Open access institutional repositories in Latin America

Michelli Pereira da Costa
Fernando César Lima Leite

Universidade de Brasília – UnB, Brasil

ORIGINAL

Abstract

Objetive. This paper identifies and analyzes institutional repositories (IR) in Latin America. The development of systems by universities and research institutes in Latin America is highlighted in literature as a way to achieve the Open Access goals considering the public funds and also the IR characteristics.
Method. In terms of methodology, this paper used five selection methods and five IR analyses dimensions.
Results. The related elements brought the selection, mapping and characterization of 84 IR listed in a set of 289 digital repositories.
Conclusions. The research data showed that Latin America obtained relevant advances in terms of IR also if compared with United States and some European countries. This research shows that IR are efficient tools to promote the open access.

Keywords

Institutional repositories; Latin America; Open access; Scholarly communication.

Repositorios institucionales de acceso abierto en América Latina

Resumen

Objetivo. Este trabajo identifica y analiza los repositorios institucionales (RI) en América Latina. El desarrollo de sistemas por parte de universidades e institutos de investigación en América Latina se destaca en la literatura como una forma de alcanzar los objetivos de acceso abierto considerando los fondos públicos y también las características de los RI.
Método. En cuanto a la metodología, en este trabajo se utilizaron cinco métodos de selección y cinco dimensiones de análisis de RI.
Resultados. Los elementos relacionados permitieron la selección, mapeo y caracterización de 84 RI listados en un conjunto de 289 repositorios digitales.
Conclusiones. Los datos de la investigación mostraron que América Latina obtuvo avances relevantes en términos de RI también si se compara con Estados Unidos y algunos países europeos. Esta investigación muestra que los RI son herramientas eficaces para promover el acceso abierto.

Palabras clave

Acceso abierto; América Latina; Comunicación científica; Repositorios institucionales

1 Introduction

Institutional repositories (IR) are information systems comprising digital collections developed from management services concerning collection, organization, dissemination, and preservation of research outputs from members of an institution (COSTA, 2014). This type of information system has been widely cited in literature regarding scholarly communication given its importance as a means towards the Open Access. This movement seeks to promote initiatives for the distribution of scientific literature on the Internet with no access costs and with as little restriction as possible. Particularly noteworthy is the need to make available in open access scientific literature resulting from researched funded by public resources.

The first actions towards making scientific literature available in open access date from the early 1990s, with the development of a digital repository currently known as ArXiv (RAMLO, 2007). Ten years later, researchers from various institutions consolidated the Open Access with declaration entitled Budapest Open Access Initiative

(SARMENTO et al., 2005). This declaration included proposals for two action strategies: one of them refers to the open access journals, and the other regards archiving journal articles in digital repositories, be them in IR or subject ones.

Facing the context of scholarly communication, IR have been considered strategic to fulfill the objectives of open access. They are also regarded as a valuable tool to manage research outputs in an institution. Therefore, given the importance of this type of information system, both for open access and for the institutions themselves, this study aimed to identify and describe the characteristics of institutional repositories in Latin America and their contributions for the development of open access in this region.

It is important to explain that the study of characteristics and contributions from institutional repositories for the development of access in Latin America is justified by at least two aspects. The first aspect is the near exclusivity of public funds for the researches developed in Latin America (UNESCO, 2010); this condition makes the role of institutional repositories more relevant for the promotion of free circulation of scientific information. The second reason regards the absence of studies aiming to portray and analyze the development of IR in Latin America focusing on their contributions for the advance of open access.

2 Metodology

This research is of descriptive nature and used the strategy of mixed methods. According to Creswell (2010), this strategy is the convergence of qualitative and quantitative approaches. The convergence stems from the acknowledgement that all methods have limitations, but the triangulation of methods may overcome some of them. The strategy of mixed methods consists in triangulating quantitative and qualitative data sources, seeking an integration in the analysis of the results. Therefore, the study was developed in two phases. The first phase consisted in identifying information systems which can effectively be considered IR in the ambit of Latin America. This is because not all open access digital repositories constitute institutional repositories, even if they call themselves so. The second phase consisted in characterizing the selected systems. Even though there were data with different natures, at times quantitative, at other times qualitative, both phases adopted the data collection as a method and the documental research as a technique for this method.

The methodology for selection and characterization of the IR followed the proposal by Costa & Leite (2015). In their work, these authors analyzed the most cited scientific literature on the theme of IR and highlighted the elements defining an IR and the elements commonly used to characterize it. For the identification of the IR, five criteria were proposed for application: i) being institutionally defined; ii) dealing with the academic production of the institution; iii) making available the complete texts of the documents; iv) being an open access initiative; and v) being interoperable within the context of open access.

In the second stage, the selected repositories were characterized according with five analysis dimensions (Table 1). The dimensions were based on the reports from researches of the same nature presented by Lynch & Lippincott (2005), Westrienen & Lynch (2005), and Rieh et al (2007).

Table 1: IR analysis dimensions and characterization elements

ANALYSIS DIMENSIONS	CHARACTERIZATION ELEMENTS
1. Size	Total quantity of documents Quantity of documents of academic production
2. Document types	Types of document presented in the report of RCAAP validator and IR communities and collection
3. Areas of knowledge	Sciences Social Sciences and Human Sciences Arts and Humanities
4. Policies	Working policy Institutional policy Preservation policy Copyright policy
5. Technologies	Software the IR uses

Source: Costa & Leite (2015), with adaptations.

The data were collected in the two research phases using the documental research method. According to Franco (2003), this sort of investigation strategy is based on the analysis of documents, seeking to identify the necessary information in documents rendered relevant for the research issues. In this study, the webpages of the repositories were considered, as well as the reports generated by the database validators, as source-documents of the necessary data. The research used two validators: RCAAP[1] and OAI-PMH Validator & data extractor Tool[2]. Both are OAI database interface analysis systems, which aim to guarantee the interoperability between the systems and the quality of the data they present.

3 Identification of IR in Latin America

In order to identify IR in Latin America, it was necessary to collect all systems in this area registered as digital repositories in seven information sources concerning this theme[3]. From the seven information sources, two are of international range, three are regional and two are nationwide. The use of these sources resulted in the identification of 298 information systems classified as open access digital repositories. The identified systems come from institutions from 18 countries in the region (Figure 1). Of the group, noteworthy are the participation of Brazil, Mexico, Argentina, Colombia, Ecuador, Venezuela, Peru, and Chile, which together represent 89% of the total. The remaining ten countries presented results lower than ten. In three countries — Honduras, Paraguay, and Trinidad and Tobago — only one system per country were found.
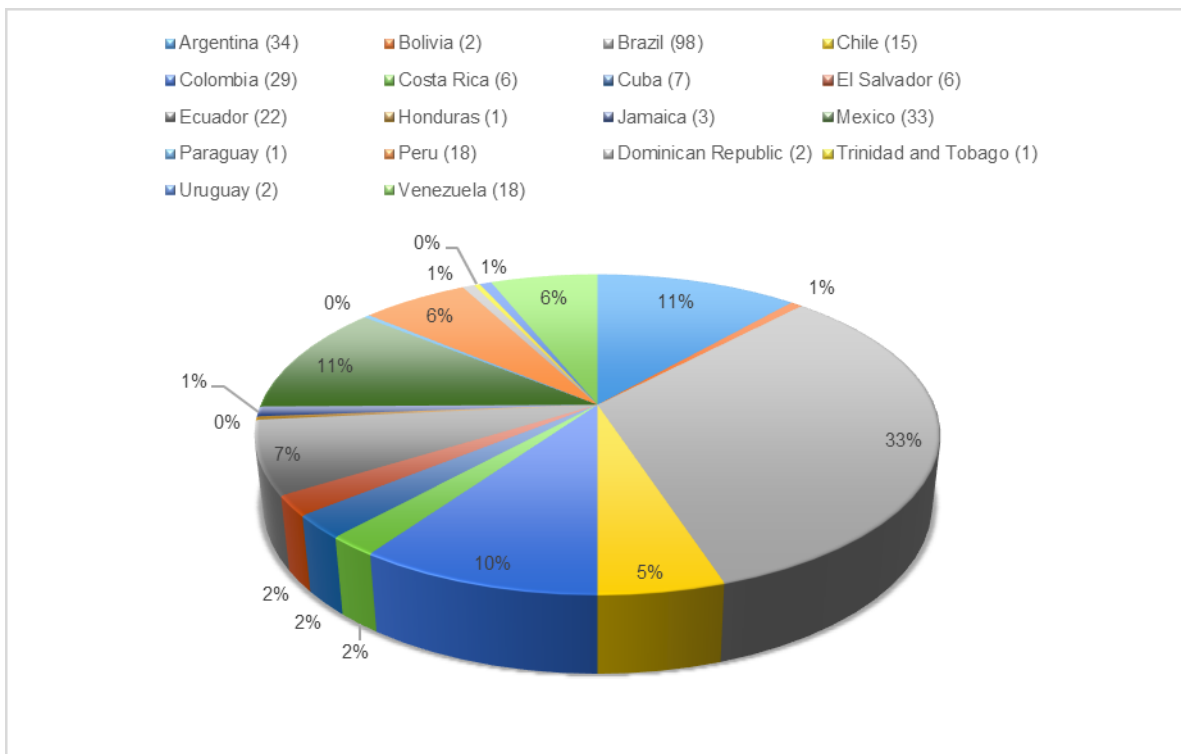

Figure 1: Digital repositories identified per country in Latin America

Source: Elaborated by the author of this study.

Of the 24 countries initially researched, six presented no result for the search strategy in the aforementioned sources. The countries with no participation in the universe of digital repositories were: Barbados, Guatemala, Guyana, Haiti, Nicaragua, and Panama. Given the lack of digital repository data in these countries, this study considered there are no IR in any of their education and research institutions. Thus, the five criteria for selection were applied to the 298 digital repositories.

For the application of the first selection criterion - institutional orientation, a verification was made to check if the repository was maintained by an institution and if it aimed to store the academic production of its members. The sources used for the analyses were the presentation texts of the repositories themselves and other documents

available in the system itself. As a result, 172 repositories classified with institutional orientation were identified (Figure 2).
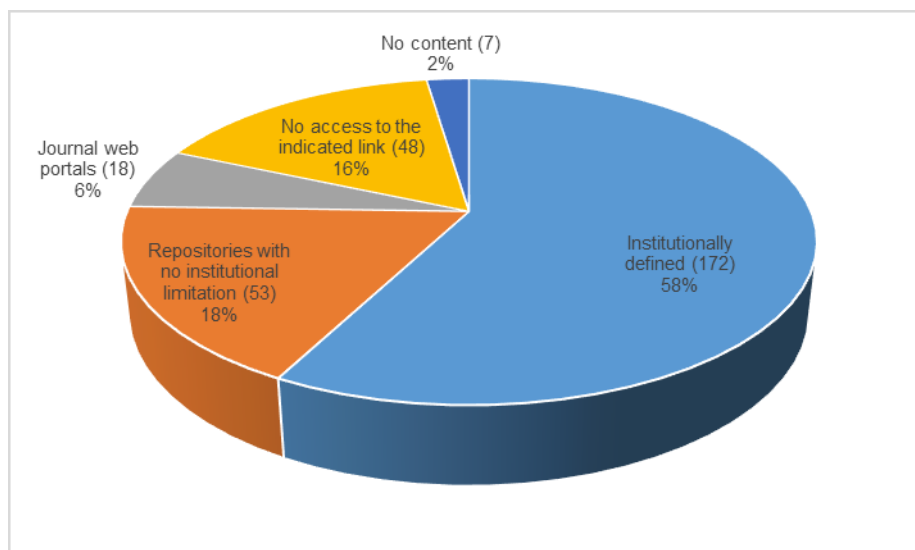

Figure 2: Institutional orientation of the DRs

Source: Elaborated by the author of this study.

In the group of digital repositories with no institutional orientation, groups were formed of systems with characteristics of subject repositories, digital libraries, and portals. Furthermore, the classification was made in this category for seven repositories that were not analyzed due to the absence of available documents in the system, and 48 repositories that did not present access to their webpages, caused by failure to load the webpage or an internal system error. According to the results, only 58% of the systems registered in international directories and national lists of digital repositories presented an institutional orientation in their digital collections. The other 42% were excluded of the following steps in the research.

The second selection criterion regards the type of content stored in the 172 repositories. As previously justified, the IR are a strategy proposed by the Open Access Movement for the dissemination of scientific articles on the Internet. Therefore, it was here considered that, in the ample group of types of documents that can be stored in the systems, the presence of scientific articles is mandatory.

In order to analyze the types of documents found in the IR, three strategies were used. The first was the analysis of the texts of presentation and description of the systems, so as to verify the purpose of the repositories and their relation with academic publications. The second strategy was the analysis of the communities, collections, and search filter of the repository, when applicable. Such strategy allowed the identification of the manner in which the contents were described in the systems, including the documental typology. In the cases when these elements did not offer sufficient data, a third strategy was adopted: the use of a report generated by RCAAP Validator.

The analysis of the types of documents allowed the identification of 89 repositories fulfilling the requirement of storing their institution's member's scientific articles (Figure 3). These repositories store, preponderantly, scientific articles, papers presented at events, and theses and dissertations. The other systems presented certain types of documents that can be characterized as scientific production; nevertheless, given the absence of scientific articles, they were removed from the analysis in the following steps.
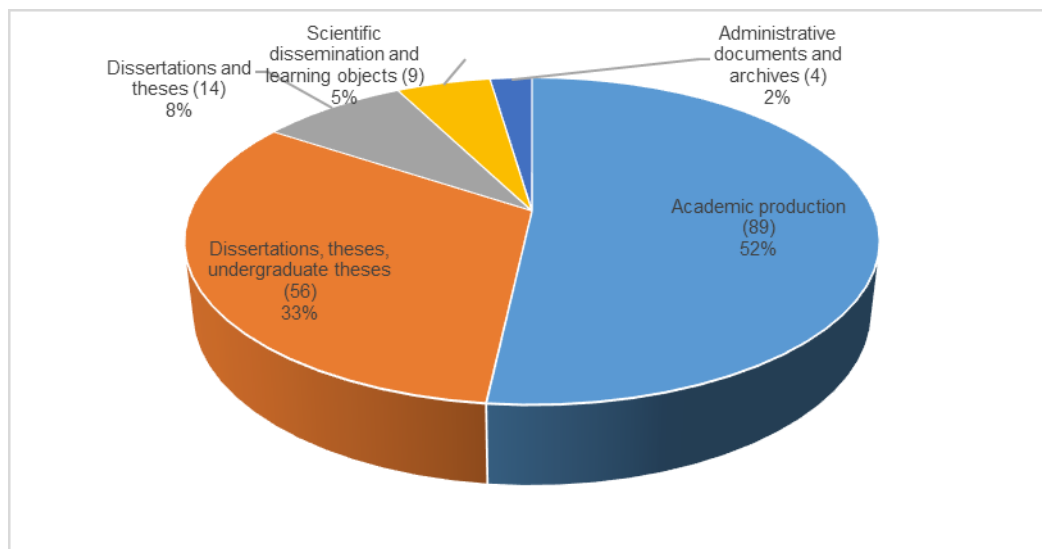
Figure 3: Range of document types from institutionally defined repositories

Source: Elaborated by the author of this study.

The third selection criterion applied to the 90 repositories was the availability of complete texts. Attempts were hence made of accessing the document in some repositories. Only five systems presented no access for any of the files attached to the descriptive register and, therefore, were classified as referential databases. Of the total repositories at this stage, 84 were considered systems making the full texts available. However, not all registers presented the document files or were accessible, as will be discussed in the fourth criterion.

The fourth selection criterion concerned the availability of the complete texts in open access. For this, the information generated by the RCAAP Validator report was checked. An assessment was made, in the reports, among other aspects, on the percentage of documents in restricted access in the repository. This information was used to calculate the percentage of documents in open access. Up to the end of January 2014, 48 report requests received a response. From the data collected, the following results were obtained:

- 6 repositories make from 1% to 50% of their registers available in open access;

- 26 repositories make from 51% to 75% of their registers available in open access;

- 17 repositories make from 76% to 100% of their registers available in open access;

- 45 repositories obtained no response from the Validator.

The last criterion of selection dealt with the interoperability of the repositories. Within the context of open access, it is defined by means of the use of Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH Protocol). Therefore, in order to assess under this criterion, the URLs of the OAI interface of the repositories were collected and had their validity tested. The OAI URL of the repositories was verified with the RCAAP Validator and OAI-PMH Validator & data extractor Tool[4]. From the two strategies, it was possible to identify that 71 (85%) of the repositories OAI URLs, the necessary condition for interoperability in the context of open access. Nevertheless, it is not possible to affirm categorically that the other 13 (15%) repositories are not interoperable, nor that they do not possess an OAI URL. This is because the OAI URL can be configured in the system with a value different from what this research tested. Considering the exposed problematic, this criterion was not treated as eliminatory and its results were used only to illustrate the current situation of the repositories in what concerns an essential characteristic for this type of system.

An application of the selection criteria resulted in the identification of the group of repositories fulfilling essential characteristics the most cited scientific literature on this theme indicated. The selection of the IR started with the collection of the 298 digital repositories. The performance of the five states of selection, corresponding to the criteria, resulted in the identification of a group of 84 IR (Figure 4).
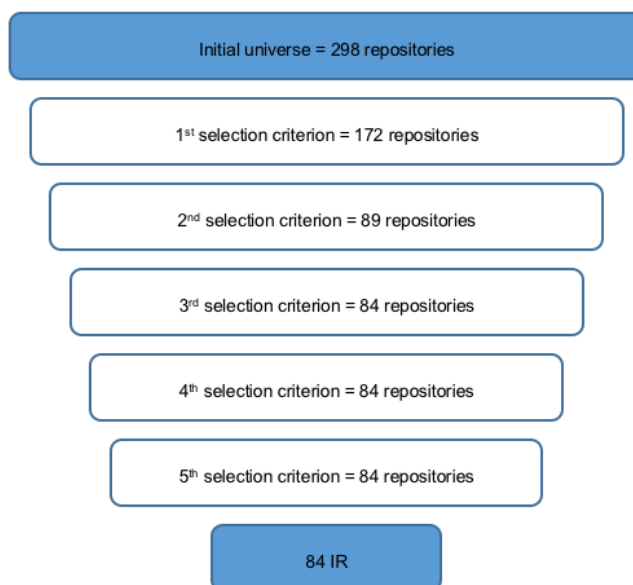
Figure 1: IR selection stages

Source: Elaborated by the author of this study.

4 Characterization of the IR of Latin America

The IR of Latin America this research mapped were described from five analysis dimensions: repository size, type of documents, areas of knowledge, policies, and technologies of the repositories.

4.1. Size of the IR

The first dimension analyzed about the IR was their size, measured from the volume of documents deposited. The verification of the total quantity of documents was calculated from the information offered by the repositories and by the RCAAP Validator report. As a result, 588 thousand documents were identified, deposited in the selected 84 IR. In general, Brazil was the country with the largest quantity of documents, with over 267 thousand, corresponding to nearly half (46%) of all the documents deposited in IR in the region. After Brazil, Colombia (17%) stands out, then Argentina (14%) and Mexico (9%). These four countries together hold 85% of the documents found in the IR of Latin America.

In addition to be the country with the largest total quantity of documents deposited in its IR, Brazil also holds three of the five IR with the largest quantity of documents. Among the institutions of the region at the top of the list are: UFRGS (Br), Universidad Icesi (Col), Embrapa (Br), USP (Br), and Universidad Nacional de La Plata (Ar). Of these five, three are Brazilian, one is Colombian, and one Argentinian. Yet, although Brazil is the country with the greatest quantity of documents deposited in the IR, the average documents per repository is not as high as some other countries in the region. In order to calculate the average, the total value of documents per country was divided by quantity of IR in each country. This, the country identified with the largest quantity of documents in average is Ecuador (12 552 documents), followed by Colombia (12 393 documents) and Chile (8 028 documents). Brazil comes fourth, with an average of 7 875 documents per repository, i.e. in proportional numbers, Brazil occupies the fourth place in quantity of documents in its IR.

These results indicate that, although Brazil, Argentina, and Mexico hold a significantly higher quantity of IR operating, in average they are not as populated as those in certain countries, such as Ecuador, for instance. Furthermore, according to Erber (2000), the three countries concentrated about 80% of scientific publications and patents of all the region. However, the data show they are not among the first with highest average of documents per repository. This datum reveals an indication that the presence of scientific publications is still low in the IR of some of the institutions in Latin America.

According to the data collected by Rieh et al. (2007), in the USA, the average of documents among the IR implemented in the country was of 3,200 documents, while in Latin America, the current average is 7,000 documents deposited. Of the total of IR investigated by Rieh et al. (2007), only 19.4% held more than 5 thousand documents. The authors consider these rates low, but stated these results are in accordance with those of previous researchers, such as Lynch & Lippincott (2005), Shearer (2004), and Bailey Jr. et al. (2006). The disparity between the reality found in this study and the reports of the North American reality should be taken with relativity, given the time distance between the two analyses.

Populating is one of the aspects analyzed on the development of IR, as pointed in the study by Lynch & Lippincott (2005). According to these authors, this aspect indicates the success and challenges for the development of the systems. Nevertheless, it is important to highlight that many of the IR analyzed hold more types of documents than what is considered by this research as academic production. Considering this, it was necessary to investigate the types of documents comprising the collection of the IR of the region.

4.2. Types of documents present in the IR

For the investigation of this aspect, data generated by the search filters of each of the systems and the information provided by the RCAAP Validator reports were collected and analyzed. Hence, it was possible to identify that the type of document with greatest presence in the IR are journal articles (36%). In second place are dissertations, theses, and undergraduate thesis (24%). At the same rate (24%) are administrative documents, archives, and other types of documents, not categorized as research output. In fourth position come works presented at academic events (9%).

In the sequence, at 5%, audiovisual documents, learning objects, documents of science dissemination, technical reports, and software were grouped. These documents, despite not being consensually classified as scientific documents, were considered within the scope of the definition given their academic nature. Finally, representing only 2% of the documents, books and chapters from books (Figure 5).
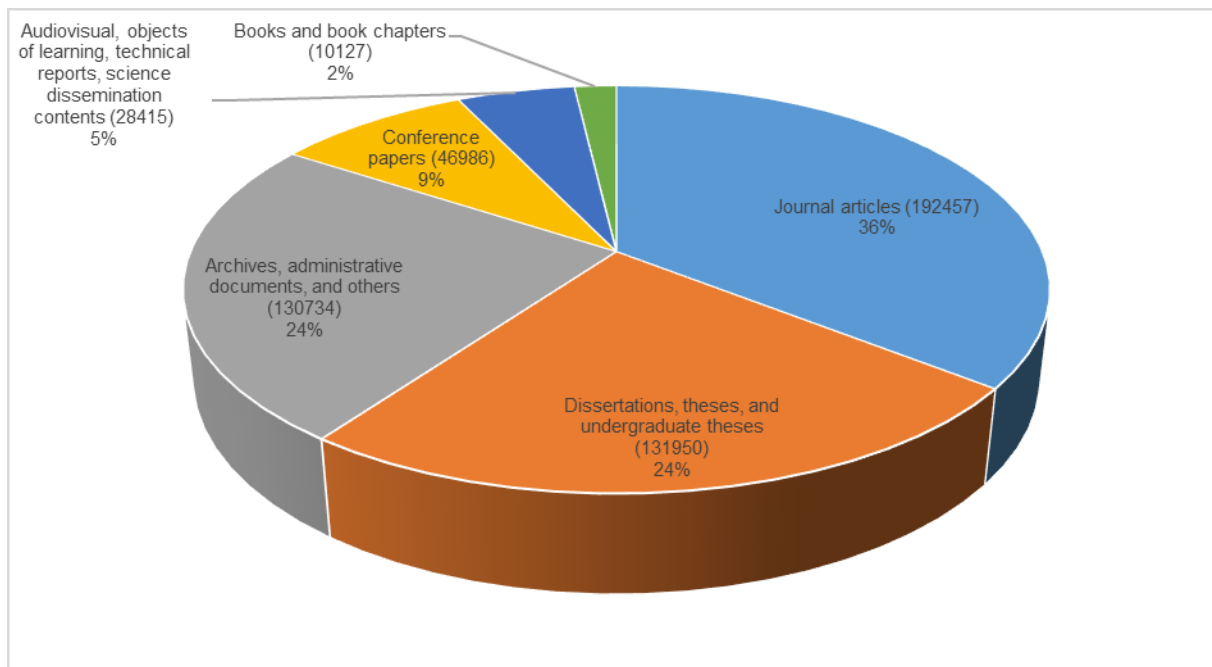

Figure 5: Types of documents found in the IR of Latin America

Source: Elaborated by the author of this study.

The predominant presence of journal articles in the IR can be explained as stemming from the conditioning adopted by this research, in which the selected repository should necessarily present journal articles. This condition was established as the research scope are the IR within the context of open access and scholarly

communication. This implies that both the origin and the aims of the repository should be linked to scientific information, which has in journal articles one of its main manifestations. Hence, considering the limitations stated, the data point at a positive result in what concerns the aims of the IR of the region regarding open access.

Dissertations, theses and undergraduate theses also present significant presence among the types of documents. In addition to this result in the IR, they had great occurrence among the repositories excluded from the analysis. In the selection stage, from the total of 172 systems, 41% were predominantly dissertations, theses, and undergraduate theses, and were excluded for not presenting journal articles. However, when adding the two groups, the results show that dissertations, theses, and undergraduate theses are the majority among the types of document in digital repositories of the region.

According to Bailey Jr. et al. (2006), dissertations and theses are the most common type of document among the 87 IR of the Association of research Libraries (ARL), comprehending university libraries of Canada and the USA. For the authors, the results are explained by the facility of recruiting such types of documents in relation to other types. In the context of Latin America, some elements can be added to explain this phenomenon: for instance, the policy of one of the main research funding agencies in Brazil, Coordination for the Improvement of Higher Education Personnel (CAPES). This agency published an Order in 2006 establishing that all dissertations and theses of Brazilian postgraduate programs it acknowledges bust be made available and possible to be accessed on the Internet (BRASIL, 2006).

 A new factor observed in the region is the presence of undergraduate theses. In none of the data collected on the development of IR in other countries was reported the existence of this type of document. The presence of undergraduate theses was more significantly found in five countries: Brazil, Chile, Colombia, Ecuador, and Peru. Among them, Chile, Colombia, and Peru were part of the initiative called Cybertesis[5]. This initiative is funded by UNESCO and counts on the partnership with several universities in Africa, Europe, and Latin American, and aims to store and retrieve dissertations, theses, and undergraduate theses, by means of the OAI-PMH protocol.

Very close to the quantity of dissertations, theses, and undergraduate theses are administrative documents. Both types represent each 24% of the total documents. This type of document makes no part of scholarly communication context and, hence, is set apart from IR assumptions for the aims of open access. The strong presence of administrative documents in the IR indicate they have also been used as solutions for managing information of another nature, or it may not have been appropriate for the purposes of open access scholarly communication systems. Westrienen & Lynch (2005) collected data on IR in 13 countries, and no category specific for administration documents was used, but an "others" category was applied to classify documents not fitting any type of academic production. Two countries stood out in this category: The Netherlands and Germany. In the Netherlands, 40% of the documents found in the repositories were classified in the "others" category; in Germany the number drops to 25%. In the data collected by Rieh et al. (2007) in the USA, in general, 20.5% of the content of the repositories are administrative documents. The results show the practice of storing administrative documents in the IR is not an isolated fact in Latin American countries, but certainly need to have their impacts assessed.

In a smaller percentage are books and book chapters, with only 2% of the universe. In the data collected by Westrienen & Lynch (2005), books represented only 1% of the documents found in the IR of Australia. One of the elements that can help explain this phenomenon, which is not exclusive to Latin America, is the participation of commercial publishers in the production and distribution of books. Furthermore, another explanation can be the low reach of open access as an alternative/complementary model of scholarly communication in certain areas, especially Social Sciences and Arts and Humanities. According to Brockman et al. (2001), the importance of books in the scholarly communication processes in these areas is still strong.

From the classification of the types of documents, it was possible to determine the presence of research outputs in the IR. This was made by subtracting from the total of documents the quantity representing the category of "administrative documents, archives, and others." As a result, approximately 409,000 documents were identified among journal articles, dissertations and theses, papers presented at academic events, books and book chapters. This value corresponds to 69.5% of the total documents stored, i.e. about 30% of the documents stored in the IR in the region dot not correspond to research outputs of their institutions.

At the end of this filtration, Brazil continued to concentrate the greatest quantity of academic production documents in its IR, given the quantity of systems in the country. Along with Brazil, Argentina, Mexico, and Chile gathered 86% of the total academic production documents (Figure 6).
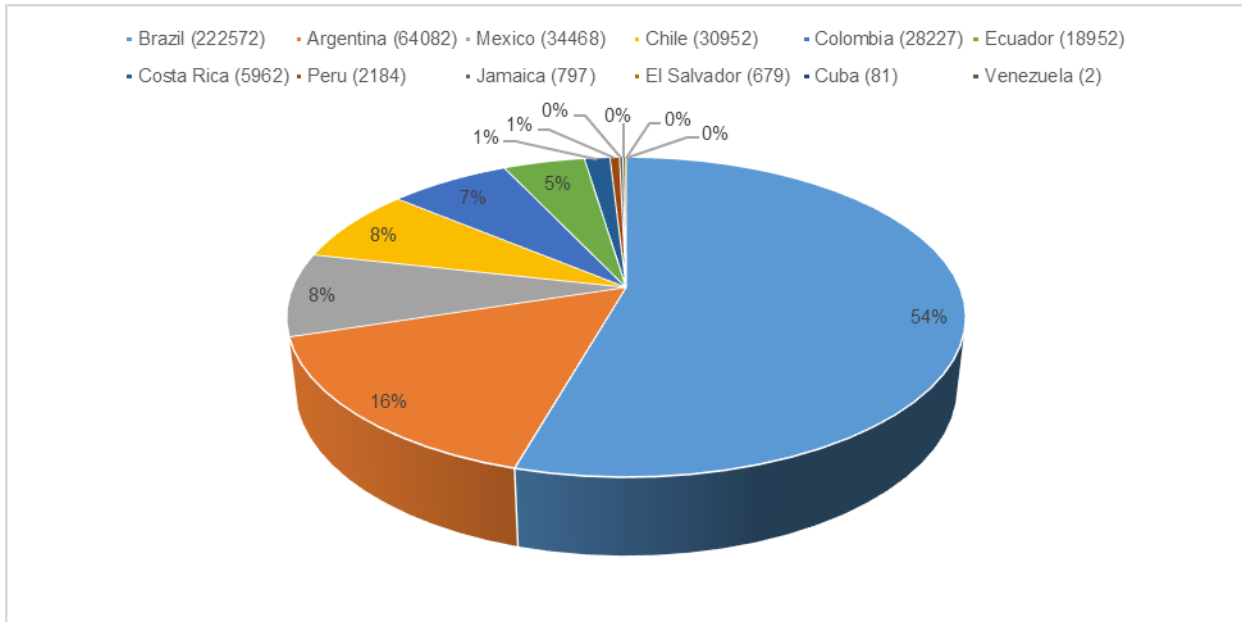

Brazil (222572) · Argentina (64082) · Mexico (34468) · Chile (30952) · Colombia (28227) · Ecuador (18952)
Costa Rica (5962) · Peru (2184) · Jamaica (797) · El Salvador (679) · Cuba (81) · Venezuela (2)

Figure 6: Academic production documents found in IR of Latin America

Source: Elaborated by the author of this study.

Among the countries with the greatest quantity of total documents, Colombia held second place; however, when only the academic production was measured, that country dropped to fifth place and Chile entered the group of the four highest countries. This indicates Colombian IR make a large quantity of documents available, but few represent the academic production of its institutions. The reduction of Colombian documents was in the order of 71.5%, the greatest of all countries. In addition to Colombia, only Costa Rica had a reduction of more than 50%. The country with the smallest reduction was Chile, with only 3.6%, which indicates most documents stored in Chilean IR are of scholarly nature.

4.3. Areas of knowledge

The investigation of the presence of the areas of knowledge in the IR made it possible to generate indications and to perceive tendencies in the areas in what concerns its participation in the open access. This was obtained using only the information available in the communities and collections, whenever it was possible to identify them. Given the different forms of organization of the contents in the systems analyzed, it was not possible to calculate the quantity of documents per area of knowledge in about 40% of the repositories.

 The categories used for the classification of the areas of knowledge were the three Costa & Meadows (2002) and Gumieiro (2009) adopted. According to these authors, the choice of the categories is justified by its simplicity and broad use in English language works. One of the categories used was "Natural and Exact Sciences", which gathered works from Exact and Technological Sciences, Biological Sciences, and Health Sciences. This category gathered 72% of the total documents analyzed. The second category gathering the most documents was "Social and Human Sciences", representing 23%. This area included in its group the disciplines of Applied Social Sciences and Human Sciences. Finally, with only 5% of the documents is the area of "Arts and Humanities". It includes the works from the disciplines of Linguistics, Letters and Arts[6]. The strong presence of documents in the area of Exact and Natural Sciences was not the same in all countries of the regions. It was preponderant in the IR in Argentina, Ecuador, Brazil, Colombia, and Chile. In Mexico and Jamaica, documents from Social and Human Sciences prevailed. Arts and Humanities were not predominant in any of the countries, but they appear with more relevance in the IR from Chile, Brazil, and Argentina.

Given the low presence of IR this research identified in some countries, associated to the difficulty of classification of the documents in subject areas, six countries did not have any of their documents classified according to their themes: Costa Rica, Cuba, El Salvador, Peru, Trinidad and Tobago, and Venezuela. The sum of the documents from these six countries totalizes less than 10% of the total; therefore, despite the absence of data on them, it is possible to perceive the predominance of documents in the areas of Exact and Natural Sciences made available in open access in Latin America.

The wide presence of documents of the areas of Exact and Natural Sciences in Latin America can be compared to the experience in the United Kingdom, according to data from Westrienen & Lynch (2005). These authors found two thirds of the documents of the region belonged to the area of Natural Sciences and Engineering. In Australia and Italy, the results indicated the IR focus on the contents of Social and Human Sciences. The other countries they investigated presented a certain uniformity among the areas or were not classified. Some factors regarding disciplinary differences in the standards of scholarly communication may contribute towards the explanation of the predominance of Exact and Natural Sciences contents in the IR of the region. Disciplines in this area disseminate the results of their researches by means of journal articles, as found in studies by Gorraiz et al. (2009), Houghton et al. (2003), Huang & Chang (2008), and Leite (2011). Although open access brings with its potential benefits for all areas, operating it in the context of disciplines that prioritize books as the primordial vehicle for publication is something more complex. It is probable that this occurs due to peculiarities involving publishing processes of the production of books. Facing this, it seems easier to promote open access to journal articles than to books, which possibly explains the predominance of the first in the IR.

4.4. IR policies

The fourth analysis dimension dealt with the availability of policies on the IR. At this stage, the policies were comprehended as documents regulating various aspects regarding the existence and the workings of the repositories. The intention was to verify the existence (or inexistence) of policies available on the webpage of the repositories. The content of these documents was not analyzed in the work. The existence of four types of policies was checked in each system: working policies; policies of information on the institution dealing with the IR; content preservation policies; copyright policies.

The working policies placed in the IR dealt with aspects regarding the operation of the systems. As a rule, they declared the aim of the system and explained the processes for the deposit, validation, access, and use of the documents. In these conditions, 31 policies were identified in the selected 84 IR. Brazil was the country presenting the largest quantity of working policies, followed by Argentina and Mexico. In Chile and Colombia, two repositories were found with this type of policy. In El Salvador, Peru, and Trinidad and Tobago, the quantity dropped to only one per country. In Costa Rica, Cuba, Ecuador, Jamaica, and Venezuela, no repositories were located making their working policies available.

According to the data collected by Bailey Jr. et al. (2006), 65% of North American IR indicated to have their working policies, and about 50% of them presented this policy for revision by an institutional authority, most of them to the University Council. Brazil presented a significant quantity of working policies, certainly due to the action of the Brazilian Institute of Information in Science and Technology (IBICT). As it induces and funds the construction of a large part of Brazilian IR, IBICT guides the elaboration of such sort of policy and its respective availability to the users of the systems (COSTA, 2008). The practice observed among the IR indicates the working policy can be a good tool to validate the repository and its processes, along with the high administrative instances and policies of the institution. Furthermore, the policies are certainly planning mechanisms and can be used as parameters to assess the performance of the systems.

In addition to working policies, the existence of information institutional policies relating with the IR was investigated. Documents presenting guidance and/or determinations on academic production of the members of a given institution were classified as institutional policies. As a rule, documents are signed by the summit directing the institution, such as rectors and directors. However, in order to be considered in this research, documents must necessarily deal with the repository and be available in their webpage. In these conditions, only eleven policies were located, nine of them in Brazilian repositories and two in Argentinian ones.

Institutional policies are important documents of formalization of the commitment established with the repository by the maintaining institution. According to Abadal et al. (2009), this type of policy aims to boost the open access

in the ambit of scholarly communication and, hence, boost initiatives in this context. According to the authors, the policy should determine or guide the researchers to deposit their works in IR or subject repositories and stimulate the results from their researches to be published in open access journals. The contents of the 11 identified policies were not analyzed, so it is not possible to state if they fulfill such aspects Abadal et al. (2009) pointed. However, it is possible to notice a low presence of legal mechanisms to formalize a commitment of the institution with the development of the repository.

Preservation policies stood out in the research, as it is an essential aspect in the constitution of a repository. They represent the planning and the commitment the repository assumed with the preservation guarantee of the documents therein deposited throughout the years. Despite its importance, this was the least present type of policy among the selected repositories. In none of them a specific policy dealing only of preservation was located. In all cases, preservation elements were dealt in the working policies and were limited to the determination of the types of document that should be deposited for their preservation to be guaranteed. Even in these conditions, only eight were found in Brazil, one in Argentina, and one in Mexico. The low presence of this type of policy was also found by Rieh et al. (2007). According to these authors, only 17% of North American managers specified in a document the types of files the repository takes and preserves. Most include the preservation for the most common types of documents, such as PDF, XML, AIFF, GIF, JPEG, and TIFF.

The last type of policy investigated concerns copyright. Documents regarding legal conditions for the deposit, access, and use of the available material were classified in this category. As in preservation policies, copyright policy elements were presented in the working policies. Some exceptions were observed from the availability of non-exclusive distribution terms or similar documents in the repository webpage. As a rule, in the terms, the IR receive the rights to make available certain documents freely and with no access restriction. Among the selected repositories, 32 policies were found, most from Brazilian systems. In addition to these, there were also six from Argentina, four from Mexico, two from Chile, and two from Colombia. El Salvador, Peru, and Trinidad and Tobago had only one policy each. In Costa Rica, Cuba, Ecuador, Jamaica, and Venezuela, as in the other types, no IR copyright policies were found.

The copyright policy in the IR, in addition to be a management mechanism within the system, should establish licenses and conditions for the access and use of the documents therein deposited. As BOAI presented, licenses are a necessary condition for the availability of documents in open access. In the context of Latin America, it was found that 62% of the repositories do not have a document available in the very system to formalize the conditions of access to the documents deposited, which can be a problem for the development of open access itself in the region.

In general, the presence of policies on the IR was considered low. In none of the types of policies did the presence of the documents reach 40% of the repositories. In only 15 repositories the existence of policies was considered satisfactory and they were all from institutions from Brazil, Argentina, and Mexico.

4.5. Technologies of the IR

In order to analyze the technology of the repositories, the software used to build and manage the system was investigated. In general, three software were identified: DSpace, Eprints, and Greenstone. In seven systems (8%) it was not possible to identify the software used. DSpace is a free software for the management of digital repositories developed by Massachusetts Institute of Technology (MIT) in 2002. According to the results obtained in the research, this program is the most widely used by IR in the region. It is employed by 68 analyzed systems, representing 82% of the total. In second place, with only seven IR (9%) came Eprints. Like DSpace, it is a free software for the same finality, developed by the University of Southampton. Greenstone was developed by the University of Waikato with the cooperation of UNESCO. It is used by only one repository among the selected ones in the region (less than 1%). The predominance of DSpace was also observed among the IR in North America in the study by Bailey Jr. et al. (2006). According to the authors, 70% of the investigated repositories used DSpace. In the indicated context, only one repository uses Greenstone and no use of Eprints was indicated in their research. The results in Baily Jr. et al (2006) concerning North American IR are quite different from those presented by Westrienen & Lynch (2005) in relation to the IR in Europe, where Eprints is used by 44 systems, while DSpace is used by only 25. However, the predominance of DSpace at a global level is confirmed by the data available on ROAR. In that directory, of the approximately 3,000 digital repositories registered, nearly half

use DSpace. Statistics on OpenDOAR, likewise, indicate greater adoption of DSpace, corresponding to 41.5% of the total of 2,577 digital repositories in all the world.

5 Conclusions

All selection criteria used in the research were pointed by the most cited scientific literature on the theme with essential elements to IR. Yet, after applying those criteria to the systems identified in directories and lists of open access repositories, only 27% of the total of 298 digital repositories remained. The portrayal herein discussed points to the existence of a relatively large distance between what literature indicates as institutional repository and the practice of developing such systems in Latin America.

The exclusion of part of the digital repositories of the group of open access IR was justified by the very aims of open access, which seek to promote the circulation of journal articles freely through the Internet. Therefore, the presence of journal articles in the IR was considered a mandatory criterion when selecting the systems. The application of this criterion brought about a reduction to nearly half the systems of the analyzed group. With this, a tendency to create digital repositories in the region is noticeable, but these systems do not store and make available journal articles and, hence, were not considered as open access IR.

Among the systems selected as open access IR, journal articles were the documents with the largest representation in the group of available documents. However, other types of research output were also identified as relevant for the constitution of the collections, especially dissertations, theses, and undergraduate theses. In addition, administrative documents made part of the documents the IR in the region made available. Such results indicate IR in Latin America have also been used to manage types of documents not foreseen or expected in their scope of action. The presence of administrative documents represents, at the same time, a deviation in the action of the IR in the context of open access and of scholarly communication, and the adoption of IR to deal with information management issues of another nature.

In general, this study presented a wide panorama of the development of IR, aiming towards the objectives of open access. Among the characteristics promoting open access stand out the quantity and the nature of the documents the IR store. Among the deficiencies of the systems are noteworthy the little formalization of the institutional commitment and the low presence of policies. Certainly, the results presented do not deplete the analysis of IR as promoters of open access in Latin America and should be discussed in the view of new perspectives and other aspects not dealt with in this research, such as the perspective of researchers, institutions, and publishers.

References

ABADAL, E.; MELERO, R. ABAD, F.; VILLARROYA, A. [2009] Políticas institucionales para el fomento del acceso abierto: tipología y buenas práticas. Bollettino AIB, 2009, vol. 49, n. 2, p. 159-170. Retrieved from: <http://eprints.rclis.org/13565/>. Access: Feb 2013.

BAILEY JR., C.; Institutional Repositories. [2006] SPEC Kit, n. 292. 2006. Retrieved from: <http://publications.arl.org/Institutional-Repositories-SPEC-Kit-292/>. Access: Feb 2013.

BRASIL. [2006] Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos. Portaria nº 013, de 15 de fevereiro de 2006. CAPES: Brasília. 2006. Retrieved from: <https://www.capes.gov.br/images/stories/download/legislacao/Portaria_013_2006.pdf> Access : Feb. 2014.

BROCKMAN, W. S. et al. [2001] Scholarly work in the humanities and the evolving information environment. Washington, DC: Council on Library and Information Resources. 38p.

CARVALHO, J.; RODRIGUES, E. [2012] Condições de agregação de recursos no portal RCAAP. RCAAP. Retrieved from: <http://projeto.rcaap.pt/index.php/lang-pt/consultar-recursos-de-apoio/remository?func=startdown&id=364>. Access: Dec 2013.

COSTA, M. [2014] Características e contribuições da via verde para o acesso aberto à informação científica na América Latina. 226 f., il. Dissertação (Mestrado em Ciência da Informação) - Universidade de Brasília, Brasília. Retrieved from: <http://repositorio.unb.br/handle/10482/15687> Access: Jun 2015.

COSTA, S.; MEADOWS, J. [2000] The impact of computer usage on scholarly communication among social scientists. Journal of Information Science, v. 26, n. 4, p. 255-262. Retrieved from: <http://repositorio.unb.br/handle/10482/632>. Access: Aug 2013.

CRESSWELL, J. W. [2010] Projeto de pesquisa: métodos qualitativo, quantitativo e misto. 3 ed. Porto Alegre: Artmed. 296p.

ERBER, F. [2000] Perspectivas da América Latina em ciência e tecnologia. Parcerias estratégicas, n. 8, p. 181-200. Maio 2000. Retrieved from: <http://seer.cgee.org.br/index.php/parcerias_estrategicas/article/view/102>. Access: Jan 2013.

GORRAIZ, J. et al. [2009] International publication output and research impact in social sciences: comparison of the Universities of Vienna, Zurich and Oslo. Research Evaluation, v. 18, n. 3, p. 221-232.

GUMIEIRO, K. [2009] Modelos de negócios para periódicos científicos eletrônicos de acesso. 157 f. Dissertação (Mestrado em Ciência da Informação)-Universidade de Brasília, Brasília. Retrieved from: <http://hdl.handle.net/10482/3251>. Access: Dec 2013.

HOUGHTON, H. W.; STEELE, C.; HENTY, M. [2003] Changing research practices in the digital information and communication environment. Canberra: Department of Education, Science and Training. 186p.

HUANG, M.; CHANG, Y. [2008] Characteristics of research output in social sciences and humanities: from a research evaluation perspective. Journal of the American Society for Information Science and Technology, v. 59, n. 11, 2008, p. 1819–1828.

LEITE, F. [2009] Como gerenciar e ampliar a visibilidade da informação científica brasileira: RI de acesso aberto. Brasília: Ibict. 120 p.

LEITE, F. [2011] Modelo genérico de gestão da informação científica para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto. 262 f. Tese (Doutorado em Ciência da Informação - Universidade de Brasília), Brasília. Retrieved from: <http://hdl.handle.net/10482/9753>. Access: Feb 2013.

LYNCH, C.; LIPPINCOTT, J. [2005] Institutional repository deployment in the United States as of early 2005. D-lib Magazine, v. 11, n. 9, p. 5. Retrieved from: <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/september05/lynch/09lynch.html> Access: Aug 2013.

RAMLO, S. [2007] arXiv. org and Physics Education. The Physics Teacher, v. 45, n. 6, p. 374-375. Retrieved from:<http://scitation.aip.org/content/aapt/journal/tpt/45/6/10.1119/1.2768698>. Access: Feb 2015.

RIEH, S.; MARKEY, K.; JEAN, B.; YAKELL, E.; KIM, J. [2007] Census of institutional repositories in the United States. CLIR, n. 140. Retrieved from: <http://www.clir.org/pubs/abstract/reports/pub140>. Access: Feb 2013.

SARMENTO, M.; MIRANDA, A.; BAPTISTA, A.; RAMOS, I. [2005] Algumas considerações sobre as principais declarações que suportam o movimento Acesso Livre. In: World Congress on Health Information and Libraries, 2005. Salvador, Bahia, Brasil, 20-23 Set. Anais. Retrieved from: <http://hdl.handle.net/10760/8512>. Access: Jan 2013.

SHEARER, K. [2004] Survey Results—Summer 2004: CARL Institutional Repositories Project. Retrieved from: http://www.carlabrc.ca/projects/institutional_repositories/pdf/survey_results_.2004-e.pdf.

UNESCO. [2010] Informe de la Unesco sobre la ciencia 2010: el estado actual de la ciencia en el mundo. Unesco. Retrieved from:<http://unesdoc.unesco.org/images/0018/001898/189883s.pdf>. Access: Feb 2013.

WESTRIENEN, G.; LYNCH, C. [2005] Academic institutional repositories: deployment status in 13 nations as of mid 2005. D-Lib Magazine, v. 11, n. 9. Set. 2005. Retrieved from: <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>. Access: Feb 2013.

Author data

Michelli Pereira da Costa

Doutora e mestre em Ciência da Informação e bacharel em Biblioteconomia pela Universidade de Brasília. Tem experiência na área de Ciência da Informação, atuando principalmente nos seguintes temas: ciência aberta, dados de pesquisa, acesso aberto, bibliotecas digitais, bibliotecas populares e bibliotecas prisionais. Foi editora pelo Brasil do E-LIS Eprints in Library and Information Science de 2012/2014 (repositório temático internacional da produção científica em Ciência da Informação). Atuou como pesquisadora no IBICT em projetos de acesso aberto entre 2011 e 2014. É Professora Adjunta do curso de Biblioteconomia da Universidade de Brasília. **michellicosta11@gmail.com**

Fernando César Lima Leite

Graduado em Biblioteconomia. Mestre em Ciência da Informação. Doutor em Ciência da Informação. Experiência na área de Ciência da Informação, atuando principalmente nos seguintes tópicos: informação em ciência e tecnologia, gestão da informação e do conhecimento científico, biblioteconomia, planejamento e implementação de serviços de informação em C&T em ambiente digital, comunicação científica, divulgação científica, acesso aberto e repositórios institucionais, periódicos científicos eletrônicos. Foi editor pelo Brasil do E-LIS Eprints in Library and Information Science de 2006/2012 (repositório temático internacional da produção científica em Ciência da Informação). Foi analista da Empresa Brasileira de Pesquisa Agropecuária (Embrapa) onde planejou e implementou atividades de gestão da informação científica, tecnológica e organizacional e coordenou iniciativa e estratégias de acesso aberto à informação científica. Consultor do IBICT entre 2006 e 2009. Especialista Visitante do IBICT entre 2010 e 2013, no âmbito do PCI/CNPQ. Foi vice-coordenador e Coordenador do Programa de Pós-Graduação em Ciência da UnB. Atualmente é professor adjunto da Faculdade de Ciência da Informação, diretor da Biblioteca Central da Universidade de Brasília e bolsista de Produtividade em Pesquisa do CNPq - Nível 2.

fernandodfc@gmail.com

---

[1] RCAAP validator is an OAI-PMH interface validation system which aims to guarantee the interoperability of systems of this nature. Source: http://validador.rcaap.pt/

[2] Source: http://validator.oaipmh.com/

[3] Registry of Open Access Repositories (ROAR) – international, 254; Directory of Open Access Repositories (OpenDOAR) – international, 232; Red de Repositorios Latinoamericanos - regional (Latin America), 76; Rede LA Referencia - regional (Latin America), 16; Comunidad Latinoamericana de Bibliotecas y Repositorios Digitales (CoLaBora) – regional (Latin America),3; Lista dos repositórios brasileiros - national (Brazil), 79; Sistema Nacional de Repositorios Digitales (SNRD) – national (Argentina), 9.

[4] Source: http://validator.oaipmh.com/

[5] Source: http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/key-organizations/latin-america-and-the-caribbean/cybertesis/

[6] The disciplines quoted are the grand areas used by CAPES to classify the areas of knowledge. Source: http://www.capes.gov.br/images/stories/download/avaliacao/TabelaAreasConhecimento_042009.pdf