

# TRUSTED DIGITAL REPOSITORIES FOR DOCUMENT PRESERVATION: A TOOLS' EVALUATION

Ana Oliveira<sup>1</sup>, Luís Moura<sup>1</sup>, Ana Lúcia Terra<sup>2</sup>

<sup>1</sup>Porto Accounting and Business School, Polytechnic Institute of Porto (PORTUGAL)

<sup>2</sup>University of Coimbra (PORTUGAL)

## Abstract

Grounded in a review of the preservation concept in trusted digital repositories, this paper aims to address how different trusted digital repositories operate while in the process of preserving digital objects. For this purpose, a literature review and an evaluation of trusted digital repositories' tools were carried out. A research was first conducted on several concepts related to this topic, namely digital preservation and its policies, reliable digital repositories and their rules. A search on theoretical background and a survey of available trusted digital repositories were performed. A set of tools was identified and then evaluated, and for that purpose several steps were outlined. First, in the criterion identification phase, the several criteria that a reliable digital repository must meet were identified, and then classified as mandatory or complementary. Then, functionalities of different software were highlighted and a comparison was made between them, resulting in a comparative table that led to the choice of DSpace software. This software is the most complete tool, and accomplishes the most important functionalities of a trusted digital repository. Furthermore, it is one of the most used software worldwide for trusted digital repositories, which grants it a high level of reliability.

**Keywords:** Digital preservation, trusted digital repositories, information preservation

## INTRODUCTION

This paper is an academic output of the curricular unit of Information Services and Resources Project, from the 6<sup>th</sup> semester of the degree of Library and Information Sciences and Technologies (LIST) at the Polytechnic Institute of Porto, Porto Accounting and Business School. In this course the development of a project was proposed to students, and from that same project, the elaboration of a scientific paper. All steps and activities performed along the project are explained in the project blog (<https://repositoriosdigitaisconfiaveis1920.car.blog/>).

Firstly, in this paper, the theme of the project will be presented, followed by a literature review that addresses several topics such as the concept of digital preservation and its policies, trusted digital repositories, standards related to those repositories and tools available to create trusted digital repositories. This literature review allowed to identify the criteria that digital repositories should meet, and then choosing in which software trusted digital repositories should be developed in.

Digital preservation is a process that has increasingly gained importance and relevance, as it allows, through different strategies, preserving digital objects.

Thus digital repositories are imperative to allow a process of long term digital preservation. In order to elaborate this project and write the scientific article, the methods used were information search about repositories, as well as the study of real cases of software application.

## METHODOLOGY

As a methodology for this project, information searches of relevant scientific literature were performed in reliable sources such as Google Scholar and Portuguese Open Access Scientific Repositories (RCAAP). These search results allowed the literature review, as well as the process of information gathering in order to elaborate a comparison between different trusted digital repositories software. The aim was to identify criteria that digital repositories should meet in order to become reliable and analyse these criteria by comparing them between tools that have the ability to create a reliable digital repository.

## LITERATURE REVIEW

For the development of the project, it was necessary to firstly elaborate a literature review, including several concepts with the purpose of gaining a better understanding of the two main topics of the project: digital preservation and repositories for digital preservation. Thus, as a first phase of the project, a research was done on the concept of digital preservation and its policies. Regarding the digital repositories, research was also carried out on their concept, as well as the standards and tools available to create trusted digital repositories.

According to Arellano (2017, p. 3), digital preservation is defined as a “series of strategies taken to promote the availability and usability of digital information over time”.

Digital preservation is seen as something necessary so that the information contained in digital objects can be read and interpreted throughout the years, without it being modified. Therefore, digital preservation’s mission is to ensure continuous access to authentic digital objects. The biggest challenge of digital preservation, when acquiring information, is to ensure its reliability, integrity, originality and access (Arakaki, et al., 2019, p. 142). According to CONARQ (2014), digital preservation must guarantee long term access to authentic archival documents, which means that certain principles must be adopted.

Digital preservation consists of two types of procedures: structural, where issues such as infrastructure, rules, choice of standards, financing and definition of metadata are defined; and operational, that are activities that aim at the physical, logical and intellectual preservation of digital documents, that is, the strategies for the preservation of documents.

Regarding digital preservation policies, Lyman and Bresser (2010, p. 337) claim that long-term digital preservation does not only require technical solutions and organizational strategies, it also needs the formation of a new mentality that values the survival of bits over time, and the initial step would be the development of digital preservation policies.

A policy is a set of rules and/or principles that guide all decisions and actions that are taken in order to achieve the results that are aimed to certain aspects or objectives (Júnior & Borges, 2015, p. 2), which must be:

- Imprescriptible;
- Technologically neutral;
- Provide support to the governance structure and organizational culture;
- Have revision and/or update cycles periodically or motivated by private actions.

According to Júnior and Mota (2012, p. 53), in public institutions, digital preservation policies aim to implement actions to preserve digital collections, so that permanent access is maintained under the bastion of the information access law.

Regarding trusted digital repositories, according to CONARQ (2014, p. 9), a digital repository is an environment to store and manage digital objects. It is an environment in which digital objects are captured, stored, preserved, through different approaches, and later accessed,

appealing to the support of the management of digital objects. They are formed by hardware, software and metadata elements, requiring an organizational infrastructure, rules for their management and various technical procedures. Despite being one of the characteristics of digital repositories, they are not presented, nor should they be seen as a computer solution for a lack of storage. The big question lies in the trusted digital repositories' reliability, and this is when the concept of Trusted Digital Repository develops. It is necessary to define this concept, in order to be able to understand how these repositories are capable of preserving information, maintaining their reliability characteristics, that is, defining how we can achieve it. It is important to clarify that reliability is the characteristic that confirms the credibility of something (CONARQ, 2014, p. 6).

Following information searches, several tools for the development of trusted digital repositories were found, with emphasis on the following: DSpace, Archivematica, EPrints, Fedora and Lockss [Table 1].

EPrints is a software developed by the University of Southampton (United Kingdom) that contains several tools for the development of repositories aimed at "literature research, scientific databases, theses, reports and multimedia" (Martins, Rodrigues & Nunes, 2008, p. 3). It shares several characteristics with document management systems, but is primarily used in institutional repositories and scientific journals.

Fedora, according to Martins, Rodrigues and Nunes (2008, p. 3), is a proposal for a repository system, developed jointly between Cornell's Information Science University and University of Virginia's Library (USA). It aims to provide free access repository software, as well as services that can serve as a start for the development of different types of information management systems.

Lockss, which stands for "Lots of Copies Keeps Stuff Safe", and that according to its official website, is a program based on Stanford University (USA) that provides open source services and technologies for safe, reliable and resistant digital preservation. This program works with its users in order to establish preservation networks based on their technologies.

Regarding to DSpace and according to Lampert (2016, p. 144), this open source software is characterized as being a solution normally used for the preservation of scientific and cultural-institutional production, fostering the dissemination of digital content produced for research. It is thus directed to access, dissemination of scientific and academic communication, since it contains a structure of metadata and submission of digital objects identical to the process of editing a journal.

Lastly, Archivematica, according to the official software website, is a web based open source software, with standards that allow its users to preserve access for long term to digital content. Its software is based in the metadata standards Metadata Encoding Transmission Protocol (METS), Preservation Metadata: Implementation Strategies (PREMIS), in Dublin Core, in the BagIt specification of the Library of Congress, as well as in other recognized standards for reproducing Archival Information Packages (AIP's).

*Table 1. Websites of tools for trusted digital repositories*

<b>Tools</b>	<b>Website</b>
EPrints	<a href="https://www.eprints.org/uk/">https://www.eprints.org/uk/</a>
Fedora	<a href="https://getfedora.org/">https://getfedora.org/</a>
Lockss	<a href="https://www.lockss.org/">https://www.lockss.org/</a>
DSpace	<a href="https://duraspace.org/dspace/">https://duraspace.org/dspace/</a>
Archivematica	<a href="https://www.archivematica.org/pt-br/">https://www.archivematica.org/pt-br/</a>

It is also relevant to mention the rules related to trusted digital repositories. It is the duty of a digital repository to manage documents and metadata in accordance with archival practices and standards, specifically related to document management, multilevel archival description and preservation. Thus, it is also possible to enable levels of interoperability with other digital repositories and computerized systems that deal with archival documents.

According to CONARQ (2014, p. 3), there are eleven standards and reference standards that digital repositories should meet:

- OAIS Reference Model;
- Report by the Research Library Group (RLG) and Online Computer Library Center (OCLC) - Reliable digital repositories: attributes and responsibilities;
- Trusted Repository Archiving Checklist - TRAC;
- Technical requirements for auditing and certification entities of candidate organizations to be trusted digital repositories;
- Preservation metadata - PREMIS;
- ISAD (G): General International Standard Archival Description;
- Metadata for Interoperability - MIP;
- Open Archives Initiative Protocol for Metadata Harvesting - OAI-PMH;
- Metadata encoding and transmission standard - METS;
  - Encoded Archival Description - EAD.

## **CRITERIA IDENTIFICATION**

As soon as the literature review process is over, it is time of identifying the criteria that a digital repository should have, classifying them as mandatory and complementary.

A trusted digital repository should attend several mandatory requirements. According to CONARQ (2014, p. 9), a trusted digital repository must store and manage documents according to its phases:

- Manage documents and metadata according to archival practices and standards, multilevel archival description and preservation (documents' management);
- Protect the characteristics of the archived document (Log in; access rights to content; back up and restoration);

A trusted digital repository should also attend to several mandatory responsibilities, related to the content of information packages. Therefore, a trusted digital repository in accordance with the OAIS model should be able to:

- Control information to enable long-term preservation (long-term information preservation);
- Ensure that the community is able to understand the preserved content information (simple architecture; availability in multiple languages);
- Follow previously documented policies and procedures (use of digital preservation strategies; OAIS / ISO Standard);
- Make preserved content information available to the designed community and allow to disseminate them (dissemination of information).

To be even more complete, a trusted digital repository should also accomplish other requirements which, although not yet proved to be as important as those already mentioned, support a trusted digital repository to become more effective and efficient. These requirements are:

- Be responsible for digital materials' maintenance (import and export data; data editing);
- Have an organizational structure that supports the long-term viability of the repositories and digital materials themselves (reliable organizational structure; interoperability);
- Demonstrate economic sustainability and administrative transparency (sustainable);
- Establish methodologies for evaluating systems that consider the reliability expectations of the community (reliability assessment methodologies).

Mandatory responsibilities are related to the acquisition of digital archival documents (information content), trying to safeguard the rights to carry out their preservation in the long-term. These will guide what will be preserved; will define the right of preserving; will keep documents procedures about preservation; and will define who will access and use this material. OAIS model understands the need to ensure possession of digital documents, as well as to know in advance the existence of a community potentially interested in preservation and access to information.

## COMPARISON BETWEEN FUNCTIONALITIES AND SOFTWARE

In order to select the software to be used in the project, it was essential to make a comparison between all the previously mentioned functionalities and criteria. Thus, a comparative table with the five chosen tools and respective functionalities was prepared, as well as a more detailed comparative analysis [Table 2].

	DSpace	Archivematica	Eprints	Fedora	Lockss
Documents Management	✓	✓	✓	✓	✓
Use of digital preservation strategies	✓	✓	✓	✓	✓
Log in	✓	✓	✓	✓	✓
Long-term information preservation	✓	✓	✓	✓	✓
Simple architecture	✓		✓		
Dissemination of information	✓	✓			
Import and export of data	✓	✓	✓	✓	✓
Data editing	✓	✓		✓	
Backup e restoration	✓	✓	✓		✓
Sustainability	✓				
Reliability assessment methodologies	✓				
Ease of installation		✓	✓	✓	✓
Effective search	✓	✓	✓		✓
OAIS / ISO Standard	✓	✓		✓	✓
Interoperability	✓	✓	✓	✓	✓
Availability in multiple languages	✓		✓	✓	✓
Viable organizational structure	✓	✓	✓	✓	✓
Access rights to content	✓	✓	✓		✓

*Figure 2. Comparison table of trusted digital repository tools*

As a digital repository, document management is in fact considered as one of the mandatory requirements and as such is present in the five chosen tools.

Together with the aforementioned functionality, there are six more features that are also present in all chosen tools:

- Use of digital preservation strategies, an extremely important functionality, because these repositories should work along with archival practices and standards, more precisely with documents' management standards;
- Need for Log in, another functionality with extreme importance that allows to control access, and provides usage restrictions for different types of users;
- Long-term information preservation, which is the main objective of digital repositories. So, this functionality allows the best preservation of all documents stored in the repositories, using several strategies for that preservation;
- Import and export of data, allowing to import content from other repositories, as well as to export the repository itself;
- Interoperability, which is the ability for the information in the repository to be read and interpreted in a different system;
- Viable organizational structure, which is something that conveys confidence and viability.

To better understand all mandatory functionalities and requirements, an informative table was developed to summarize them [Table 3].

Manage documents and metadata according to archival practices and standards.	Log in; Access rights to content; Back up and restoration.
Protect the characteristics of the archival document.	Documents' Management.
Negotiate and accept the appropriate information with the producer.	Long term information preservation.
Obtain control of the content information provided to enable its long-term preservation.	Simple architecture; Availability in multiple languages.
Ensure that the designated community is able to understand the preserved content information without the need of special resources or producers' help.	Use of digital preservation strategies; OAIS / ISO Standard.
Follow previously documented policies and procedures.	Dissemination of information; Availability in multiple languages.
Turn preserved content information available to the designed community and allow to disseminate them as a copy or traceable.	

**Figure 3.** Informative table with mandatory functionalities and requirements

There are many features extremely important, but the most significant one is the adoption of the OAIS model. Out of the five selected tools only Eprints does not have that functionality. However, comparing this tool with the others, Eprints, together with DSpace, have the simpler architecture, which will facilitate the usage process, without having to resort to external help, such as the producers. The three remaining tools (Archivematica, Fedora and Lockss) have a more complex architecture, which indicate a possible difficulty in using them, despite all the positive features.

Concerning data and data processing (editing), the importance of data import and export functionality has already been mentioned, which is present in all tools. Nevertheless, Eprints and Lockss do not have the ability to edit data.

Archivematica is the only platform that solely uses English, which can make it less popular among users. The functionality of access rights to the content is of the utmost importance since in a trusted digital repository there must be a tight control and rules previously established. Fedora is the only tool that does not have this functionality.

Acknowledging that one of the user's needs is to make an effective search, this was considered a feature of extreme relevance, and only Fedora does not score in this item. In addition to not having this functionality, it does not have, yet, the option of backup and restoration, which turns out to be a danger in its use.

Consequently, we conclude that DSpace is the most complete and suitable tool for building a trusted digital repository. Together with all the features already mentioned, this is the only tool that also has ability for sustainability and reliability assessment methodologies. Thus, and considering the fact that it is one of the most widely used open source digital repository software worldwide, we conclude that all these characteristics grant DSpace, a reliable, professional and secure software.

An informative table was developed in order to better understand the information above:

Accept, on behalf of its depositors, responsibility for the maintenance of digital materials.	Import and export data; data editing; access rights to content.
Have an organizational structure that supports not only the long-term viability of their repositories themselves, but also the digital materials under their responsibility.	Viable organization structure; Interoperability.
Demonstrate economic sustainability and administrative transparency.	Sustainability.
Establish methodologies of systems that consider the expectations of reliability expected by the community.	Reliability assessment methodologies.
Observe some factors related to the organizational responsibilities of the repositories.	Back up and restoration.
Others.	Ease of installation; Effective search.

*Figure 4. Informative table with complementary functionalities and requirements*

## CONCLUDING REMARKS

Despite some difficulties with the installation process, DSpace is one of the best software for developing a trusted digital repository. Besides being a widely used tool, it checks out all the functionalities, requirements and criteria.

A Demo provided by the developer (LyraSis, s.d.) was used to have a more practical experience. This demo's main objective is to provide a complementary help in understanding the operation of the original software, trying to prevent users from starting from scratch and not being able to understand how the software works. Furthermore, a user manual and some parameters were elaborated for the Demo DSpace. The user manual focuses on all the features that are present in the demo and the offered possibilities, from the home page; creation, edition and elimination of deposits; different levels of users; and the search options.

With this work, the identification and analysis of the environment surrounding digital preservation, digital repositories and tools that ensure their functioning were accomplished. The importance of models and rules that these tools must follow to achieve the necessary reliability and security was highlighted. A digital repository that is not trustworthy is unlikely to be considered a viable option for users, because if they do not feel confident that they are using credible information, the repository will no longer have the users' trust and will lose members.

Now more than ever, it is necessary to protect and preserve information. It is increasingly necessary to have locations and/or tools that provide this protection with the highest quality. More and more information is produced and stored digitally, and the objective of a trusted digital repository is not only to preserve information, but also to make it available to all users.

## REFERENCES

- Arakaki, F., Gonçalez, P., Coneglian, C., Santos, P., & Segundo, J. (2018). Web Semântica e a Preservação Digital: o padrão de metadados PREMIS na proposta do Linked Data. *Informação & Tecnologia*, 5, (1) 141-156. <https://periodicos.ufpb.br/index.php/itec/article/view/38118/22115>
- Arellano, M. (2017). A preservação de repositórios Digitais dentro do modelo OAIS. *FGV SB – Seminário Repositórios Digitais de Acesso Aberto: Práticas e Desafios*. <http://hdl.handle.net/10438/19050>
- Conselho Nacional de Arquivos. (2014). Câmara Técnica de documentos eletrônicos. Diretrizes para a implementação de repositórios digitais confiáveis de documentos arquivísticos. Rio de Janeiro: Arquivo Nacional. [http://www.conarq.arquivonacional.gov.br/media/publicacoes/resol\\_conarq\\_39\\_repositorios.pdf](http://www.conarq.arquivonacional.gov.br/media/publicacoes/resol_conarq_39_repositorios.pdf)
- Júnior, L., & Borges, M. (2015). Políticas de Preservação Digital: estruturação e características. *Desafíos y oportunidades de las Ciencias de la Información y la Documentación en la era digital*. Madrid: Universidad Complutense de Madrid. <https://estudogeral.sib.uc.pt/bitstream/10316/29995/1/politicas%20preservacao%20digital.pdf>
- Júnior, L., & Mota, V. (2012). Políticas de preservação digital no Brasil: características e implementações. *Ciência da Informação*, 41, (1), 51-64. <http://revista.ibict.br/ciinf/article/view/1351/1530>.
- Lampert, S. (2016). Os repositórios DSpace e Archivematica para documentos arquivísticos digitais. *Acervo*, 29, (2), 143-154. <http://oaji.net/articles/2017/3932-1484338415.pdf>
- Lyman, P., & Bresser, H. (2010). Defining the Problem of Our Vanishing Memory: background, current status, models for resolution. In R. Parry (Ed.), *Museums in a Digital Age* (pp. 336-341). Abingdon: Routledge.
- LyraSis. (s.d.). *About DSpace*. Disponível em: <https://duraspace.org/dspace/about/>.
- Martins, A. B., Rodrigues, E., & Nunes, M. B. (2008). Repositórios de informação e ambientes de aprendizagem: criação de espaços virtuais para a promoção da literacia e da responsabilidade social. *Rede de Bibliotecas Escolares*, (3). <https://www.rbe.mec.pt/news/newsletter3/repositorios.pdf>.