AtoZ
novas práticas em informação e conhecimento

ISSN: 2237-826X
PAPER | ARTIGO | ARTÍCULO

Scientific data curation in information science: national scenario survey

A curadoria de dados científicos na Ciência da Informação: levantamento do cenário nacional

Liliane Chaves de Resende[1], Marcello Peixoto Bax [2]

[1] *Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil. ORCID: https://orcid.org/0000-0002-4650-0540*

[2] *Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil. ORCID: https://orcid.org/0000-0003-0503-3031*

Mail to/*Autor para correspondência/Correo a:* Liliane Chaves de Resende, lilianederesende@gmail.com

Submitted/*Recibido*: September 12, 2019; **Approved/*Aceptado*:** August 16, 2020

Abstract

Introduction: For contemporary science, data sharing is a key element for its progress and collaboration among scientific communities. In the context of open science and e-science, information science investigates solutions to the challenges of treating and using information. The current lack of knowledge of the importance of digital curation activities of scientific data and the degree of adherence to this new scenario, cause difficulties for the information professional to adapt in this context and contribute solutions to the challenges of digital curation. **It aims** to investigate the importance that the digital curation activities of scientific data have for the area of Information Science in Brazil, in face of the international trend of scientific knowledge management. **Methodology:** The research follows a quantitative approach, it is applied in nature and exploratory and descriptive. It uses procedures of a survey with Survey, to obtain information from the Digital Curator through the opinion of Brazilian researchers in the area of Information Science. **Results:** The information analyzed reveals that the field of Brazilian Information Science is seeking to adapt in this scenario. **Conclusions:** The data allow us to affirm that there is not a sufficiently assimilated awareness by the researchers of Information Science regarding the need for more conscious and committed involvement in Digital Curation activities. New research is indicated to deepen this discussion.

Keywords: Digital Curation; Scientific data; Information Professional; Information Science; Survey; e-Science.

Resumo

Introdução: *para a ciência contemporânea, o compartilhamento de dados constitui elemento primordial para seu progresso e colaboração entre comunidades científicas. No contexto da ciência aberta e e-science, a ciência da informação investiga soluções para os desafios do tratamento e uso da informação. O desconhecimento atual da importância das atividades de curadoria digital dos dados científicos e do grau de adesão com esse novo cenário, causam dificuldades para o profissional da informação se adequar nesse contexto e contribuir com soluções para os desafios de curadoria digital.* ***Objetivo:*** *investigar a importância que as atividades de curadoria digital de dados científicos têm para a área da Ciência da Informação no Brasil, frente à tendência internacional do gerenciamento do conhecimento científico.* ***Método:*** *pesquisa de abordagem quantitativa, de natureza aplicada e do tipo exploratória e descritiva. Utiliza procedimentos de uma pesquisa com Survey, para obter informações da curadoria digital por meio da opinião dos pesquisadores brasileiros da área da Ciência da Informação.* ***Resultados:*** *As informações analisadas revelam que o campo da Ciência da Informação brasileira está buscando se adaptar nesse cenário.* ***Conclusões:*** *os dados permitem afirmar, que ainda não há uma conscientização suficientemente assimilada pelos pesquisadores da Ciência da Informação em relação à necessidade de envolvimento mais consciente e comprometido nas atividades de Curadoria Digital. Novas pesquisas são indicadas para aprofundar esta discussão.*

Palavras-chave: *Curadoria Digital; Dados Científicos; Profissional da Informação; Ciência da Informação; Survey; e-Science.*

INTRODUCTION

The Digital Curatorship of Scientific Data (CDDC) is gaining importance with research institutions. Academic communities are becoming aware of the need to maintain scientific information as a resource for continuity of research or new opportunities for knowledge discovery.

In several international scientific communities, the use and practice of CDDC activities has already been carried out in universities in the USA, UK, Europe, Australia, Asia and Africa. There is a notable participation of librarians in these projects, carrying out activities aimed at providing support to researchers in the management of scientific data, assisting in appropriate storage and dissemination activities of scientific data. The execution of activities necessary for CDDC practices can transform academic research libraries into a locus of scientific data management and curation.

In Brazil, this scenario is still in embryonic process. There are few initiatives of digital curatorship activities being effectively implemented in the Brazilian scientific communities. It is noticeable that there is great interest from institutions in developing the CDDC, but there is still no knowledge of a solid policy, with clear definitions on how this can become a mandatory activity in the institutions.

According to Kouper (2016), the initiative to obtain information on Digital Curatorship (CD), directly from those who research subjects in the area, expands the existing knowledge of the main competencies of the CD and improves the understanding of the knowledge, values and daily experience of researchers in the area of Information Science (CI). Therefore, one may ask, in order to declare the problem questioned by this research: what is the scenario of the IC in front of the movement of digital curatorship of scientific data, at national level ?

The objective of this research is to describe the current scenario of the DC area in Brazil through the opinion and attitudes of CI researchers regarding the basic activities of storage, use and reuse of raw data produced in their researches.

BRIEF OVERVIEW OF DIGITAL CURATORSHIP

Digital curatorship (CD) is an emerging discipline, applied and interdisciplinary, with academic and professional dimensions, progressing from the technical dimension of digital information as an object of research.

The CD is aimed at the digital environment whose specific function is to "care" for digital resources. As its main objective, it seeks to develop strategies to solve problems in the flow of digital information. This is its core activity. However, there are implications that are caused by the multiplicity and diversity of elements involved with digital information. This makes different levels of curatorship appropriate for different producers and consumers of digital information (Higgins, 2018; National Research Council, 2015).

The evolution of the CD area has grown from practices originated from traditional curatorship, appearing together with the definition of e-Science, or electronic science, defined by Jim Gray, Szalay, Thakar, and Stoughton in 2002, revolutionizing the format of scientific research. For Gray, researchers who are specialists in a particular field and want to perform curatorial activities should also learn relevant concepts and techniques used by librarians, such as the techniques of annotation, description and preservation of data (Gray et al., 2002; National Research Council, 2015).

Kouper (2016) states that the CD area grows at the intersection of librarianship, archival and information technology and their practices require a combination of skills from these areas and new ones to be defined. The desired competence includes technical familiarity with digital technologies and CD tools, the ability to initiate and manage projects, programs and services, with strong interpersonal and organizational skills.

A contribution to the evolution of the CD was made with the publication of the Open Archiving Information System (OAIS) Reference Model in 2003 by the International Organization for Standardization (ISO). This model, although it does not directly prescribe DC practices, proposes a structuring of responsibilities, processes and functions for archiving, preservation and long-term access to digital information.

Therefore, the Digital Scientific Data Curatorship (CDDC) appears in this scenario as an opportunity for the IC area and information professionals. CDDC activities constitute a resource that transforms the cycle of scientific communication, corroborating new discoveries of knowledge through collaboration among researchers (Sales & Sayão, 2012) despite presenting itself as a complex challenge in the face of the demands of managing the raw scientific data produced by the researches.

The proposal of the CDDC is, specifically, to assign value to the data produced by scientific research. This added value refers mainly to the maintenance of data that are impossible to recreate, aiming at the most diverse purposes: use in teaching, validation of published research results, analyzing those that may or may not be discarded, maintaining their integrity and quality for future reuse in other research, and may generate new knowledge (Poole, 2016).

The U.S. National Research Council, in its intensive CD study, concluded that the demand for readily accessible, accurate, useful and usable digital information has led to limitations related to digital data management. And to solve this, it is necessary to develop policies, services, technologies and knowledge in CD, as a way to articulate better opportunities to reduce costs and increase benefits to the whole knowledge society (National Research Council, 2015).

According to Nielsen and Hjørland (2014), digital search data are different from printed documents. The maintenance of such data sets requires scientific knowledge of the domain and advanced knowledge in technology to organize and store them for preservation and reuse. For authors, "data is often understood as the raw material of information processing and knowledge acquisition" (Nielsen & Hjørland, 2014, p. 3), specific knowledge on how to manage research data for a given domain is required. The various priorities, considerations and estimates from domain perspectives that lead to a requirement for deeper competencies of managers and trustees.

In this sense, the CDDC will be able to contribute to the demand for the management of the data produced by the research, proposing alternatives for its long-term preservation, guaranteeing access and reuse. Even institutional repositories are already considered the main instruments for knowledge dissemination, supporting the creation of new scientific discoveries (Sales & Sayão, 2012).

Research libraries are also moving to develop new roles and implement new service models. Librarians with specific information technology functions can take on leadership roles in scientific data management, working in partnership with researchers (Cox, Searle, Wolski, Simons, & Richardson, 2015).

Scientific advancement increasingly provides the need for access to research data, causing impacts on the efficiency and effectiveness of scientific activities. In this sense, librarians will need to become domain specialists, proposing means and resources for research needs in different domains, forming a partnership with researchers. They will be able to assist them in matters of articulating research infrastructure within the institutions, and in data curatorship activities (Lee & Stvilia, 2017; Nielsen & Hjørland, 2014, p. 3).

Institutional repositories are already considered as the main instruments for knowledge dissemination, supporting new scientific discoveries. Especially with the Open Science movement, which has provided the democratization of scientific information. But their creation is still recent and their use is little disseminated (Shintaku, Duque, & Suaiden, 2015).

The need to integrate CD activities into the activities of researchers can transform CDDC as part of the mission of CI information professionals. Lee and Stvilia (2017) interviewed professionals who work in institutional repositories and discovered that there are different levels in the CDDC activities performed in institutional repositories and that an initial meeting with the researcher is very important to define the extent of the CDDC needs. Examples of these activities are the issues of conceptualization, planning, creation, loading or uploading, storage and publication of scientific data. Throughout this process, the researcher is needed and will interact with various professionals, primarily librarians, archivists, IT professionals, acting according to each activity of the stages of the scientific data management life cycle. It is through this interaction that the needs of CDDC will be identified (Lee & Stvilia, 2017).

The role of institutional repositories, in this sense, has aligned with the traditional view of libraries in safeguarding and offering access to information. However, it has an even greater potential than the traditional library in providing support, generation, preservation and dissemination of scientific information. Therefore, the use of digital repositories and research libraries to organize scientific data is a necessary and highly indicated outcome for academic institutions (Lee & Stvilia, 2017; Shintaku et al., 2015; Tibbo & Hank, 2015).

In the Brazilian scenario, the CDDC is still in a process of recent and embryonic implantation. Some initiatives of digital curatorship activities are being effectively implemented in the Brazilian and governmental scientific communities. It is noticeable that there is great interest from institutions in developing the CDDC, but there is still no knowledge of a solid policy, with clear definitions on how this can become a mandatory activity in these institutions.

During the execution of this research, a total of nine repositories specifically focused on the storage, use and reuse of scientific data in Brazil, made available by Re3data (www.r3data.org), a global register of scientific data repositories that includes data sets for researchers, funding agencies, editors and academic institutions. The objective of Re3data is to promote the culture of sharing and access to research data. As some projects in progress in Brazil aiming the same objective, we can mention:

- The CarpeDIEN platform of Information on Nuclear Energy of the Institute of Nuclear Engineering (IEN/CNEN), linked to the Spanish Higher Council of Scientific Research (CSIC) since 2014, proposes to integrate research institutions in the nuclear area, with more than 45 Brazilian institutions already participating, including USP, Fiocruz and EMBRAPA. The purpose of this platform is to promote the preservation, curatorship and dissemination of the technical-scientific digital memory produced by the partner institutions. The researcher Dr. Luana Sales, librarian and one of the creators of the platform, participated in this project;
- The Center for Documentation and Digital Collection of Research (CEDAP), an auxiliary organ of the School of Librarianship and Communication (FABICO) of the Federal University of Rio Grande do Sul (UFRGS), whose purpose is to provide support to institutional scientific research. It offers several services for storage, access and sharing of scientific data;
- The Brazilian Scientific Data Network (RDP Brasil) initiative is a partnership between the National Teaching and Research Network (RNP) and the Brazilian Institute of Information on Science and Technology (IBICT) in collaboration with the Federal University of Rio Grande do Sul (UFRGS) and the Federal University of Rio Grande (FURG). It aims to develop a joint project on Open Access to Research Data (AADP), whose purpose is to develop activities that contribute to the identification of open access practices to research data in Brazilian institutions, mapping requirements and developing a prototyping of AADP systems to raise a comparison of services and technological solutions of AADP that aims to facilitate the dissemination of scientific information;
- The CRUESP Repository of Scientific Production, of the Council of Rectors of the State Universities of São Paulo - USP, UNICAMP and UNESP, whose purpose is to gather, preserve and provide open, public and integrated access to the scientific production of professors, researchers, students and servants of the

participating universities. Integrated to CRUESP, are the institutional repositories of intellectual production (scientific, artistic, academic and technical) of USP, the Digital Library of Intellectual Production of the University of São Paulo (BDPI), the Virtual Library of FAPESP, the Digital Library of Intellectual and Scientific Production of UNICAMP and the Institutional Repository of UNESP.

In addition to these projects, there are already proposals in Brazil to raise awareness of the participation of information professionals in CDDC projects. But it is important to find out what the engagement of researchers in the field of Information Science is with this process and how they are organizing themselves to contribute to the development of the DC area in the face of the demands of the scientific field, as is already happening in other countries. Based on this assumption, obtaining opinions about the CDDC activities performed by Brazilian CI researchers is essential to describe how the CD area in Brazil is evolving and its current state, specifically for the CI field.

METHODS

The research follows a quantitative approach, is of an exploratory and descriptive nature. A research with Survey is used to obtain information from the Digital Curator through the opinion of Brazilian researchers in the area of Information Science. Table 1 below presents the characterization of the survey with Survey:

Research Technique	Research with Exploratory Survey.
Methods	Survey of opinions.
Objective	Describe the population investigated at a given time to analyze the context of the CDDC at the national level.
Nature	Quantitative, exploratory and descriptive.
Population	Researchers in the area of information science participating in graduate programs stricto sensu at Brazilian universities recommended by CAPES to the National Council of Education - CNE/MEC, registered in the Sucupira platform.
Sample Selection	Random probability sampling (equal probability of population selection), self-selected sampling.
Collecting Instrument	Online questionnaire.
Methods of analysis	Statistical analysis by means of absolute frequencies of the obtained answers, describing attitudes, preferences and population trends.
Validity of the instrument	Application of the questionnaire as an intentional pre-test with a random sampling design applied to 3 researchers.

Table 1. Survey Characterization *Survey*.
Source: Prepared by the authors, 2019.

The target population was composed by all the Brazilian researchers in the area of CI, linked to graduate programs *stricto sensu* (Master and Doctorate) in the area of Information Science, selected in the Sucupira Platform of the Coordination for the Improvement of Higher-Level Personnel (CAPES), with equal chances of participation.

In this study, the sampling method used was simple random sampling. The sampling was self-selected, that is, composed of researchers who accepted the invitation to participate in the survey. According to Oliveira and Grácio (2005), it is also necessary to ensure that the sample has the same general population characteristics. Therefore, the calculation of the sample size is determined based on the number of elements needed to compose the sample, in order to obtain valid results.

The researchers were contacted through e-mail messages sent to their respective mailboxes, individually. The e-mails were obtained through public information, made available on the official pages of graduate programs in the area of CI or through articles published in open access journals. The questionnaires were sent by e-mail to 390 members of the target population. About 78 members of the target population were unable to participate because it was not possible to locate e-mail references on institutional, publicly accessible websites.

According to the rate of respondents obtained, through 97 respondents, it was possible to obtain a representative range of researchers in the survey. Marconi and Lakatos (2003) state that questionnaires sent to respondents reach, on average, 25% of return. Oliveira and Grácio (2005) suggests that for a sample to be representative, it must cover a fixed percentage of the population, of approximately 10% to 20% and that this percentage must represent at least 30 to 40 elements of the population, otherwise it is considered too small. Therefore, the quantitative obtained from respondents suggests a representative and valid sample, since it is within the estimate suggested by the authors, i.e., 25% of the target population, and sufficient to generalize the findings.

The data was collected during the months of May, June, July and August 2019 through a link to access the Google Forms web form. This tool was chosen because of its simplicity, ease of use and agility to obtain data,

since it is perfectly possible to use this tool to obtain a reasonable number of data, in a short period of time and in different geographical areas, without cost. Even to allow more flexibility to the respondent (Gil, 2008).

A free trial version of the SPSS Statistics Subscription (Statistical Package for the Social Sciences) software was used to tabulate the responses obtained, which were analyzed using their absolute frequencies. The questions developed in the research instrument used totaled 26 multiple choice questions and two open-ended questions (Appendix 1 - Data collection research tool).

This survey requires a basic initial understanding of CDDC concepts by respondents. Given the emerging nature of the discipline, this may have been a factor that has limited or not motivated more respondents to participate in the survey.

RESULTS

To describe the current state of the CD in the Brazilian scenario, data were obtained for: 1) characterization of the profile of Brazilian researchers in the IC area who work in graduate, master and doctorate programs; 2) the level of knowledge of these researchers about the CD; 3) the level of involvement in the activities of the CDDC when developing their research; 4) trends and opinions of these researchers about the CDDC.

Profile of the Brazilian Researcher in Information Science

To present the researcher's profile in the IC area, personal data, basic academic data and CD activities were collected. The data obtained are presented in Table 2 below.

		N	Cases
Gender	Female	57	58.8
	Male	39	40.2
	Didn't inform	1	1.0
Academic qualifications	Doctorate	85	87.6
	Master	12	12.4
Age group	21 to 30	1	1.0
	31 to 40	24	24.7
	41 to 50	26	26.8
Research experience (years)	2 to 5	16	16.5
	6 to 10	28	28.9
	11 to 20	22	22.7
	More than 20	31	32.0
Locus of perfomance	Public sector	44	46.3
	Private sector	21	22.1
Research locus	Place of work	80	84,2
	Brazilian Universities	56	58.9
	International University	29	30.5
	With Research Institutes	36	37.9
	Non Governmental Org.	11	11.6
	Government Entity	1	1.1
Financing locus	National promotion	33	34.7
	International promotion	5	5.3
Areas of knowledge of graduation	Exact Sciences	13	13.4
	Human Sciences	12	12.4
	Engineering	9	8.1
	Health Sciences	2	2.1
	Linguistics, Literature and Arts	7	7.2
	Applied Social Sciences	68	70.1

Table 2. Respondent profile.
Source: Research data.

In short, the participant is over 40 years of age, is mostly female and almost all have PhD degrees, with over 10 years of experience in academic research. He/she works mainly in public sector universities, and most of them seek to promote research with national entities. In total, 58.2% interact with other Brazilian universities

and 32.8% with international universities, demonstrated relatively low internationalization. As for scientific education, it is clear the participation of graduation in other areas of knowledge.

		N	%
Is digital curatorship an opportunity for the information professional?	I strongly disagree	1	1.0
	I disagree	1	1.0
	No opinion	5	5.2
	I agree	43	44.3
	I strongly agree	47	48.5
Participation in events	No	47	48.5
	Yes	49	50.5
	I can't say	1	1.0
Participation and training	No	79	81.4
	Yes	18	18.6
Policy making National standards	No	83	85.6
	Yes	13	13.4
	I can't say	1	1.0
MANAGEMENT of scientific data produced	Personal PC storage	76	78.4
	Store on Google Drive	72	74.2
	Stay with the students	1	1.0
	External HD storage	3	3.1
	Disposal	2	2.1
	GP Platform*	1	1.0
	Journal submitted article	1	1.0
	I make available on my own website	1	1.0
	I make available Repositories	1	1.0
	Partner platform work	2	2.1
	Use of Institutional Assets	1	1.0
	IE work platform	21	21.6
Cited Scientific Data	Yes	49	50.5
	No	46	47.4
	I can't say	3	3.1
Published scientific data	Yes	47	48.5
	No	52	53.6
	I can't say	1	1
Scientific data produced by the researcher	Publish along with the articles	54	55.7
	Disclosure at scientific events	48	49.5
	Stores personal computer	68	70.1
	Available in repository	24	24.7
	Discard the data	2	2.1
	Other	8	8.1
Use of metadata standards	DataCite Metadata Schema	4	4.1
	DDI	4	4.1
	Dublin Core	40	41.2
	ISO 19115	2	2.1
	Other	2	2.1
	I can't say	50	51.5
Sharing scientific data in Storage Technology Platform	DSpace	44	45.4
	Zenodo	7	7.2
	Dataverse	8	8.2
	FigShare	5	5.1
	Tainacan	2	2.1
	CKAN	5	5.1
	EPrints	15	15.5
	I have never used	45	46.4
	Other	6	6.2

		N	%
ACTIONS on data produced in the research	Stored	96	99.0
	Accessed	77	79.4
	Shared	61	62.9
	Made available	45	46.4
	Reused	59	60.8
	Storage for own use	1	1.0
	Sharing with GP*	1	1.0
	I published	1	1.0
Brazilian Repositoriesof data	BCD/ UFPR	35	36.1
	CIS	14	14.4
	GLOBE	14	14.4
	IBGE	70	72.2
	IBICT	61	62.9
	Other	16	16.3
	None	2	2.1
REUSE of data in Technology Platform	DSpace	41	42.3
	Dryad	1	1.0
	Zenodo	9	9.3
	Dataverse	9	9.3
	FigShare	6	6.2
	Tainacan	4	4.1
	CKAN	6	6.2
	EPrints	14	14.4
	EUDAT	2	2.1
	I have never used	49	50.5
	BDC/UFPR	1	1.0
	Dadosabertos.info	1	1.0
	Webmuseum	1	1.0
	Wordpress	1	1.0
	My own articles	1	1.0
	Comunidade FIOCRUZ	1	1.0
Life cycle model as reference	DCC	17	17.5
	DDI	2	2.1
	UK Data Lifecycle	3	3.1
	CVD-CI	2	2.1
	DataONE	7	7.2
	Research360	1	1
	OAIS	14	14.4
	IE work permit	2	2.1
	Institution of promotion	11	11.3
	I can't say	46	47.4
	No models	3	3.1
	Other	3	3.1
Surveys are more cited when making the data available	I strongly agree	22	22.7
	I agree	52	53.6
	No opinion	15	15.5
	I disagree	8	8.2
	I strongly disagree	0	0

* Research Group

Table 3. Basic knowledge on digital curatorship of scientific data.
Source: Research data (2019).

Involvement in Scientific Data Digital Curatorship Activities

In order to describe if the participant has allusion and basic knowledge of digital curatorship, 93% say they know the general definition of CDs and almost all (82%) have never participated in the elaboration of national policies or norms directed to CDs. It is noticeable that, from the suggestions presented by some researchers, there is already a movement to use research data repositories, mainly in institutional repositories, but 50.5% reported never having used this resource. As for the participation in specific events on the subject, 49% said they had participated, but few (18%) researchers have already participated in training or specific training in the area.

Basic CD activities cover several practices that enable the administration and management of digital information,

among them: collect, describe, represent, archive and preserve data (Thompson, Senseney, Baker, Varvel, & Palmer, 2013). Regarding these activities, almost all (99%) researchers reported that they have already stored, 79.4% accessed, 60.8% reused and 62.9% shared their own research data at some time.

On average, 51.9% of researchers claim to store their scientific data on personal devices. However, 46.4% of researchers have already made their research data available, and few (1%) have reported having published their raw research data. The data obtained are presented in Table 3.

Some information differs from the use of digital repositories to store and reuse research data. About 98% of researchers have confirmed to have already used digital repositories. This information suggests that researchers are aware of the importance of data repositories for scientific research.

However, it is perceived that it is necessary to develop an effective national policy for the management and sharing of scientific data, since 46.4% said they have never used a digital repository to store their own research data, even storing them on personal computers (70%). Only 24.7% claim to make them available and 2% even discard them when they finish the research. The practice of data sharing is not a common activity for Brazilian researchers (Sayão & Sales, 2016; Vanz et al., 2018).

In relation to the reuse of research data, both their own and others', they also confirm the interest of 60.8% of researchers. The facts suggest the understanding that Information Science "is a collaborative, open and digital environment" (Oliveira & Silva, 2016, p. 11).

However, even knowing the trend of open access to scientific data, 46.4% make it openly available. And when investigated if they have already cited data from other researchers, 49% said yes and 47% say they have already published their own data. The researchers (76.3%) also agree that studies that make research data available may be more cited.

Poole (2016) argues that the citation and publication of scientific data still generates many discussions. The researcher states that the citation of scientific data is important because the set of data generated in research are considered first-class research products and should be publicly and perpetually available. Citing data generates other challenges such as granularity, micro attribution, contribution identifiers and citation placement (Poole, 2016).

The data obtained confirm this view which is justified with the argument that "in Brazil the lack of uniform instruments and guidelines for the preservation and safety of research data - the raw material of the scientific enterprise" (Vanz et al., 2018, p. 21). Oliveira and Silva (2016) allege that the advance in the Brazilian legal-regulatory framework, referring to institutional repositories, is slow and that restrictive factors are beyond technological and technical issues; they are mainly based on political, legal, economic and cultural spheres.

The evidence regarding the availability of data also corroborates the statement of Vanz et al. (2018) by the low percentage of respondents, 24.7% who claim to make their data available in research data repositories. This figure is still well below the percentage of respondents who reuse survey data, 66.6% (Oliveira & Silva, 2016; Vanz et al., 2018).

It is noticeable that there is a lack of understanding about the practices of CD de facto (share, make available, publish, quote). Especially because 46.4% claim they have never used repositories to share their research data, 51.5% can't say or don't use metadata standards to describe their scientific data and can't even say which reference model to use in the elaboration of the Research Data Management Plan (PGP).

About the techniques related to CDDC activities on scientific data description, Poole (2016) clarifies that studies indicate that researchers are unfamiliar with creating or documenting metadata, and that they cannot predict the needs of those who would reuse their data. It also reinforces that structured metadata is essential for the sharing and reuse of data. It is important that CI researchers understand that these CDDC practices play an essential role in the management of scientific data and, through them, can assist other researchers in their research (Poole, 2016; Tripathi, Shukla, & Sonkar, 2017).

Therefore, in order to move forward with the CDDC, efforts are needed in several instances, mainly involving "differentiated action and decision making, both internal and external to science, ranging from the individual researcher and research teams to the macro level of public policies and international regulations, as well as the same level of scientific institutions and development agencies" (Abagli, 2015, p. 21).

Scientific Data Digital Curator Preferences and Opinions

Table 3 presents data from participants. In relation to the preferences of researchers for the area of the CDDC, it was considered that ethical issues in the processes of scientific research are more important in the performance of CD activities (89.7%). About 75.3% of the researchers identified that the guarantee of usability and accessibility of scientific data is an important CD activity for the information professional, and that finding data and publications for reuse purposes is the main activity that the information professional can perform to help other researchers in their researches.

Researchers state that for an information professional to act in digital repositories of scientific data, he or she must master important skills on scientific data, such as understanding legal, regulatory and copyright policy frameworks, forms of access to digital data and intellectual property issues, with 86.6%, 87.6% and 84.5% respectively of responses obtained.

The participants considered the sustainability and maintenance of scientific data associated with the evolution, development and constant technological changes as the most critical issue for Brazilian DC research, but considered it uncritical to pay attention to education, training and development of professionals who can support DC activities.

		N	% Cases
Activities that the information professional can help the researcher	Store and preserve data	52	53.6
	Develop a data management plan	45	46.4
	About legal issues related to my research data	42	43.3
	Share data	45	46.4
	Creation of metadata	47	48.5
	Find data and publications for reuse	53	54.6
	Data Citations	45	46.4
	I can't say	3	3.1
	I agree with all the above	43	44.3
	Data manipulation and visualization generation	1	1.0
	Ethical and legal milestones of the CD	2	2.1
Important CD activities for the information professional	Life cycle management / continuum of digital objects, perhaps even the creation of the record keeping system for these materials	61	62.9
	Active involvement over time and over the long term of the creators of the scientific data	48	49.5
	Evaluation and selection of scientific data for availability in digital repositories	68	70.1
	Development and provision of access to scientific data	53	54.6
	Ensure the preservation (usability and accessibility) of scientific data	73	75.3
	Data analysis and visualization	1	1.0
	I don't know	2	2.1
Digital repository skills	Ethical processes of scientific research	87	89.7
	Research methods	79	81.4
	Forms of Scientific Communication	80	82.5
	Intellectual Property	82	84.5
	Ways to access digital data	85	87.6
	Metadata standards	81	83.5
	Legal and Regulatory Milestones and Copyright Policies	84	86.6
	All previous options	72	74.2
	Information and knowledge organization	1	1.0
	Knowledge of international practices and models	1	1.0

Table 4. Preferences and Opinions of CDDC.
Source: Research data (2019).

The Digital Curatorship of Scientific Data Scenario in the opinion of CI Researchers

About the stage of evolution of the CDDC in Brazil, the researchers affirmed that the CDDC is in its embryonic form, taking the first steps. Claims that the CDDC is "in need of further studies to develop it" are a sample of this scenario.

One respondent mentions that there is still no national policy for managing and sharing scientific data. But that there are governmental movements to discuss and develop a draft Decree for open science in Brazil, through which the role of development agencies, access to international resources for research development, availability of data in international repositories have been discussed.

Therefore, according to CI researchers, it seems a long way to go for the CI area to take over the CD as part of its mission. However, responses indicated that the CDDC in Brazil, and not only in the IC, is "in its expansion phase, although it is not widely discussed within institutional spaces, as well as in specific disciplines" and "becoming an issue that needs the researchers' attention, because the organization and access to open digital object data is an eminent need".

FINAL CONSIDERATIONS

This survey presented the Survey to survey and describe the current scenario of the Digital Curatorship of Scientific Data (CDDC) for the field of Information Science (CI) in Brazil, through the attitudes, opinion and preferences of its research representatives.

Contributes with the presentation of the professional and academic performance scenario of the CI researchers in CDDC. It analyzes how the field of CI, through its research representatives, perceives the important role it plays in this phenomenon and also how it is following and interacting with the international movement of CDDC. From this presentation we can begin a more in-depth discussion about strategies to increase the relevance of the CDDC activities carried out by these professionals, and consequently by the IC itself..

Through the data collected and the opinion of the researchers, it is inferred that there is an initial movement for adherence to the digital curatorship of scientific data. But there is an understanding that the CDDC's mission goes far beyond simply making data available and preserving it, but knowing how to administer and manage how the data can be reused, contributing to new scientific discoveries. The survey allows us to conclude that there is a low performance of researchers on the topic.

It was clearly found that there is still no effectiveness of CD practices of the data produced by these researchers themselves. It is clearly observed that there is not enough involvement with the CDDC on the part of the researchers to be able to claim that the national IC will take the CDDC as a pillar or a founding part of its mission. This requires a mastery of these activities and the concepts involved to carry them out.

Thus, it was concluded, in general, that a considerable evolutionary change in the theoretical, practical and technical disciplinary training of these researchers is still necessary. Such conclusion is even more relevant for the strengthening of CI as an area of knowledge as a whole, because it is these same researchers who form the future information professionals who will act in professional practices and technical information services.

One limitation of this research is that the participation of national CI researchers could have been even greater and also more comprehensive. It is believed that the objectives initially proposed, even if preliminarily, were achieved.

As future works, we suggest new researches that seek to generate even more detailed information to subsidize and deepen this discussion, complementing, corroborating or refuting the results achieved so far. The detailed survey of the descriptive information of the IC scenario in front of the CDDC will allow the establishment of strategies for the evolution of the treatment of this subject by the IC area in the country. It will also allow the discussion, in an empirically based way, about gaps in the disciplinary formation of the information professional, both at the undergraduate and graduate levels.

Another suggestion could be to carry out research that goes deeper into more specific themes in the field of the CDDC, such as the integration, reuse, publication and preservation of scientific data; the use of scientific data repository tools and not just texts (articles, monographs, theses and dissertations). The collection of more detailed information on these activities and tools will allow CI to reflect on strategies and means to raise awareness in the academic community of the importance of the CDDC for Information Science, as well as the importance of making the data produced in its research available. It is believed that such discussion will allow the IC to make a definite contribution to improving data reuse rates and conditions for reproduction of scientific research results in all areas of knowledge, as well as of its own researchers.

APPENDIX 1 - DATA COLLECTION RESEARCH TOOL

1. CARACTERIZAÇÃO DO RESPONDENTE

Trabalha em qual Instituição:

Qual Formação acadêmica de graduação?

Titulação acadêmica
( ) Mestrado
( ) Doutorado

Idade?
( ) 21 a 30
( ) 31 a 40
( ) 41 a 50
( ) Mais de 50

Gênero?
( ) Feminino
( ) Masculino
( ) Não quero informar

Tempo de atuação como pesquisador?
( ) 2 a 5 anos
( ) 6 a 10 anos
( ) 11 a 20 anos
( ) Mais de 20 anos

Entidades onde realiza suas pesquisas? Selecione todas as aplicáveis.
( ) Universidade onde trabalho
( ) Universidade no Brasil
( ) Universidade internacional
( ) Instituto de pesquisa
( ) Organização não governamental
( ) Instituição de fomento à pesquisa nacional
( ) Instituição de fomento à pesquisa internacional
( ) Setor público
( ) Setor privado
( ) Outro:

2. INFORMAÇÕES SOBRE CURADORIA DIGITAL

A Curadoria Digital é um conceito abrangente que inclui o gerenciamento de dados científicos ao longo ciclo de vida das pesquisas. Inclui atividades de preservação de dados para acesso, compartilhamento, uso e reuso futuro. Este campo está se tornando uma oportunidade de carreira para o profissional da informação. Qual sua opinião mediante esta afirmativa?

( ) Concordo fortemente
( ) Concordo
( ) Sem opinião
( ) Discordo
( ) Discordo fortemente

Atualmente no Brasil, algumas instituições de pesquisa utilizam repositórios digitais de dados científicos de experimentos, medições ou entrevistas, para colaboração entre pesquisadores. Quais dos repositórios de dados científicos você já ouviu falar. Marque todas as opções aplicáveis.

( ) BDC/UFPR - Base de Dados Científicos da UFPR
( ) CIS – Consórcio de Informações Sociais
( ) GLOBE – Global Collaboration Engine
( ) IBGE – Instituto Brasileiro de Geografia e Estatística
( ) IBICT Dataverse Network
( ) Outro:

Como você faz o gerenciamento dos dados produzidos em suas pesquisas?

( ) Armazeno e mantenho no meu computador pessoal
( ) Armazeno e mantenho no Google Drive, Box ou outro similar
( ) Utilizo a plataforma fornecida da instituição em que trabalho
( ) Utilizo uma plataforma específica de armazenamento parceira da instituição em que trabalho
( ) Descarto quando termina a pesquisa
( ) Outro:

Em relação aos dados produzidos em suas pesquisas, marque as ações que você já realizou. Selecione todas as opções aplicáveis.

( ) Armazenou
( ) Acessou
( ) Compartilhou

( ) Disponibilizou
( ) Reutilizou
( ) Nenhuma das acima

Quais dessas plataformas tecnológicas você já utilizou para reutilizar dados de suas pesquisas? ( ) DSpace
( ) Dryad
( ) Zenodo
( ) Dataverse
( ) FigShare
( ) Tainacan
( ) CKAN
( ) EPrints
( ) EUDAT
( ) Outro:
( ) Nunca utilizei plataforma de gerenciamento de dados

O IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia) lançou em 2016 o Manifesto de Acesso Aberto a Dados de Pesquisa Brasileira para a Ciência Cidadã, para apoiar movimentos e iniciativas para a Ciência Aberta no Brasil, traduzidos pelo amplo e irrestrito acesso a fontes primárias de pesquisa possibilitando compartilhamento e reutilização de dados de pesquisa. Você tem conhecimento desse Manifesto?

( ) Sim, mas não o conheço bem
( ) Sim, e o conheço bem
( ) Não sei do que se trata
( ) Não sei dizer

Já participou da elaboração de políticas e normas para o desenvolvimento das práticas de Curadoria Digital de dados científicos?

( ) Sim
( ) Não
( ) Não sei dizer

Já participou de algum evento que abordou como tema principal a Curadoria Digital de dados científicos?

( ) Não
( ) Sim
( ) Não sei dizer

Já participou de treinamento técnico (curso, capacitação) em práticas de Curadoria Digital de dados científicos?

( ) Sim
( ) Não
( ) Não sei dizer

3. PREFERÊNCIA, VISÃO E EXPERIÊNCIA

Identifique qual a importância das disciplinas/componentes curriculares que poderiam ser incluídos no currículo de graduação dos profissionais da informação (bibliotecários, arquivistas, cientistas da informação) para atuarem na área de Curadoria Digital? (1 – menos importante a 5 – mais importante)

	5 Extremamente importante	4 Muito importante	3 Moderadamente importante	2 Ligeiramente importante	1 Nem um pouco importante
Teoria de metadados					
Preservação digital					
e-Science					
Aprendizado de máquina					
Banco de dados					
Bibliotecas digitais					
Modelagem da Informação					
Análises estatísticas					
Big data					
Visualização de dados					
Linguagem de programação					
Ontologias					
Organização da informação					
Gerenciamento e análise de sistemas					
Recuperação da informação					

Quais habilidades você considera importante para um profissional da informação deve dominar em repositórios digitais de dados científicos?

( ) Processos ética da pesquisa científica
( ) Métodos de pesquisa
( ) Formas de Comunicação científica
( ) Propriedade intelectual
( ) Formas de acesso a dados digitais
( ) Padrões de metadados
( ) Marcos legais e regulatórios e Políticas de direito autoral
( ) Todos as opções anteriores
( ) Nenhuma das opções anteriores
( ) Outros:

Das plataformas tecnológicas de armazenamento de dados abaixo, qual ou quais você já utilizou para compartilhar dados de suas pesquisas?

( ) DSpace
( ) Dryad
( ) Zenodo
( ) Dataverse
( ) FigShare
( ) Tainacan
( ) CKAN
( ) EPrints
( ) EUDAT
( ) Nunca utilizei plataforma de gerenciamento de dados
( ) Outro:

Qual padrão de metadados você utiliza para representar seus dados científicos?

( ) DarwinCore
( ) DataCite Metadata Schema
( ) DDI (Data Documentatio Initiative)
( ) Dublin Core
( ) ISO 19115
( ) Outro:
( ) Não sei dizer

Que modelo de ciclo de vida de dados você tem como referência para elaborar seu próprio Plano de Gerenciamento

de Dados de Pesquisa?

( ) Digital Curation Centre (DCC)
( ) Modelo de Atividade de Pesquisa Idealizada I2S2
( ) DDI Combined Lifecycle Model
( ) UK Data Archive Data Lifecycle
( ) CVD-CI
( ) DataONE Data Lifecycle
( ) Research360
( ) OAIS
( ) Modelo padrão oferecido pela Instituição de Fomento à pesquisa
( ) Não sei dizer
( ) Outro:

Você já citou dados científicos produzidos por outros pesquisadores? (Cadernos de anotações de Laboratório ou campo, diários e transcrições textuais, dados de entrevistas e de questionários, Fluxos de trabalho e metodologias, registros de projetos de pesquisa, Fotografias, Arquivos de áudio, software, dentre outros).

( ) Não
( ) Sim
( ) Não sei dizer

Você concorda que estudos científicos que disponibilizam os dados produzidos podem ser mais citados do que estudos semelhantes que não o fazem.

( ) Concordo fortemente
( ) Concordo
( ) Sem opinião
( ) Discordo
( ) Discordo fortemente

Você já publicou dados científicos, por exemplo cadernos de anotações de Laboratório ou campo, diários e transcrições textuais, dados de entrevistas e de questionários, Fluxos de trabalho e metodologias, registros de projetos de pesquisa, Fotografias, Arquivos de áudio, software, dentre outros?

( ) Não
( ) Sim
( ) Não sei dizer

Em relação aos dados produzidos em suas pesquisas, normalmente o que você faz?

( ) Publica junto com os artigos
( ) Divulga em eventos científicos
( ) Armazena no computador pessoal
( ) Disponibiliza em repositórios de dados de pesquisa
( ) Descarta os dados
( ) Outro
( ) Não sei dizer

Qual(ais) necessidade(s) de formação técnica curricular que você precisa para praticar a Curadoria Digital de dados científicos?

4. ACOMPANHAMENTO DE TENDÊNCIAS

Em quais atividades de Curadoria Digital um profissional da informação pode auxiliar um pesquisador? Selecione todas as opções aplicáveis.

( ) Armazenar e preservar dados
( ) Elaborar um plano de gerenciamento de dados
( ) Sobre questões legais relacionadas aos dados de minhas pesquisas
( ) Compartilhar dados
( ) Criação de metadados
( ) Encontrar dados e publicações para reuso
( ) Citações de dados

( ) Não sei dizer
( ) Concordo com todas as anteriores
( ) Outra:

Quais das atividades de Curadoria Digital abaixo são mais importantes para que um profissional da informação possa interagir junto aos pesquisadores no desenvolvimento de suas pesquisas científicas?

( ) Gerenciamento do ciclo de vida / continuum dos objetos digitais, talvez até mesmo a criação do sistema de manutenção de registros desses materiais
( ) Envolvimento ativo ao longo do tempo e a longo prazo dos criadores dos dados científicos
( ) Avaliação e seleção dos dados científicos para disponibilidade em repositórios digitais
( ) Desenvolvimento e provisão do acesso aos dados científicos
( ) Garantir a preservação (usabilidade e acessibilidade) dos dados científicos
( ) Outro:

Na sua opinião, qual das preocupações relacionadas abaixo é mais crítica para a pesquisa científica brasileira?

( ) A sustentabilidade e manutenção dos dados científicos em ambientes tecnológicos a longo prazo
( ) A criação ou avaliação, seleção, aquisição e ingestão e manutenção dos dados científicos para uso futuro
( ) Evolução, desenvolvimento e mudanças tecnológicas constantes
( ) Disponibilização dos dados científicos (publicação)
( ) Gerenciamento de acesso e outros controles sobre os dados científicos, primando pela base ética e legal desses controles
( ) A capacidade de usar, combinar, recombinar, interagir, processar, anotar, discutir e revisar os dados científicos ao longo do tempo (reuso para novas descobertas do conhecimento)
( ) Gerar ligação, contexto e metadados relacionados a dados científicos em softwares específicos
( ) Manter as informações de autenticidade, integridade, proveniência e contexto computacional relacionadas aos dados científicos
( ) Preservação digital dos dados científicos, incluindo preservar o acesso a estados passados dos dados científicos
( ) Descarte e destruição deliberada e / ou acidental dos dados científicos, em períodos de tempo potencialmente ampliados, embora os prazos possam ser comparativamente de curto ou médio prazo
( ) Reconhecer os impactos dos orçamentos financeiros necessários e potenciais mudanças futuras na política do país
( ) Prestar atenção à educação, treinamento e desenvolvimento de profissionais que possam dar suporte às atividades necessárias de Curadoria Digital.

Na sua opinião, a Curadoria Digital de dados científicos, atualmente, no Brasil está ...

REFERENCES

Abagli, S. (2015). Ciência aberta em questão. In *Ciência aberta, questões abertas* (pp. 09–25). Brasília: IBICT.

Cox, A., Searle, S., Wolski, M., Simons, N., & Richardson, J. (2015). Librarians as partners in research data service development at griffith university. *Program: Electronic Library and Information Systems.*, *49*(4), 440–460. doi: 10.1108/PROG-02-2015-0013.

Gil, A. C. (2008). *Métodos e técnicas de pesquisa social.* (6th ed.). São Paulo: Atlas.

Gray, J., Szalay, A. S., Thakar, A. R., & Stoughton, C. (2002). Online scientific data curation, publication, and archiving. In *Virtual observatories* (Vol. 4846, pp. 103–107). Redmond, WA: International Society for Optics and Photonics. Retrieved from https://arxiv.org/pdf/cs/0208012.pdf.

Higgins, S. (2018). Digital curation: The development of a discipline within information science. *Journal of Documentation*, *74*(6), 1318–1338. doi: 10.1108/JD-02-2018-0024.

Kouper, I. (2016). Professional participation in digital curation. *Library and Information Science Research*, *38*(3), 212–223. doi: 10.1016/j.lisr.2016.08.009.

Lee, D. J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PlosOne*, *12*(3), e0173987. doi: 10.1371/journal.pone.0173987.

Marconi, M. D. A., & Lakatos, E. M. (2003). *Fundamentos de metodologia científica.* (5th ed.). São Paulo: Atlas.

National Research Council. (2015). *Preparing the workforce for digital curation.* Washington, DC: National Academies Press. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK293663/.

Nielsen, H. J., & Hjørland, B. (2014). Curating research data: the potential roles of libraries and information professionals. *Journal of Documentation*, *70*(2), 221–240. doi: 10.1108/JD-03-2013-0034.

Oliveira, & Grácio, M. C. C. (2005). Análise a respeito do tamanho de amostras aleatórias simples: uma aplicação na área de ciência da informação. *DataGramaZero: Revista de Ciência da Informação*, *6*(3), 01–11.

Oliveira, & Silva, E. M. d. (2016). Ciência aberta: dimensões para um novo fazer científico. *Informação & Informaçã*, *21*(2), 05–39. doi: 10.5433/1981-8920.2016v21n2p5.

Poole, A. H. (2016). The conceptual landscape of digital curation. *Journal of Documentation*, *72*(5), 961–986. doi: 10.1108/JD-10-2015-0123.

Sales, L. F., & Sayão, L. F. (2012). O impacto da curadoria digital dos dados de pesquisa na comunicação científica. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, *7*(2), 118–135. doi: 10.5007/1518-2924.2012v17nesp2p118.

Sayão, L. F., & Sales, L. F. (2016). Curadoria digital e dados de pesquisa. *AtoZ: novas práticas em informação e conhecimento*, *5*(2), 67–71. doi: 10.5380/atoz.v5i2.49708.

Shintaku, M., Duque, C., & Suaiden, E. J. (2015). Análise da adesão às tendências da ciência pelos repositórios institucionais brasileiros. *InCID: Revista de Ciência da Informação e Documentação*, *6*(2), 148–169. doi: 10.11606/issn.2178-2075.v6i2p148-169.

Thompson, C. A., Senseney, M., Baker, K. S., Varvel, V. E., & Palmer, C. L. (2013). Specialization in data curation: Preliminary results from an alumni survey. *Proceedings of the American Society for Information Science and Technology*, *50*(1), 01–04. doi: 10.1002/meet.14505001151.

Tibbo, H. R., & Hank, C. (2015). Digital data curation essentials for data scientists and data curators and librarians. In *Proceedings of the 15th acm/ieee-cs joint conference on digital libraries* (pp. 293–294). New York: Association for Computing Machinery. doi: 10.1145/2756406.2756928.

Tripathi, M., Shukla, A., & Sonkar, S. K. (2017). Research data management practices in university libraries: a study. *DESIDOC Journal of Library and Information Technology*, *37*(6), 417–424.

Vanz, S. A. d. S., Passos, P. C. S. J., Caregnato, S. E., Pavão, C. G., Borges, N., Rocha, R. P. d., ... Azambuja, L. A. B. (2018). *Acesso aberto a dados de pesquisa no brasil: práticas e percepções dos pesquisadores: relatório 2018.* Retrieved from http://hdl.handle.net/10183/185195.

How to cite this article (APA):
Resende, L. C. de & Bax, M. P. (2020). . *AtoZ: novas práticas em informação e conhecimento, 9*(1), 111 – 126. Retrieved from: http://dx.doi.org/10.5380/atoz.v9i1.69190