms to the Libraries. The constant updating is in line with

the scientific production of the university and the possibility of improvement in the automation of the process is under analysis.

[...] we are automating, we have already automated a lot of things but we want to automate even more for (...) greater agility in the insertion of new terms ... ~ I would say that we have had good results, we have problems, it is a continuous work of improvement, of improvement (G-USP)

And as ... the idea was to be a vocabulary that was valid for the whole USP ... because the other used was well ... general, it was quite general, were terms ... I don't know exactly where it came from, but I only know that the indexing was to be desired because they were very general matters. (...) The management group really think it was a little further, because then we saw the need to have people to ... to be following and to be feeding the new terms that were ... (...) In this vocabulary management we have representatives from three areas, which is Human, Exact and Biological. (...) The subjects are requested through the ... n, a platform, and are requested by the Libraries. Each area has two, from two to three representatives who analyze this request ... will consult the sources, because they have to come with the request and with a note of scope we will verify if it is suddenly a subject to be inserted even or if it is not a remissive of another already existing (I-USP)

At UNESP, the UNESP Language Group is formed by the General Coordinator of Libraries (CGB) and librarians representing the Libraries. Requests for new terms are analyzed by the group and, if approved, are incorporated to the Thesaurus UNESP. At UNICAMP new terms are opened, provided that they are adequately referenced with the respective sources and remissive in the Table of Issue Authorities of the Acervus online catalog.

It is, if I make this request I believe that ... be evaluated and accepted or not in two or three days, let's say so. (...) When we make the request we already indicate a source (...) to support, let's say, the quality of the ... of the term, (...) So they usually accept yes. (And you have to do the ... you have to describe the term in Aleph right, do the remissive, indicate the source). Yes. (The tree of authority of the subject?) Yes. It would be in those cases; it would be that (I-UNESP)

What they already have in SophiA they are using what they have in SophiA that is already standardized, what they don't have in SophiA if need be they are creating or going after references to put, with the definitions. ... So we are using the same policy (I-UNICAMP)

As for the number of assigned terms, at USP the guideline is to use no more than six terms, with a variation from one to six assigned subjects. At UNESP, the policy leads to the assignment of at least three subjects, with no limit of assigned terms, with the use in Portuguese and its version in English. At UNICAMP the minimum is one and the maximum of six subjects, the same orientation applied to theses and dissertations.

Look, actually we advise you not to ... that it does not exceed 6 ... 5, 6 terms ... but ... has no limit, the person can put as many subjects as he thinks necessary (I-USP)

[...] I don't know this criterion, but I don't (...) ... we work with these three terms in Portuguese and ... (...) one in English, which is 650 zero, so we ... we end up having to put this one too. (So the limit of terms in Aleph is three, is that it?) Yeah, that's the minimum. (...) (And the maximum, you have a maximum?) (...) Sometimes there is more, a job with more terms, sometimes 4, 5 or 6, right? (...) But there is no maximum (I-UNESP)

There we have ... é ... Portuguese ... from 1 to 6. Minimum 1, maximum 6 (I-UNICAMP)

Although the evaluation of subject indexing is not yet implemented with users in the Library where the cataloger-indexing librarian at USP acts, as a member of the VOCAUSP Management Group she realizes that the request for new terms in certain areas of knowledge is not recurrent, with the subject indexing being maintained in a more general way, without much specificity in the terms. In order to solve this problem, training and area reviews have been carried out, in search of continuous improvement of the subject indexing process. At UNESP, the cataloger-indexer librarian is unaware of the carrying out of subject indexing evaluation tests in the Athena online catalog or in the institutional repository with users. At UNICAMP, although the cataloger-indexer librarian points out that doubts and suggestions for improvement from professionals or users are promptly answered, there is also no evaluation of subject indexing.

[...] we realize that there are areas that don't request much and this reflects in the register that we see that it is very general ... and we notice some mistakes ... Every time we notice this kind of thing we look for what to do, first do trainings, because in the vocabulary group we have a teacher who ... We also have a teacher who guides us, who is from ECA. (...) And we also do area reviews. (...) ... We also [created] a methodology to do area revisions. So every time we notice these deficiencies, not on the part of the user, in fact, we even make area revisions for ... and trainings for indexing, to see if it tries to improve the indexing (I-USP)

No ... I don't remember either having something like this happen, having individual, like, I believe not (I-UNESP)

[...] we don't do evaluation (I-UNICAMP)

With regard to possible improvements in relation to thematic representation, in the view of the cataloger-indexing librarian at USP at the moment there are no significant improvements around controlled vocabulary. However, she stresses the need for greater collaboration of librarians in the Libraries in updating the controlled vocabulary and standardizing remissive during the subject indexing process.

The UNESP cataloger-indexer librarian suggests that the UNESP Thesaurus be available to graduate students during the self-archiving of theses and dissertations in the institutional repository. He notes that, after the implementation of this process, there has been an increase in the number of assigned terms, favoring the professional who performs the validation of the thematic metadata in the online catalog Athena a broader view of the topics covered in the research, improving the compatibility of natural language with controlled language. Before the self-archiving, such situation was not possible, since there was a limit in the established quantity of terms in the cataloging request, which generated a certain dissatisfaction of graduate students, because on several occasions the suggested terms were not compatible with controlled language. The self-archiving allows more freedom in assigning the subjects of theses and dissertations, because the terms remain in the records of the institutional repository, while in the Athena online catalog there is validation and compatibility of terms according to a controlled language. By professional perception, currently the terms assigned in natural language are both general and very specific, not yet consolidated in literature.

[...] it's a recent thing, right? ...] It's a recent thing, but when it comes to filling out the metadata they have at the disposal of the thesaurus, maybe some kind of campaign (...) but that would be even from the sections of powders to make them aware of trying to make an effort to ... the use of the metadata, né ... the use of the thesaurus for ... for an improvement in this sense (...) although they are not needed many times we get ... to have a little better idea sometimes, nah, by the amount that it can do in this filling. (...) It's something that ... not always, not all cases are like this but many times it can happen (...) So we would arrive, elaborate, I would return to them (...) ... and they spoke, but these were not the key words that I put, the key words that I put, then I

explained look, it is not always possible to use this or that word for a matter of ... of the institution's policy (...) ... *sometimes they insisted that that word was very important (...)* Sometimes I even put it not being valid and then in cataloguing ... we used ... what was best for them (I-UNESP)

The establishment of an indexing policy at UNICAMP is relevant, a situation observed by the cataloger-indexing librarian. Although it is not yet formalized in an indexing manual, for the other informational resources the same guidelines are being adopted for theses and dissertations regarding the standardization of subjects. In relation to the new informational resources from the automatic collections, they are being standardized by the Directorate of Information Treatment (DTRI) of the UNICAMP Library System (SBU), while the retrospective informational resources verified by librarians of the Libraries.

[...] the ideal would be to have an indexing policy at UNICAMP, which we don't have yet. But what we are doing is using the same policy that we use for theses and dissertations throughout the repository (...) is what we are doing now. Only, what is the reality? It's something that will take us 100%. It's not something fast. And something that ... I ... we are doing on account of the repository: we are not letting up anything that is not really standardized in SophiA. (...) (...) This we're doing, this care we're taking, with all the subjects. (...) ... we are doing all the things; these things are being done at DTRI (I-UNICAMP)

In general, the results of the semi-structured interviews with managers and librarian cataloger-indexers of the institutional repositories of USP, UNESP and UNICAMP reveal a lack of consistency between the three universities regarding information policies and the thematic representation of the institutional repositories, being slightly more consistent between USP and UNESP and less consistent between them and UNICAMP (Chart 1):

Chart 1. Summary of interviews with cataloger-indexer managers and librarians at USP, UNESP and UNICAMP institutional repositories

Aspect	USP	UNESP	UNICAMP
Cataloguing theses and dissertations	Graduate students or Librarians in the Digital Library of Theses and Dissertations (BDTD) and imported to the online catalog DEDALUS	Graduate students in the institutional repository and imported into the Athena online catalog	Librarians cataloger-indexers in the Acervus online catalog and imported into the institutional repository
Type of language adopted for standardization of thesis and dissertation subjects	Controlled Language (USP Controlled Vocabulary - VOCAUSP)	Controlled language (Thesaurus UNESP)	Controlled language (mixed vocabulary), according to each field of knowledge: LC, BN, DeCS, FSTA Thesaurus, Ei Compendex, among others
Self-archiving of theses and dissertations	Graduate students at the Digital Library of Theses and Dissertations (BDTD)	Graduate students in the institutional repository	There are none
Cataloging of others informational resources	Institutional repository management team by automatic collection of PubMed, SciELO, Scopus and Web of Science databases. After that, the informational resources are sent to the	Institutional repository management team by automatic collection of PubMed, SciELO, Scopus and Web of Science databases	Institutional repository management team by automatic collection of PubMed, SciELO, Scopus and Web of Science databases. After that, the informational resources are sent to the

	Libraries for validation and insertion in the institutional repository		Libraries for validation, which return them validated to the management team for insertion
Self-archiving of other informational resources	Librarians	There are none	There are none
Type of language adopted to standardize the subjects of other informational resources	Controlled Language (USP Controlled Vocabulary - VOCAUSP)	Natural language in the institutional repository Controlled language in the Athena online catalog (Thesaurus UNESP)	Controlled language, composed of mixed vocabulary, according to each area of knowledge, being generally: LC, BN, DeCS, Ei Compendex, FSTA Thesaurus, among others
Controlled language management group	yes	yes	No
Formalized indexing policy/ Subject indexing manual	DEDALUS online catalog	Athena Online Catalog	No

Source: Own elaboration.

4 FINAL CONSIDERATIONS

The purpose of subject indexing in institutional repositories is to determine the thematic content of informational resources and express it in terms of indexing, allowing compatibility between informational resources and the information needs of the academic community. The diverse and growing complexity that permeates the representation and subject retrieval of informational resources, the university library should incorporate standards, tools and policies of organization, making the other actors involved in the process active partners in the growing interconnectivity of academic knowledge, aiming at proposing improvements in access, use and reuse of informational resources in institutional repositories.

In practice, the technical aspects surrounding institutional repositories need to be developed, managed and promoted by the librarian, involving not only the university library but the entire academic community. In this context, university libraries have the necessary experience for the establishment and management of institutional repositories, because more than being closer to the authors who produce the knowledge, they are institutions historically marked by the acquisition, treatment, dissemination and preservation of academic informational resources, which are increasingly relevant.

The research sought to conduct a diagnostic study of the situation of thematic representation in institutional repositories in the context of university libraries, making it possible to characterize, in a comparative manner, the organizational context of USP, UNESP and UNICAMP. Based on its results, some recommendations are outlined, which may be adopted in other institutional repositories that share the same characteristics:

- construction of a unique vocabulary at UNICAMP (UNICAMP Controlled Vocabulary) that encompasses the specificities of its areas of knowledge;

- elaboration and formalization at UNICAMP of an indexing policy in a manual that provides guidelines to professionals who perform the thematic representation of informational resources in the Acervus online catalog and institutional repository;
- implementation of self-archiving of theses, dissertations and other types of materials in UNICAMP's institutional repository;
- implementation of self-archiving of other types of materials at UNESP, with validation of thematic metadata by cataloger-indexer librarians;
- adoption of the Athena online catalog indexing policy at UNESP's institutional repository;
- formation of a Working Group (WG) of thematic representation (Indexing WG), by cataloger-indexer librarians and teaching specialists representing the areas of knowledge, with the establishment of partnerships with the VOCAUSP Management Group and the Working Group on Indexing Policy of the UNESP Library Network;
- formation of a Working Group (WG) of thematic representation (Indexing WG) at CRUESP, by managers and cataloger-indexer librarians representing the three universities;
- making available the controlled language used in the subject indexing of the informational resources to users in the pages of the institutional repositories;
- development of application profiles to generate reliable production indicators, meeting the demand of the three universities.

In this scenario, the role of the librarian in the creation and management of thematic metadata stands out, enabling adequate representation and retrieval of subjects in institutional repositories. However, such challenges require a professional performance consistent with this (new) role, based on better training and continued education both in aspects of creation and management of metadata of informational resources and in the theoretical-methodological foundations in indexing issues of Knowledge Organization.

Given the significant advance in the organization of informational resources in digital environment, in institutional repositories of university libraries the question of the description of thematic metadata remains open in the Knowledge Organization. Some points discussed here lead to other questions and demonstrate that the critical reflection of the process of indexing issues in institutional repositories in the context of university libraries is a relevant theme in the scope of Knowledge Organization.

REFERENCES

ALMEIDA, Maria Christina Barbosa de. **Planejamento de bibliotecas e serviços de informação**. 2. ed. rev. e ampl. Brasília: Briquet de Lemos, 2005.

BOCCATO, Vera Regina Casari. **Avaliação do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias**: um estudo sociocognitivo com protocolo verbal. 2009. 303 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2009.

CRUESP. Portal. Available at: <http://www.repositorio.unicamp.br>. Access on: 18 abr. 2019.

DAL'EVEDOVE, Paula Regina. **O tratamento temático da informação em abordagem sociocultural: diretrizes para definição de política de indexação em bibliotecas universitárias**. 2014. 268 p. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.

DAVID, Amos. *et al.* ISKO and Knowledge Organization's 25th anniversary: the future of Knowledge Organization and ISKO Panel Discussion. Reported by Rebecca Green. **Knowledge Organization**, v. 41, n. 4, p. 327-331, 2014.

FUJITA, Mariângela Spotti Lopes. Aspectos evolutivos das bibliotecas universitárias em ambiente digital na perspectiva da rede de bibliotecas da UNESP. **Informação & Sociedade: Estudos**, v. 15, n. 2, p. 97-112, 2005.

GNOLI, Claudio. *et al.* ISKO and Knowledge Organization's 25th anniversary: the future of Knowledge Organization and ISKO Panel Discussion. Reported by Rebecca Green. **Knowledge Organization**, v. 41, n. 4, p. 327-331, 2014.

GOMES, Fabio Andrade. **Padronização de metadados na representação da informação em repositórios institucionais de universidades federais brasileiras**. 2015. 277 f. Dissertação (Mestrado em Ciência da Informação) - Instituto de Ciência da Informação, Universidade Federal da Bahia, Salvador, 2015.

GONÇALVES, Marcos. Digital libraries. *In*: BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern information retrieval: the concepts and technology behind search**. 2nd ed. Harlow: Pearson, 2011.

GUIMARAES, José Augusto Chaves. Organização do conhecimento: passado, presente e futuro sob a perspectiva da ISKO. **Informação & Informação**, Londrina, v. 22, n. 2, p. 84-98, maio/ago., 2017.

HERNON, Peter; MATTHEWS, Joseph R. **Reflecting on the future of academic and public libraries**. Chicago: Ala Editions, 2013.

LEITE, Fernando César Lima. **Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto**. Brasília: IBICT, 2009.

MAI, Jens-Erik. Deconstructing the indexing process. **Advances in Librarianship**, v. 23, p. 269-298, 2000.

MAI, Jens-Erik. Folksonomies and the new order: authority in the digital disorder. **Knowledge Organization**, v. 38, n. 2, p. 114-122, 2011.

MORALES DEL CASTILLO, José Manuel. **Hacia la biblioteca digital semántica**. Ediciones Trea: Gijón, 2011.

PICKARD, Alison Jane. **Research methods in information**. 2nd ed. London: Facet Publishing, 2013.

SISTEMA DE BIBLIOTECAS DA UNICAMP (SBU). Repositório institucional. Available at: <http://www.sbu.unicamp.br/sbu/repositorio-institucional/>. Access on: 13 jul. 2020.

TARGINO, Maria das Graças; GARCIA, Joana Coeli Ribeiro; PAIVA, Maria José Rodrigues. Repositórios institucionais brasileiros: entre o sonho e a realidade. **Revista FSA**, Teresina, v. 11, n. 1, art. 6, p. 117-133, jan./mar. 2014.

TARTAROTTI, Roberta Cristina Dal'Evedove. **Avaliação do processo de indexação de assuntos em repositórios institucionais pela abordagem da recuperação da informação**. 2019. 370 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2019.

TARTAROTTI, Roberta Cristina Dal'Evedove; DAL'EVEDOVE, Paula Regina; FUJITA, Mariângela Spotti Lopes. Avaliação da consistência da indexação em bibliotecas universitárias federais da Região Nordeste do Brasil. **Anales de Documentación**, v. 20, n. 1. p. 1-19, 2017.

UNIVERSIDADE DE SÃO PAULO (USP). **Resolução nº 6444**, de 22 de outubro de 2012. Available on: <http://www.leginf.usp.br/?resolucao=resolucao-no-6444-de-22-de-outubro-de-2012>. Access at: 10 jul. 2020.

UNIVERSIDADE ESTADUAL PAULISTA (UNESP). **Regulamento Interno do Repositório Institucional**. Available at: <https://repositorio.unesp.br/handle/11449/144653>. Access on: 10 jul. 2020.



Article submitted to the similarity system

Submitted: 06/08/2020 – Accepted: 05/09/2020 – Published: 08/11/2020
