Long-term digital preservation supported by action plans
a systematic literature review

Luciana Gonçalves Silva Souza¹
Elisângela Cristina Aganette²

¹Universidade Fed. Minas Gerais, Belo Horizonte, Minas Gerais, Brasil / e-mail: lucianags9@gmail.com
²Universidade Fed. Minas Gerais, Belo Horizonte, Minas Gerais, Brasil / e-mail: elisangelaaaganette@gmail.com

ABSTRACT
The production of digital content is a usual drill practice in organizations. In order ensure reliability and access in long term to digital content, it is essential to implement digital preservation actions. Proper planning of preservation activities includes the development and institutionalization of a digital preservation policy and the implementation of guidelines established through the digital preservation plan, with the definition of procedures related to the maintenance and preservation of digital records for the long term. Thus, ones find out identify structures and experiences of digital preservation plan. Since, the method was used the systematic literature review. For the research, eight databases were selected, two specific from the Computer Science area, three from the Information Science area, one from the education area and two multidisciplinary databases. The searches were carried out on scientific articles evaluated by peers. The recovered contents include guidance on procedures related to the preparation of the digital preservation plan and its structural elements. The review showed the reduced exploitation of the term digital preservation plan and its associated contents in the scientific literature, an aspect that possibly indicates the recent use of the instrument as a guide for the development of digital preservation actions in institutions.

KEYWORDS

A preservação digital em longo prazo amparada por planos de ações
uma revisão sistemática de literatura

RESUMO
A produção de conteúdos digitais é prática cada vez mais frequente nas organizações. Para garantir a confiabilidade e o acesso futuro aos conteúdos digitais é essencial a implementação de ações de preservação digital. Um planejamento adequado das atividades de preservação inclui o desenvolvimento e institucionalização de uma política de preservação digital e a implementação das orientações estabelecidas por meio do plano de preservação digital, com a definição dos procedimentos relativos à manutenção e à preservação dos registros digitais por longo prazo. Assim, busca-se identificar estruturas e experiências de elaboração do plano de preservação digital. Para tal, apresenta-se como método a revisão sistemática de literatura. Para a pesquisa foram selecionadas oito bases de dados, entre elas duas específicas da área de Ciência da Computação, três da área de Ciência da Informação, uma da área da educação e duas de bases de dados multidisciplinares. As buscas foram realizadas em artigos científicos avaliados pelos pares. Os conteúdos recuperados incluem a orientação de procedimentos relativos à elaboração do plano de preservação digital e seus elementos estruturais. A revisão evidenciou a reduzida exploração do termo
plano de preservação digital e dos conteúdos a ele associados na literatura científica, aspecto que, possivelmente, indica a recente utilização do instrumento como orientador para o desenvolvimento das ações de preservação digital nas instituições.

PALAVRAS-CHAVE.

JITA: JH. Digital preservation
1 INTRODUCTION

The production of digital content is an increasingly frequent and profuse practice in organizations, and these often become the only documentary records of processes and activities performed in an institution and will need to be kept as a source of proof and consultation for the development of new actions or improvements in routine processes. To ensure reliability and future access to digital content, document management and implementation of digital preservation actions by the institution is essential (SANTOS; FLORES, 2015; SILVA, 2017; SILVA JUNIOR, 2017; SMITH, 2002).

However, implementing digital preservation actions is not an easy and routine task. It is necessary to plan and operate the actions to be successful (INNARELLI, 2012). Proper planning of digital preservation activities includes the development and institutionalization of a digital preservation policy, a document that includes the commitments assumed by the organization regarding the maintenance and preservation of its digital records for the long term. From this instrument it is possible to realize the planned actions (NOONAN, 2014).

Based on the guidelines established by the policy, the digital preservation plan is developed, whose elaboration and implementation is another important and indispensable step for the operationalization of digital preservation of documents. It is a guiding document composed of organizational and technical definitions, which should be unfolded into strategies and action plans.

From this perspective, many managers have realized the importance of thinking about digital preservation in their institutions and have begun to develop their policies and plans for digital preservation. This article aims to identify, through a systematic review of literature, experiences in the development of digital preservation plans. Its relevance is justified by synthesizing essential contents to guide the process of formulation and development of digital preservation actions, practices that are still excipient and that demand a planning supported by specific norms and guidelines.

Eight databases were selected to be available at CAPES's journal portal, three of which are from the Information Science area, two multidisciplinary databases, two databases with contents related to Computer Science and one from the Education area. The searches were carried out on scientific articles evaluated by peers who, after applying inclusion and exclusion criteria, returned seven documents that were analyzed.

This article was developed based on partial results obtained from ongoing doctoral research, with which it seeks to analyze actions to operationalize the preservation of digital archival documents, proposed by Brazilian Higher Education Institutions.

The other sections of the article are organized as follows: section two presents concepts related to digital preservation, including principles and fundamentals. Similarly, section three presents concepts related to the policy and plan of digital preservation and its importance to current institutions. Section four presents the systematic literature review, the method used in this paper. Section five includes the presentation and discussion of the results obtained, and section six presents the final considerations and perspectives for future work.

2 DIGITAL PRESERVATION

The preservation process includes care required to maintain materials for a specific or indefinite period. It was defined by Conway (2001, p. 14) as "the acquisition, organization and distribution of resources in order to prevent further deterioration or to renew the possibility of
using a select group of materials”. It covers the execution of techniques aimed at keeping contents accessible for certain periods of time.

The preservation of printed content is composed of procedures that aim to protect the media and information together, since there is no separation between the two aspects, while in digital preservation the preservation of "the content of the digital object and its authenticity, regardless of the media in which it is available", as noted by Grácio (2011, p. 68), because during the life cycle of the digital document it becomes necessary to transfer it from support, adapting it to existing technologies to ensure permanent access.

Duranti (2010) defines digital preservation as a set of principles, policies, rules and strategies aimed at prolonging the existence of the digital object, keeping it in suitable conditions for use, either in original format or in a more persistent format, ensuring the protection of the identity and integrity of the object, i.e. its authenticity (DURANTI, 2010, p. 157, our translation).

Thus, digital preservation includes processes to guarantee the continuity of materials for previously established periods, and accessibility to content is the purpose of this action. Preservation strategies should include access to authentic information (UNESCO, 2005).

Therefore, institutions holding digital objects should be attentive to the development of preservation of their contents. According to Grácio, Fadel and Valentim (2013, p. 113) the concept is linked to the "process of organizational management that encompasses several activities necessary to ensure that a digital object can be accessed, recovered and used in the future, from the ICTs existing at the time and with guarantees of authenticity". It thus involves a large part of the institution and demands the cooperation of its members.

However, this is a concern still under development, often understood only after losses of important content for the institution due to inadequate management and preservation, as Innarelli (2015) points out. Archivists, together with administrators and Information and Communication Technology (ICT) professionals, are responsible for good practices in digital preservation, as the author puts it: "Archivists of institutions, responsible for the management and preservation of conventional archival documents, should also assume responsibility for the management and preservation of the digital archival document" (INNARELLI, 2015. p. 30). The involvement of these professionals, aware of the importance of preserving digital information, will contribute to safeguarding digital content for long-term access.

Thus, for the preservation to be effective it is essential that the organization develops and institutionalizes formal documents that will guide its actions and establish its commitment to maintain digital content. These documents are the policy and plan of digital preservation, explored in the following section.

3 DIGITAL PRESERVATION POLICIES AND PLANS

The policy and the plan are the necessary instruments to plan and operationalize digital preservation in an institution. Politics is a term that can be used in different senses, having, however, a close relationship with the sense of power. Oliveira (2007) clarifies the concept of the object explored in this article, when he defines policies such as those that are "[...] derived from the objectives, challenges and goals of the company and are established by top management with the purpose of guiding subordinates in their decision-making process. They usually correspond to strategic or tactical issues". They establish the commitment of senior management with the outlined principle and guide decision making. For Interpares ([201-], p. 1) a policy is "a formal statement of direction or guidance on how an organization will fulfill
its mandate, functions or activities, motivated by certain interests or programs.

Thus, the policies applied in the context of digital preservation aim to formalize the commitment of the administration of institutions to continuously maintain their projects of management, preservation and access to digital resources for the long term. The constitution of this official document, besides institutionalizing the responsibility, allows the accomplishment of planned actions. Noonan (2014, p. 1, our translation) states that "To effectively carry out digital preservation activities, an institution or organization must have a digital preservation policy that articulates and institutionalizes its commitment to its preservation strategies and actions.

Thus, the definition of preservation policies can guarantee the consolidation of the institution's purposes and investments with the digital resources produced and that need to be preserved for the long term for future access. However, for its implantation, complementary documents should be considered, which operationalize the definitions contained there, such as the preservation plans, as Santos and Flores (2018, p. 43) state: "After defining a preservation policy, the strategies that will be part of the digital preservation plan should be chosen, always considering the related diplomatic requirements".

Bountouri, Gratz and Sanmartin (2018) conceptualize the Digital Preservation Plan (PPD) as a set of documented strategies for the preservation of collections. It is an instrument that makes it possible to describe how the institution fulfills its obligations regarding preservation actions, by explaining and documenting the vision and strategy of the preservation service performed. It includes the legal basis on which the institution relies to provide the digital preservation service and all the definitions that will make the implementation of the digital preservation policy accurate and complete. Thus, the preservation policy is an important instrument of the preservation plan, together with the procedures and standards.

The PPD comprises the detailing of projects and "defines a series of preservation actions to be taken by a responsible institution due to a risk identified for a certain set of digital objects or records (called collection)" (BECKER et al., 2009).

Silva and Flores (2018) clarify that PPD can be employed in any organization that produces digital documents and seeks a planning for the preservation of these contents, after identifying the need to use the information produced for long periods. It is understood that the process requires investment and technical knowledge, and, as the authors point out, it needs to be accompanied by document management practices, using instruments such as the Classification Plan of Archive Documents and the Table of Temporality and Destination of Documents.

Thus, the policy of digital preservation can ensure the consolidation of the purposes and investments of the institution in the digital resources produced and that need to be preserved for the long term for future access. However, it should be noted that for its implementation, documents that operate the definitions contained therein should be deployed in digital preservation plans.

In view of the above, the present article, through the application of the Systematic Literature Review (RSL) protocol, intended to identify experiences in the elaboration of the digital preservation plan and verify the publication of initiatives of implementation of the instrument in institutions producing digital content. In the following subsection the process will be detailed.
4 METHOD: RSL

The method used in this article is the integrative Systematic Literature Review (SLA), which uses elements of quantitative and qualitative research. The systematic literature review comprises a procedure of reference selection using explicit and replicable criteria, with the objective of minimizing the risk of bias, attributing scientific validity to the discussion produced (EVANS; PEARSON, 2001).

Biolchini et al (2005, p. 2) explain that the term "Systematic Review [...] is used as a specific research methodology to collect and evaluate existing evidence related to a delimited topic". The research is constructed from a central question that is represented by specific terms and concepts and worked on as a structured and previously defined question.

The methodological guidelines developed by the Ministry of Health (BRAZIL, 2014, p. 15) point out systematic reviews as:

[...] a type of secondary study that uses a comprehensive literature review process, in an impartial and reproducible manner, to locate, critically evaluate and synthesize the set of evidence available in the scientific literature to obtain an overview of a given research question. It differs from the narrative or traditional review in that it includes a detailed description of the methods and criteria used to select and evaluate the included articles.

For Moher et al (2009, p. 335) the RSL, as well as meta-analysis, can guide the constitution of practices, in addition to enabling the identification of the need for new scientific explorations. They define it as "a review of a clearly formulated question, which uses systematic and explicit methods to identify, select and critically evaluate relevant research, and collect and analyze data from these studies that are included in the review.

According to the Cochrane Institute, an entity constituted by a network of researchers dedicated to the elaboration of systematic reviews based on the methodology of scientific evidence developed, entitled Cochrane, the systematic review is established through seven stages:

a) formulation of the problem;
b) localization and selection of studies;
c) evaluation of the quality of studies;
d) data collection;
e) analysis and presentation of results;
f) interpretation of the results;
g) improvement and updating of the review (ALDERSON; GREEN; HIGGINS, 2004, p. 14).

In the context of this article, this RSL seeks to answer the following questions: "Have digital preservation plans been instruments of operationalization of digital preservation? What are the procedures for preparing the instrument and structure used?".

As a search term, the concept "digital preservation plan" was used. It was chosen the use of this compound term, included in quotation marks, for restriction and recovery of contents that approached exactly the expression. Besides this, searches were also performed with the translation of the concept and similar expressions used in literature in the two other languages established for the search: Spanish and English. Thus, the expression "plan de preservación digital", "digital preservation plan" and "digital preservation strategic plan" were used. We chose to include the expression "digital preservation strategic plan" in English, differently from the expressions in Spanish and Portuguese, due to the verification of its use during the initial stage of identification of the expression to be used in the RSL. Some English-speaking institutions have used the Odum Institute Data Archive concept (ODUM INSTITUTE, 2017),...
International Federation of Library Associations and Institutions (IFLA, 2016), or named their plan that way, such as York University (2013). Thus, it was verified that the expression could also be considered for the identification in searches by the digital preservation plan.

It should be noted here that most of the systematic reviews already published consider the use of the search expression in the English language. In this research, it was initially thought to adopt this practice, however, a test was carried out with the use of the expression only in English and with the inclusion of keywords in the three search languages. From the tests with the use of only the keyword in English there was a reduction in the total of documents recovered. Thus, we opted for the inclusion of the expression in the three languages, seeking greater coverage of content that dealt with the subject addressed.

As mentioned, eight databases were selected for the composition of the articles, arranged in Chart 1 below.

Chart 1. Selected databases and areas of knowledge covered

<table>
<thead>
<tr>
<th>Database</th>
<th>Area of knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM Digital Library</td>
<td>Computer Science</td>
</tr>
<tr>
<td>Educational Resources Information Center (ERIC)</td>
<td>Education</td>
</tr>
<tr>
<td>IEEE Xplore</td>
<td>Computer Science</td>
</tr>
<tr>
<td>Information Science &amp; Technology Abstracts (ISTA)</td>
<td>Computer Science</td>
</tr>
<tr>
<td>Library and Information Science Abstracts (LISA)</td>
<td>Computer Science</td>
</tr>
<tr>
<td>Library, Information Science &amp; Technology Abstracts (LIST)</td>
<td>Computer Science</td>
</tr>
<tr>
<td>Scopus</td>
<td>Multidisciplinary</td>
</tr>
<tr>
<td>Web of Science</td>
<td>Multidisciplinary</td>
</tr>
</tbody>
</table>

Source: the authors.

The bases used were defined through the selection of those that offer relevant contents to the area of Information Science and Computer Science, since they are databases that integrate scientific contents evaluated by experts, available in current and reference publications for the areas, such as articles from journals and annals of national and international events. Considering also the interdisciplinarity of the subject, multidisciplinary databases were included, which index a great part of the most relevant journals from several areas of current knowledge. Thus, the databases were selected from the consultation of works already developed (GONÇALVES; DELAMARO; NUNES, 2014; PEDRO; NUNES; MACHADO-LIMA, 2013).

The Capes Periodical Portal, a system developed by the Brazilian Ministry of Education, the Coordination for the Improvement of Higher Education Personnel (Capes), was used to carry out the searches, which includes "more than 45 thousand titles with full text, 130 reference bases, 12 bases dedicated exclusively to patents, in addition to books, encyclopedias and reference works, technical standards, statistics and audiovisual content. The access to the portal was made through the agreement between the service and the university where this research was carried out.

The procedure for selecting the documents was carried out from 24 to 28 April 2020, through the "advanced search" option available in all databases. Contents arranged in the title, abstract, keywords and subject terms (or indexed terms, according to the name of each database) were selected for retrieval. This selection was chosen in order to fulfill two objectives:

a) to standardize the search in all databases, considering that some do not offer the option of full text search;
b) to recover documents that probably deal with digital preservation plans, since the expression is included in the most representative metadata of the document's content.

The aim was to keep the same expression for all databases, but it was necessary to make changes to adapt the expressions to the pattern adopted by the search machines. To this end, the guidelines established by Abrantes and Travassos (2007, p. 6) were followed: "(1) the derived string should logically be equivalent to the original string, or (2) in the impossibility of maintaining exact equivalence, the derived string should be more comprehensive to avoid loss of potentially relevant documents".

After the search in each database, documents were generated in .RIS and .BIB formats, according to the availability of the application in each database. These formats were defined as standards developed to allow the exchange of bibliographic data between different systems. The .RIS and .BIB formats include their own structure for the presentation of references obtained from searches in the databases and are the formats read by the software used to support the realization of this RSL, the StArt (State of the Art through Systematic Reviews). It is a free platform developed by the Software Engineering Research Laboratory (LaPES) of the Federal University of São Carlos (UFScar). It is characterized as a tool to support the development of systematic literature reviews by offering support to the researcher, helping the application of the technique.

In this research StArt was used in the process of organization and selection of the documents, since the importation of the data obtained with the searches in the databases, because it is a tool that allows the classification of the items according to the criteria established in the revision protocol, performing the organization in the stages of selection and extraction of the contents. The software version 3.3 Beta was used.

The inclusion criteria used in the analysis stage of the articles were:

a) peer-reviewed scientific publications;
b) publications that characterize the digital preservation plan and detail the instrument construction process;
c) publications that explore the implementation of the digital preservation plan.

The exclusion criteria were based on the following items:

a) publications presented in poster format, lectures, workshop, abstracts, reviews;
b) publications that do not mention digital preservation plans in their content;
c) publications that only mention digital preservation plans, but do not conceptualize or detail their construction or implementation.

The technical reading, performed to select the final sample, included the analysis of the sections that represent in a more synthetic way the content of the documents, which, according to Lancaster (2003), are: title, abstract, synopsis, introduction and conclusions, besides section titles, underlined words and illustration subtitles. In addition, the content locator of the software used for reading (Adobe Acrobat, Word, internet browser) was used to identify the use of the expression "digital preservation plan" in the texts.

Finally, for the analysis and categorization of the studies, a table format synthesis matrix was developed that included the reference metadata, place of application of the practice, reported procedures and results obtained allowing an overview and interpretation of the data. The results achieved with the technique and the analysis will be described in the following section.
5 RESULTS

The first step of the selection process took place with the recovery of the full text of the documents. As already reported, eight databases were selected for the search, of these only three are characterized as databases offering the full text of the productions (ACM Digital Library, LISTA and IEEE Xplore), while the others are defined as reference databases with summaries (ERIC, ISTA, LISA, Scopus, Web of Science). Nevertheless, most of the texts were retrieved by searching the databases, even the referential ones. Only two documents were not found in searches in the databases and in Internet search engines, because the publishers and sources of access to the content were not available. Therefore, it was necessary to contact the authors by e-mail, and they kindly forwarded their complete text to continue the search.

Figure 1 explains the selection procedures used in this review.

Figure 1. Flowchart of the selection steps of the studies

Source: the authors, based on Marini's model; Lourenço; Barba (2017).

The database search selected a total of 40 documents. After listing them in the StArt software, 16 duplicate documents were excluded. The remaining 24 were analyzed regarding the first exclusion criterion: "Publications presented in the format of advertising reports, lectures, poster, reviews, summaries, workshop". It was chosen to include this exclusion criterion, as the publications that fit these categories do not offer full text for the extraction of the contents, besides some cases not having scientific evaluation. For this stage, access to the full text of all articles was sought. Most of them were retrieved in the databases, even some that, at first, were only available in the reference databases: ERIC, ISTA, LISA, Scopus and Web of Science, were recovered in their entirety. Only two documents were not located, and for access to the texts, a contact was made by e-mail with the authors, who forwarded the full text for RSL. Thus, it was possible to carry out this step by reading all the complete texts and
evaluating them according to the structure they present. Nine documents were rejected because they fit these categories.

For the following two stages, 15 documents were evaluated. In this process the full text was read, and both the second exclusion criterion: "Publications that do not mention digital preservation plans" and the third criterion: "Publications that only cite digital preservation plans, but do not conceptualize or detail their construction or implementation" could be evaluated. Of these, three texts did not address the subject sought, despite being recovered by the search criteria of the databases, and five documents only cited plans of digital preservation, without conceptualizing or detailing their construction or implementation.

Seven items remained, then, which were the ones analyzed in the content extraction stage. A second reading of the complete text, more detailed and considering the criteria for inclusion of the revision, was performed. For the organization of the information retrieved it was used the strategy of categorization by inductive analysis, when defining the categories from patterns identified in the literature used in the research. Based on the conditions established in this review, it can be affirmed that, although the documents have different objectives, their main subject is digital preservation and were selected for citing, detailing information and/or applying a digital preservation plan. The PPD was considered by the authors in different perspectives.

The following table presents an overview of the seven documents recovered, including the main references: author, title, year, document typology and method used in the study.

**Chart 2. Characterization of the analyzed documents**

<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
<th>Year</th>
<th>Documental typology</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Becker, Christoph; Kulovits, Hannes; Guttenbrunner, Mark; Strodl, Stephan; Rauber, Andreas; Hofman, Hans</td>
<td>Systematic planning for digital preservation: evaluating potential strategies and building preservation plans</td>
<td>2009</td>
<td>Journal article</td>
<td>Review article</td>
</tr>
<tr>
<td>Bishoff, Liz</td>
<td>Digital preservation plan: ensuring long term access and authenticity of digital collections</td>
<td>2010</td>
<td>Journal article</td>
<td>Scientific Informatio</td>
</tr>
<tr>
<td>Bountouri, Lina; Gratz, Patrick; Sanmartin, Fulgencio</td>
<td>Digital Preservation: How to Be Trustworthy</td>
<td>2018</td>
<td>Journal article</td>
<td>Case study</td>
</tr>
<tr>
<td>Breedstraet, Els</td>
<td>Preserving the European Union’s digital publications heritage: lessons learnt on our journey to making the past accessible for the future</td>
<td>2019</td>
<td>Event Article</td>
<td>Case study</td>
</tr>
<tr>
<td>Sáenz Giraldo, Andrés</td>
<td>La preservación digital en Colombia: un análisis desde la perspectiva normativa</td>
<td>2019</td>
<td>Journal article</td>
<td>Case study</td>
</tr>
<tr>
<td>Nascimento, Andréa Gonçalves do; Queiroz, Claudete Fernandes de; Araújo, Luciana Danielli de</td>
<td>Garantindo acervos para o futuro: plano de preservação digital para o Repositório Institucional Arca</td>
<td>2019</td>
<td>Journal article</td>
<td>Case study</td>
</tr>
<tr>
<td>Silva, William; Flores, Daniel</td>
<td>Política arquivística de preservação digital: um estudo sobre sua aplicabilidade em instituições públicas federais</td>
<td>2018</td>
<td>Journal article</td>
<td>Applied research</td>
</tr>
</tbody>
</table>

Source: revision data.
The following describes the characteristics identified, related to document identification aspects, such as place and date of publications, authorship and also comparative analyses related to the content addressed in the documents.

5.1 Characterization of the authors and the method used

Each of the authors wrote only one article, or worked together with other authors on only one document. It is pointed out, however, that this conclusion cannot be extended beyond the contents retrieved in this review, because, due to the inclusion criteria adopted in the search, it is possible that there are publications by these authors with similar contents that did not meet the definition of the review. Having said that, we have tried to identify them by nationality, as shown in Chart 3 below.

<table>
<thead>
<tr>
<th>Country</th>
<th>Qty. authors</th>
<th>Qty. articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Brazil</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Colombia</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>United States of America</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Source: review data.

It was found that Austria is the country with the largest number of authors, but it is a publication that includes all six professionals. Five authors are Brazilian, and are linked to the two articles published about the subject in the databases consulted. Four authors are from Luxemburg, and two documents regarding the country were analyzed. Colombia and the United States of America include one author each, who published the documents in single authorship.

Given these data, it can be stated that there was no concentration of publications in a specific region. The content has been explored in different locations, considering the American and European continents. With the analysis of the authors' affiliation it was also possible to notice that nine of them are linked to universities, three to a research institute, other three are linked to an inter-institutional publishing body and two to consulting and service companies in the area of digital preservation, as shown in Chart 4 below.

<table>
<thead>
<tr>
<th>Author</th>
<th>Connection</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Becker, C.</td>
<td>Vienna University of Technology, Vienna, Austria</td>
<td>Review article</td>
</tr>
<tr>
<td>Kulovits, H.</td>
<td>Vienna University of Technology, Vienna, Austria</td>
<td>Review article</td>
</tr>
<tr>
<td>Guttenbrunner, M.</td>
<td>Vienna University of Technology, Vienna, Austria</td>
<td>Review article</td>
</tr>
<tr>
<td>Strodl, S.</td>
<td>Vienna University of Technology, Vienna, Austria</td>
<td>Review article</td>
</tr>
<tr>
<td>Rauber, A.</td>
<td>Vienna University of Technology, Vienna, Austria</td>
<td>Review article</td>
</tr>
<tr>
<td>Hofman, H.</td>
<td>Vienna University of Technology, Vienna, Austria</td>
<td>Review article</td>
</tr>
<tr>
<td>Bountouri, L.</td>
<td>Publications Office of the European Union, Luxembourg</td>
<td>Case study</td>
</tr>
</tbody>
</table>
Relating, also, the data regarding the authors' affiliation and the method used to produce the documents, it can be seen that the predominance of authors linked to the universities coincides with the contents of review and applied research (eight of the nine authors), while the case studies were the most frequent typology written by authors from the research institution, publishing body and consulting firms (seven of the eight authors).

5.2 Characterization of studies by date and method used

The relationship between the date of publication and the method used to conduct the studies is shown in Chart 5 below.

**Chart 5.** Authorship by year of publication and method used

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Becker, C. et al.</td>
<td>2009</td>
<td>Review article</td>
</tr>
<tr>
<td>Bishoff, L.</td>
<td>2010</td>
<td>Scientific Information</td>
</tr>
<tr>
<td>Bountouri, L.</td>
<td>2018</td>
<td>Case study</td>
</tr>
<tr>
<td>Breedstraet, E.</td>
<td>2019</td>
<td>Case study</td>
</tr>
<tr>
<td>Sáenz-Giraldo, A.</td>
<td>2019</td>
<td>Case study</td>
</tr>
<tr>
<td>Nascimento, A.; Queiroz, C.; Araújo, L.</td>
<td>2019</td>
<td>Case study</td>
</tr>
<tr>
<td>Silva, W.; Flores, D.</td>
<td>2018</td>
<td>Applied research</td>
</tr>
</tbody>
</table>

Source: revision data.
It is possible to see that the review article (BECKER et al., 2009) and the scientific report (BISHOFF, 2010) are, among the documents recovered, those that were published first. This last typology was used here due to the fact that it is a referential document with concepts and experiences reported on the subject, but without referring to theoretical references. The researches that include practical application of the digital preservation plan are the most current: Silva and Flores (2018), who present a research about the implementation of digital preservation actions in federal public institutions in Brazil, considered in this thesis as an applied research, and Bountouri, Gratz and Sanmartin (2018); Breedstraet (2019); Sáenz-Giraldo (2019) and Nascimento, Queiroz and Araújo (2019), who present case studies.

Considering the actuality of the subject and its little exploration in the world literature, it can be inferred that the constructions and practical applications of the digital preservation plan started in the last years. The low rate of review articles on the subject is perceived by the authors as indicative of the little exploration of the subject in professional and academic environments, which are dedicated to expanding knowledge about PPD.

5.3 Characterization of studies by document typology

The vast majority of the documents, six out of seven, were published as journal articles. However, it is believed that this is not a data to be considered regarding a possible preference of authors over the chosen form of communication, but rather a feature regarding the choice of sources used for the search of contents, since they are databases that index a large number of scientific journals. Besides these, an event article was selected.

5.4 Summary of the contents treated in the studies

The following is a brief summary of the contents presented in the documents analyzed.

The article written by Becker and collaborators (2009) presents aspects that should be considered during the decision making process in the planning of digital preservation, such as choice of objects, compliance and application of the OAIS model in institutional repositories, definition and implementation of the preservation plan. The document describes and details a method for monitoring and updating the preservation plans and includes experiences from three case studies conducted with the objective of identifying the appropriate strategy for conducting digital preservation in digitized image collections.

Bishoff (2010) describes, in his article, conceptual and descriptive aspects related to the digital preservation plan. The document includes objectives and components of the PPD and details each content to be included, according to the needs of the institution and the collection to be preserved. It cites examples and indicates documents already produced by institutions, as a reference point for the implementation of the practice.

Bountouri, Gratz and Sanmartin (2018) and Breedstraet (2019) present cases of the Publications Office of the European Union, an institution that includes a wide collection of digital content produced within the European Union, such as official journals, international agreements, treaties, scientific publications, websites and data sets.

Bountouri, Gratz and Sanmartin (2018) analyze preservation aspects for application in their digital collections, including the digital preservation plan, product consumers, representation information, fix policy, preservation metadata. Breedstraet (2019) presents the experience of structuring the digital preservation plan in the context of the Publications Office, describing the process of preparing the document, the difficulties faced and the challenges that
had to be faced in the process of achieving digital preservation of content.

In this perspective, Sáenz-Giraldo (2019) reports the effort to regulate the actions of digital preservation in public institutions in Colombia. For this purpose, the Ministry of Culture and the General Archives of the Nation defined norms and procedures to be adopted by the archival institutions of the country, which included the implementation of integrated systems, metadata structures, policies and plans for digital preservation and reliable digital repositories. The author analyzes the norms and points out reflections on the lack of support to the institutions by the General Archives of the Nation regarding the inclusion of the adoption of accepted standards, such as the OAIS Model and ISO norms, such as NTC-ISO 16363: 2017, in the regulatory context of the country.

Nascimento, Queiroz and Araújo (2019) present the experience of the Arca - Institutional Repository of Fiocruz, a Brazilian scientific research institution, in the preparation of its Digital Preservation Action Plan. The document details the stages of the plan's construction, including the diagnoses made to adapt the service to the practice of digital preservation and the strategies of action to be adopted to maintain the guarantee of authenticity and integrity of the contents stored on the Ark.

Finally, Santos and Flores (2018) explain the results of a survey conducted with Brazilian federal public institutions regarding the acceptance and implementation of digital preservation policies and plans. The findings indicate that most of the country's institutions still do not have policies and plans implemented, which indicates the need for the researched agencies to adopt the recommendations of the National Council of Archives, which, in turn, lacks awareness actions and dissemination of the regulations and guidelines produced.

Thus, in view of the explorations presented, an analysis of the common aspects was made, which can help in the process of building and understanding the digital preservation plan, described in the following subsection.

5.5 Comparative analysis of selected documents

In this section we sought to synthesize the content covered in the articles and identify common aspects that contribute to the understanding and structuring of digital preservation plans.

To this end, three categories were defined:
- a) Objectives of the digital preservation plan;
- b) Structure of the digital preservation plan;
- c) Methodology for creating the digital preservation plan.

5.5.1 Objectives of the digital preservation plan

A common content of the articles is the declaration of the objectives of a digital preservation plan: it was made explicit by six of the seven documents analyzed. For better identification, the concepts were summarized in Chart 6 below.

<table>
<thead>
<tr>
<th>Document</th>
<th>Declaration of the objective of the digital preservation plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Becker et al. (2009)</td>
<td>It specifies an action plan to preserve a specific set of objects for a certain purpose.</td>
</tr>
</tbody>
</table>
It will provide the mission, specific goals and objectives and policies and procedures. It will define preservation strategies, standards, digital content depositors, staff, funding, roles and responsibilities and users.

It defines the legal basis on which the PB is based to provide the digital preservation service, as well as all important definitions that will make the implementation of the digital preservation policy accurate and complete.

It details the processes, methodologies and tools applied and documents the scope of the repository in terms of the content covered. It also covers governance, financing, service level and quality expectations.

They aim to implement programs, strategies, processes and procedures, aiming to guarantee the long-term preservation of electronic archive documents, maintaining their characteristics of authenticity, integrity, confidentiality, inalterability, reliability, interpretation, understanding and availability over time.

Systematize the steps and procedures necessary to meet the minimum requirements for the preservation of digital objects deposited in the institutional repository.

Through the objective statements one can perceive similarities between the actions assigned to the PPD: the verbs systematize, detail, define, specify, implement and provide, which express the idea of action, execution, as is the proposal of the document.

The definitions of all authors in relation to the objective of the preservation plan converge, to a greater or lesser degree, when they understand that the preservation plan details procedures and methods for implementing the actions of preservation of digital objects. They include, among the objectives of a preservation plan, the definition of procedures, processes, strategies, actions. Two of them cite the repository as a detailed object of the PPD (BREEDSTRAET, 2019; BIRT; CHEESE; ARAUJO, 2019). One of the documents explains that the PPD is the document that promotes the implementation of the preservation policy (BOUNTOURI; GRATZ; SANMARTIN, 2018).

In view of the objectives presented, it is possible to state that the preservation plan is intended to outline and formalize processes, procedures and strategies for digital preservation, based on requirements established in policies and administrative acts that express the values of the implementing organization.

5.5.2 Structure of the digital preservation plan

This subcategory aimed to identify the structuring elements of digital preservation plans, according to experiences and/or references cited in the analyzed documents. Five works included information in this respect. Thus, in each document we selected the excerpts that made reference to the content contained in the plans.

For better identification, a grouping of these concepts was carried out using, as reference for the nomination of the main categories, two structures of preservation plans already published: The University of British Columbia Library, Wheaton College Library and Archives and Barbedo (2019). Seven groups and 35 topics mentioned in the analyzed articles were identified, which may integrate the sections of the digital preservation plan or allow the delimitation of those that best fit the needs of the institutions at the time of construction of the PPD.

Chart 7 presents the summary of topics grouped by categories or possible sections of the digital preservation plan.
With the identification and grouping of topics it is expected to provide elements that can integrate the institutional digital preservation plan, from the proposals and experiences defined in the analyzed documents.

The Introduction includes the presentation of the institution and some essential concepts about digital preservation, necessary for understanding the presented document. As topics included in this category are:

a) foundation of digital preservation (BISHOFF, 2010);

b) institutional setting (BECKER et al., 2009; NASCIMENTO; QUEIROZ; ARAÚJO, 2019);

c) organizational commitment (BISHOFF, 2010);

d) financial commitment (BISHOFF, 2010).

The topic Contextualization includes elements that present the institutional context and were the driving force behind the proposal to develop the digital preservation plan. They are:

a) current state of the repository (BIRTH; CHEESE; ARAUJO, 2019);

b) mission of the institutional repository, involves the mandate, the purpose of the institution with its implementation (BECKER et al., 2009);

c) information on representation and provenance (BREEDSTRAET, 2019);

<table>
<thead>
<tr>
<th>Section</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>Justification of digital preservation</td>
</tr>
<tr>
<td></td>
<td>Institutional scenario</td>
</tr>
<tr>
<td></td>
<td>Organizational Commitment</td>
</tr>
<tr>
<td></td>
<td>Financial Commitment</td>
</tr>
<tr>
<td>Contextualization</td>
<td>Current status of the repository</td>
</tr>
<tr>
<td></td>
<td>Mission of the institutional repository</td>
</tr>
<tr>
<td></td>
<td>Information on representation and provenance</td>
</tr>
<tr>
<td></td>
<td>Limitations and restrictions: - legal - financial - technical</td>
</tr>
<tr>
<td></td>
<td>Plan status and change alert</td>
</tr>
<tr>
<td>Preservation actions</td>
<td>General methodology of creation and use</td>
</tr>
<tr>
<td></td>
<td>Preserved collections</td>
</tr>
<tr>
<td></td>
<td>Procedures and relevant flows</td>
</tr>
<tr>
<td></td>
<td>Costs</td>
</tr>
<tr>
<td></td>
<td>Representation and provenance</td>
</tr>
<tr>
<td></td>
<td>Mechanisms of cooperation</td>
</tr>
<tr>
<td></td>
<td>Permanent articulation and coordination between technology areas</td>
</tr>
<tr>
<td>Resources and preservation</td>
<td>Strategies against obsolescence / Strategies for preservation</td>
</tr>
<tr>
<td>strategies</td>
<td>Formats</td>
</tr>
<tr>
<td></td>
<td>Creation of metadata</td>
</tr>
<tr>
<td>Rules and responsibilities</td>
<td>Standards for information management</td>
</tr>
<tr>
<td></td>
<td>Legal, operational and preservation policies</td>
</tr>
<tr>
<td></td>
<td>Roles and responsibilities</td>
</tr>
<tr>
<td></td>
<td>Legal obligations</td>
</tr>
<tr>
<td></td>
<td>Preservation rights</td>
</tr>
<tr>
<td>Risks</td>
<td>Risk and threat assessment</td>
</tr>
<tr>
<td></td>
<td>Risk Management</td>
</tr>
<tr>
<td>Orientation, control and updating</td>
<td>Contract maintenance and access</td>
</tr>
<tr>
<td></td>
<td>User needs / Training and education</td>
</tr>
<tr>
<td></td>
<td>Quality control</td>
</tr>
<tr>
<td></td>
<td>Good Practices</td>
</tr>
<tr>
<td></td>
<td>Monitoring and review</td>
</tr>
</tbody>
</table>

Source: revision data.
e) plan status and change alert (BECKER et al., 2009).

The actions in preservation are composed of elements that structure the procedures to be adopted by the institution for the realization of digital preservation. They include:

- a) general methodology of creation and use (SÁENZ-GIRALDO, 2019);
- b) preserved collections (BECKER et al., 2009; BREEDSTRAET, 2019; NASCIMENTO; QUEIROZ; ARAÚJO, 2019);
- c) procedures and relevant flows (BECKER et al., 2009; BREEDSTRAET, 2019);
- d) costs (BECKER et al., 2009; SÁENZ-GIRALDO, 2019; BISHOFF, 2010);
- e) representation and provenance (BREEDSTRAET, 2019);
- f) cooperation mechanisms (SÁENZ-GIRALDO, 2019);
- g) articulation and permanent coordination between technology areas (SÁENZ-GIRALDO, 2019).

In the category Resources and preservation strategies are grouped the topics related to the technological resources adopted by the institution in the construction and development of the plan, the formats and the preservation strategies used. They are the following:

- a) preservation strategies (BECKER et al., 2009; NASCIMENTO; QUEIROZ; ARAÚJO, 2019);
- b) file formats (BREEDSTRAET, 2019);
- c) metadata creation (BISHOFF, 2010).

Among the rules and responsibilities are topics that include legal and regulatory aspects related to the digital preservation process, which are:

- a) rules for information management (SÁENZ-GIRALDO, 2019);
- b) legal, operational and preservation policies (BECKER et al., 2009; SÁENZ-GIRALDO, 2019);
- c) roles and responsibilities (BECKER et al., 2009; BISHOFF, 2010);
- d) legal obligations (SÁENZ-GIRALDO, 2019);
- e) preservation rights (BECKER et al., 2009).

5.5.3 Methodology for creating the digital preservation plan

Three of the documents analyzed described the PPD creation methodology. The proposals of Nascimento, Queiroz and Araújo (2019) and Silva and Flores (2018) were based on the evaluation of the preservation system available at the institution to characterize the plan. The first proposal is the result of the experience of building a preservation plan. It included, initially, the verification of national and international regulations on the actions of digital preservation and the consultation of instructional literature on the process of plan elaboration. Based on this understanding, an initial structure of the document was elaborated, considering essential elements and desirable items, as established in the norms. It was then started to verify the situation of the information system used in relation to the requirements established in the norms and procedures of preservation, carrying out diagnoses that allowed the definition of recommendations for the adequacy of the system to the practice of digital preservation. Thus, based on the definition of the essential elements and the diagnoses made, the preservation strategies for the system used were defined, described in the digital preservation plan.

Silva and Flores (2018) presented a methodology based on the authors Barbedo, Corujo and Sant’ Ana (2011). Thus, for a better understanding of the steps, the cited document was used. The methodology presented initially includes the structuring of the proposal, with the presentation of the phases, responsible for preparing the plan and the references and instructions.
to be used for the following phases. The second step includes selecting, among the existing information systems, those that will be included in the preservation plan. Following the proposal, the third stage includes the system's archival evaluation, establishing deadlines and destinations for the objects contained. Thus, through the selections, the contents that need a preservation plan will be defined. In the following phase a description is made of the characteristics of the systems that were selected to receive preservation plans. Through the definitions of the system it becomes possible to identify the most appropriate preservation strategy, which is the last step of the process and leads to the elaboration of the digital preservation plan.

The third proposal of the preservation plan construction methodology, by Breedstraet (2019), is part of a practical experience, as well as the first one exposed previously. The elaboration of the plan started from an initial proposal elaborated with the orientation of an external consultant and the proposal was dismembered among the members of the institution's team, seeking to attribute to each professional contents that were more similar to their area of work. The next step included discussing the document in meetings with specialists in digital preservation and long-term archiving. Following the discussions, the proposal was approved. From the meetings held there was the refinement and enrichment of the proposal, which, after approval by the institution's top management, the Management Committee, resulted in the first edition of the preservation plan.

**Chart 8. Methodologies for building the digital preservation plan**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Verification of national and international regulations on digital preservation actions;</td>
<td>1 - proposal structuring: presentation of the phases, responsible for preparing the plan and the references and instructions to be used for the following phases;</td>
<td>1 - elaboration of the initial proposal, based on the orientation of an external consultant</td>
</tr>
<tr>
<td>2 - Consultation of instruction literature on the plan elaboration process;</td>
<td>2 - selecting, among the existing information systems, those that will be included in the preservation plan</td>
<td>2 - dismemberment of the proposal among the members of the institution's team for analysis, seeking to assign to each professional content that most resembled their area of expertise</td>
</tr>
<tr>
<td>3 - Elaboration of the initial structure of the document;</td>
<td>3 - archival evaluation of the system, establishing deadlines and destinations for the objects contained</td>
<td>3 - discussion of the document in meetings with experts in digital preservation and long-term archiving</td>
</tr>
<tr>
<td>4 - Verification of the status of the information system used in relation to the requirements established in the norms and procedures of preservation</td>
<td>4 - archive evaluation of the system, establishing deadlines and destinations for the objects contained</td>
<td>4 - approval of the proposal by the experts</td>
</tr>
<tr>
<td>5 - diagnosis for system adequacy to the practice of digital preservation</td>
<td>5 - identify the most appropriate preservation strategy</td>
<td>5 - refinement and enrichment of the proposal</td>
</tr>
<tr>
<td>6 - definition of the preservation strategies for the system used</td>
<td>6 - preparation of the digital preservation plan</td>
<td>6 - approval of the proposal by the institution's top management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 - first edition of the preservation plan</td>
</tr>
</tbody>
</table>

Source: revision data.

The three methodologies presented highlight important aspects in the construction of digital preservation plans:

a) The initial consultation to local, national and international regulations on digital preservation, with the purpose of ensuring an alignment of the actions to be developed;

b) The performance of a diagnosis from the mapping of activities related to the subject
under development in the institution for a better understanding of the needs and real possibilities of development of actions in preservation;

c) The elaboration of an initial version of the digital preservation plan, an action that can be advised by specialists in digital preservation external to the institution. However, after the conclusion of this stage, it becomes necessary to present and discuss the document among the members of the institution, if possible experts in different areas of knowledge, for improvement and adaptation to local reality;

d) Based on the knowledge acquired about the institution and the discussions held, the digital preservation plan is finalized.

With the realization of this RSL it was possible to understand, from the questions that raised the application of the technique, that the plan of digital preservation has been constituted as an element of operationalization of digital preservation for the institutions that have already started the movement of implementing the procedure. Much linked to the instrumentalization of the process, it may even be referred to the organization's digital repositories. The other question raised, regarding the procedures for structuring the instrument, can be fully answered by the review, because the content recovered presented essential elements and, in some cases already employed by the institutions, related to the preparation process and the basic structure of the document.

In addition, the result of the present literature review evidenced the reduced exploration of the term digital preservation plan and the contents associated with it in the scientific literature: the search for expression in eight databases returned the quantitative of 40 documents, and these, after eliminating the duplicates, resulted in 24 different documents. It can be said that this is evidence of the gap in the literature on the subject and, possibly, indicates the recent use of the instrument as a guide for the development of digital preservation actions in institutions, despite the growing production of digital content.

However, it is important to emphasize that the low rate of recovery of contents that deal with digital preservation plans cannot be defined as the absence of use and application of the instrument, considering, mainly, that:

a) Because it is a tool for the operationalization of digital preservation practices, companies and professionals may choose not to report their experiences in scientific works, not publishing or circulating them only in the internal context of the organization;

b) Practices that were not successful in their realization, in great part of the times, are not reported, nor even come to the knowledge of all the members of the organization, and normally are not described in scientific format;

c) We tried to include in the review reference databases for the area of information science and computing, but due to the vast number of publications available worldwide, there is a large quantity of publications that were not consulted in this research, and which may contain relevant studies on the subject investigated.

Among the documents recovered, it was verified that they include essential and guiding elements related to the digital preservation plan, presenting experiences in the construction of the instrument, related concepts and indicating adequate practices for its development. Thus, it is expected to contribute to the systematization of guidelines on the characterization and structuring of the PPD. However, it is emphasized that for the development of the digital preservation plan, it is essential that the institution has a policy of digital preservation approved by the administration, and seeks, from the PPD, systematize the strategies established, following the commitment assumed to perform the preservation of digital content that is produced.

It can be said that the practice of long-term preservation is still a recent activity, which
requires investment and institutional support. Many institutions are initiating their actions and seek support and guidance for this. It is believed that the recovered documents provide important subsidies for the understanding of the processes and can be fully consulted by those interested in developing digital preservation policies and actions. In view of this, it is presented as suggestions for future work the application of the RSL technique for the identification of specific processes of preservation of digital contents, such as images, videos, research data or other objects.

REFERENCES


