UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília

OFICINA UNIVERSITÁRIA
CULTURA ACADÊMICA Editora

Perspectives on preservation in digital libraries in the light of metadata: paths for the sustainability of Luso-Brazilian heritage

Gabriela Aparecida da Cunha Yamane
Fabiano Ferreira de Castro
Israel Guerra Yamane

Como citar: YAMANE, Gabriela Aparecida da Cunha; CASTRO, Fabiano Ferreira de; YAMANE, Israel Guerra. Perspectives on preservation in digital libraries in the light of metadata: paths for the sustainability of Luso-Brazilian heritage. *In:* TERRA, Ana Lúcia; FUJITA, Mariângela Spotti Lopes (org.). **Integrating Information Science for Sustainable Development:** Topics and Trends. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2025. p. 573-595. DOI: https://doi.org/10.36311/2025.978-65-5954-624-4.p573-595

Perspectives on preservation in digital libraries in the light of metadata: paths for the sustainability of Luso-Brazilian heritage

Gabriela Aparecida da Cunha Yamane [1]
Fabiano Ferreira de Castro [2]
Israel Guerra Yamane [3]

Abstract: The curation of rare book collections involves the protection of cultural heritage through methods such as microfilming, reissuing, and digitization. Digitization creates administrative and descriptive metadata, transforming analog data into digital data. Administrative metadata, such as creation dates and formats, are generated automatically, while descriptive metadata-titles, authors, subjects, and keywords-are added later for resource retrieval. How can we ensure the long-term accessibility of digital library metadata, as defined by the World Commission on Environment and Development in 1991? Preservation initiatives and curators of rare collections are coordinating digitization for unified access. For example, the Luso-Brazilian Digital Library integrates collections from the National Libraries of Brazil and Portugal, the National Register of Digital Objects, and the Brazilian Virtual Memory Network. This study aims to investigate the

[1] Federal University of São Carlos, UFSCar, São Carlos, Brazil.
E-mail: gabrielacunha91@gmail.com | ORCID ID: https://orcid.org/0000-0002-2906-0168

[2] Federal University of São Carlos, UFSCar, São Carlos, Brazil.
E-mail: fabianocastro@ufscar.br | ORCID ID: https://orcid.org/0000-0002-8712-2654

[3] Anhembi Morumbi University, São Paulo, Brazil.
E-mail: isragel@gmail.com | ORCID ID: https://orcid.org/0000-0003-0585-4809

https://doi.org/10.36311/2025.978-65-5954-624-4.p573-595

alignment of the Luso-Brazilian Digital Library's metadata standards with W3C best practices, preserving metadata in digital collections for the memory and sustainability of the Luso-Brazilian heritage.

Keywords: metadata standards; digital preservation; digital collections

Introduction

The preservation of and access to cultural heritage, especially in the digital environment, have become constants as historical and cultural legacies are increasingly constituted in digital formats and made available to the public via the Internet. In this scenario, digital libraries are responsible for ensuring the availability and preservation of their collections, prioritizing the dissemination of cultural heritage.

The term "digital library" is defined by Reitz (2004) in the Online Dictionary for Library and Information Science (ODLIS) as: "A library in which a significant portion of the resources are in machine-readable format (as opposed to print or microform) and accessible by computer. Digital content may be stored locally or accessed remotely via computer networks." (Reitz, 2004,).

For the Digital Library Federation (1998),

> Digital libraries are organizations that provide the resources, including specialized staff, to select, structure, provide intellectual access to, interpret, distribute, preserve the integrity of, and ensure the longevity of collections of digital works so that they are readily and economically available for use by a defined community or set of communities. (p.1).

In this context, digital libraries consist of digital collections that can be more accurately described as sets of items "[...] converted into machine-readable format for preservation or electronic access." (Reitz, 2004). This conversion to digital formats is directly related to the idea of digital curation, which involves careful selection, creation of digital objects, and ensuring reliable long-term access, along with the need to make items electronically accessible, which is an essential aspect of digital

curation when preserving and making resources available. According to Reitz (2004), digital curation refers to

> [...] active management, enhancement and preservation of trusted digital research data throughout its lifecycle. Digital curation involves verifying the integrity of digital data, selecting trusted digital data for its long-term value, creating digital objects and associated metadata, transferring digital objects to trusted digital repositories for secure storage, providing access to designated users, and periodically re-evaluating digital formats to prevent obsolescence.

For instance, in institutions dedicated to safeguarding cultural heritage with rare collections, the curation mission includes, among other things, policies that cover the content of cultural goods, including the safeguard of the goods through processes of microfilming, reissuing, digitization, description, and availability (Pinheiro, 2015). From this perspective, in which digitization and availability are understood as part of the digital curation process, and in which different types of administrative and descriptive metadata are generated, the transformation of analog information into digital data and information requires ensuring the long-term accessibility of this metadata in digital libraries.

In the model developed for the MoA II testbed project, Hurley et al. (1999) classify metadata into three categories:

- Descriptive metadata: which aims to facilitate the discovery and identification of objects;
- Structural metadata: which provides support for the display and navigation of objects;
- Administrative metadata: which comprises all the essential management information for the object, covering details about the creation process, storage formats, the origin and provenance of the objects, as well as the intellectual property rights associated with them. (p.21).

The authors mention categories that describe the different types of metadata, indicating that each typology provides specific information according to its function and usage. Structural metadata consists of organizational details that facilitate understanding the internal architecture of digital resources, offering insights into the structure of the content itself. On the other hand, descriptive metadata must be recorded subsequently to provide information enabling the search and retrieval of digital resources. Meanwhile, administrative metadata can be automatically captured, identifying attributes such as creation date, file format, resolution, and image size.

In the contemporary digital context, the metadata importance is recognized as an essential part of the discovery, management, and reuse of all types of objects, whether digital or not.

As defined by Yee (2009) and elaborated by Castro (2012), data structure encompasses the intangible layer within the representation and description process of along with metadata formats and standards. These interconnected elements promote the interoperability of digital information environments to ensure access, retrieval, use and reuse of information resources by human and machine users.

The tangible layer, according to Castro (2012), pertains to data provision in the Web environment. It represents the final stage and involves the ability to present them as they were constructed and stored (input), thus becoming available for retrieval. Therefore, it can be inferred that the tangible layer concerns data presented to users in an accessible and visible manner.

Given the definition proposed by the World Commission on Environment and Development (Comissão Mundial…, 1991), which states that sustainable development is the ability to meet the needs of the present without depriving or compromising restricting future generations from meeting the same needs in the future, the following question arises: how can we ensure that digital library metadata is accessible, reusable, and reproducible in the long term?

Considering the challenges present in the current technological digital preservation scenario, such as the metadata conservation in digital libraries, the general objective of this research is to identify the metadata standards used in the Luso-Brazilian Digital Library, aligned with the Best Practices for publishing data on the Web proposed by the World Wide Web Consortium (W3C), in order to ensure the integrity and continuous access to digital resources over time, promoting the sustainability of cultural heritage. Thus, the specific objectives are: a) Highlight the importance of digital preservation in promoting sustainability in cultural heritage collections; b) Identify metadata standards adopted by the LBDL; c) Map W3C's best practices in the LBDL.

In response to this demand, it is imperative to ensure, through the metadata and metadata standards available in the digital preservation process, in conjunction with the use of technological tools, that the data stemming from the digitalization and provision of cultural heritage remains accessible and available in the long term.

Literature review

The digitization of resources and the availability of institutions' digital collections on a single platform are coordinated by cultural heritage preservation programs in digital environments, functioning as digital curators of rare collections. As an example in this research, the Luso-Brazilian Digital Library (LBDL) is mentioned, which comprises a portal aggregating digital collections from the National Libraries of Brazil and Portugal, the National Register of Digital Objects and the Brazilian Virtual Memory Network.

The project of the Luso-Brazilian Digital Library (BDLB), launched in February 2014 by the national libraries of Brazil and Portugal, aims to coordinate the digitization of collections and provide all digital material from the two institutions on a single platform. This represents a significant initiative to expand the online reach of cultural content shared by the common history among Portuguese-speaking countries (BDLB, 2024).

Although termed a digital library, the project has a digital repository structure, with digital aggregator services characteristics, functioning as gathers and provides digital content from four different institutions in one place. According to Oliveira and Carvalho (2009), "[...] aggregators collect metadata from different data providers and make it available to join different service providers." (p.9). Thus, the main function of the LBDL is to aggregate and centralize data related to these collections, simplifying access to content scattered across the Internet.

Digital Repositories (DR) are online platforms that organize scientific production from institutions or thematic areas, capable of storing different file formats. These platforms offer various benefits, such as visibility for research, preservation of scientific production development and can be institutional or thematic, encompassing specific productions of an institution or a specific area. (Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT, 2020)

In this sense, the LBDL can be considered a thematic repository as it is an online environment for the preservation, dissemination, access, and storage of production from the shared cultural heritage by history and language among Portuguese-speaking nations.

The software used by the Luso-Brazilian Digital Library is DSpace, which enables advanced search through Collections (National Digital Library of Portugal, BNDigital of Brazil, Rede Memória, RNOD) and the application of filters such as Authors, Titles, Subjects, Document Type and Languages.

According to the IBICT (2023), the DSpace software is a free open-source software originally developed by the Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP). The software's global use is promoted, developed and supported by the DuraSpace organization, while in Brazil this role is played by the IBICT. Through DSpace storage, management, preservation and visibility assurance of intellectual production in digital repositories and libraries are enabled. Additionally, the management of various types of digital documents is allowed by the software, ensuring long-term access and offering the possibility of

customization of multiple functionalities to meet the informational demands of institutions, their users, and interoperability with other software and information environments.

In Brazil, the DSpace software has been used by other institutions to provide digitized collections and metadata resulting from the digitization process of rare collections, for example. Garcia (2021) highlights the use of DSpace at the Biblioteca Brasiliana Guita e José Mindlin, an interdisciplinary center for information and documentation, research and scientific dissemination at the University of São Paulo (USP).

The increasing availability of data and information in digital repositories in the Web era raises questions about long-term access and the data and metadata's integrity. In this context, it is essential to ensure access to content by promoting data interoperability in these informational environments. The World Wide Web Consortium (W3C) (2017) has established recommendations for "Best Practices for Data on the Web", which are guidelines that aim not only to standardize the publication of data on the Web, but also to ensure the consistency, quality, and long-term data and metadata's preservation.

Overall, the goal is for data publishers to share information, whether in an open or controlled manner, while consumers, who can also play the role of publishers, aim for data to be found, used, and related accurately, updated, and always accessible. In this sense, the W3C Recommendations propose guidelines for publishers so that data management becomes more consistent, promoting the use and reuse of shared data.

Each Best Practice described in the recommendations proposes ways for publishers and data consumers to overcome challenges in publishing and consuming data on the Web. For each challenge in the document, one or more Best Practices are presented, with at least one indicating its relevance (Lóscio et al., 2017).

As described by Lóscio et al. (2017) and illustrated in Table 1, each good practice has a corresponding set of benefits. According to Lóscio et al. (2017), "[...] a benefit represents an improvement in the way datasets are made available on the web. A best practice (BP) may have one or more

benefits." However, upon analyzing the set of benefits, it was found that in each set of BPs, each BP has at least two benefits, with most being related to data reuse. The benefit of reuse can be directly related to the issues of preservation and sustainability discussed in this research.

In Table 1, the benefits are represented by codes that can be associated with each Best Practice and relate to: Reusability (R), Comprehensibility (C), Discoverability (D), Processability (P), Trust (T), Linkability (L), Interoperability (I) and Accessibility (A). It is also noted that some benefits are associated with the different Best Practices, highlighting the multifunctionality of these practices to improve different aspects of the management and use of data on the Web. The contribution of Best Practices in facilitating comprehension, discovery, and reuse of data sets, as well as in fostering trust, interoperability, and accessibility, is emphasized. These practices provide a comprehensive set of essential guidelines for ensuring the quality, accessibility, and usefulness of data available on the Web, promoting standardization, and enhancing the user experience with digital data.

Table 1: Publishing Data on the Web: requirements and benefits

BP	Requirements	Benefits							
1	Provide metadata	R	C	D	P				
2	Provide descriptive metadata	R	C	D					
3	Provide structural metadata	R	C		P				
4	Provide data license information	R				T			
5	Provide data provenance information	R	C			T			
6	Provide data quality information	R				T			
7	Provide a version indicator	R				T			
8	Provide version history	R				T			
9	Use persistent URIs as identifiers of datasets	R		D			L	I	
10	Use persistent URIs as identifiers within datasets	R		D			L	I	
11	Assign URIs to dataset versions and series	R		D		T			
12	Use machine-readable standardized data formats	R			P				

13	Use locale-neutral data representations	R	C						
14	Provide data in multiple formats	R			P				
15	Reuse vocabularies, preferably standardized ones	R	C		P	T		I	
16	Choose the right formalization level	R	C					I	
17	Provide bulk download	R							A
18	Provide Subsets for Large Datasets	R			P		L		A
19	Use content negotiation for serving data available in multiple formats	R							A
20	Provide real-time access	R							A
21	Provide data up to date	R							A
22	Provide an explanation for data that is not available	R				T			
23	Make data available through an API	R			P			I	A
24	Use Web Standards as the foundation of APIs	R		D	P		L	I	A
25	Provide complete documentation for your API	R				T			
26	Avoid Breaking Changes to Your API					T		I	
27	Preserve identifiers	R				T			
28	Assess dataset coverage	R				T			
29	Gather feedback from data consumers	R	C			T			
30	Make feedback available	R				T			
31	Enrich data by generating new data	R	C		P	T			
32	Provide Complementary Presentations	R	C			T			A
33	Provide Feedback to the Original Publisher	R				T		I	
34	Follow Licensing Terms	R				T			
35	Cite the Original Publication	R		D		T			

Legenda
R = Reuse
C = Comprehension
D = Discoverability
P = Processability
T = Trust
L = Linkability
I = Interoperability
A = Access

Source: Adapted from W3C (2017).

The application of Best Practices is intended to help in the promotion of data reuse and to establish greater trust in data among developers, regardless of the technology they use, thereby enhancing the role of the open platform that the Web represents by establishing effective connections.

Each of the benefits identified in Table 1 represents an improvement in the amount of data available on the Web. According to Lóscio et al. (2017), the improvements are described as:

> Comprehension (C): people will have a better understanding of the structure and meaning of the data, the metadata and the nature of the dataset.
>
> Ease of Processing (P): machines will be able to automatically process and manipulate the data within a dataset.
>
> Discoverability (D): machines will be able to automatically discover a dataset or data within a dataset.
>
> Reuse (R): the possibilities of reusing the dataset by different groups of data consumers will increase.
>
> Trust (T): the trust customers place in the dataset will increase.
>
> Linkability (L): it will be possible to create links between data resources (datasets and data items).
>
> Ease of Access (A): people and machines will be able to access up-to-date data in a variety of ways.
>
> Interoperability (I): it will be easier to reach consensus between data publishers and consumers.

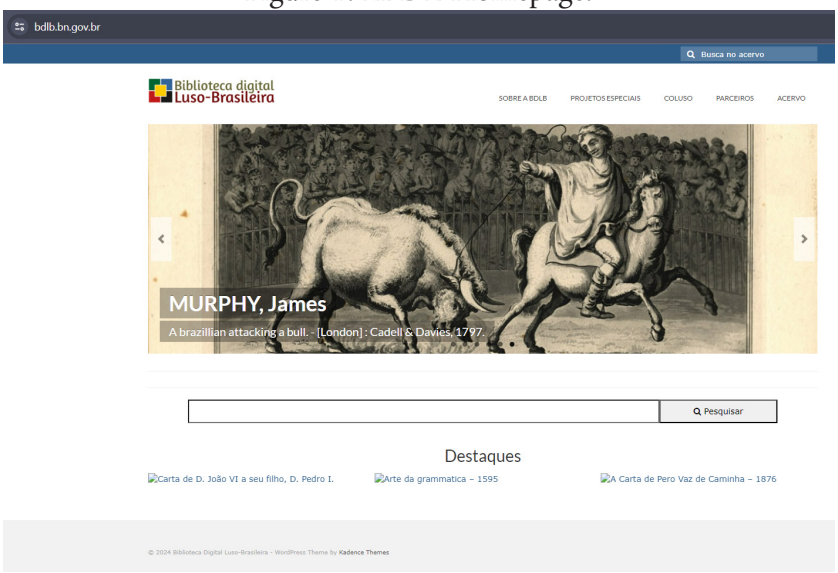
Given the requirements presented and the data management improvements offered by the W3C Best Practices, it is understood that these guidelines are fundamental for institutions that manage cultural heritage in the Web environment, such as libraries or digital repositories. These practices improve data discovery, facilitate processing, enhance search and retrieval, ensure data trust, and connect disparate resources. In addition, they promote data reuse, ensure its access, use and long-term preservation, enable the digital delivery of cultural heritage, and promote sustainability.

Although the requirements listed in the Best Practices for Publishing Data on the Web were not designed specifically for digital libraries, it is understood that their adoption can ensure greater consistency and data quality.

In this research, the example of the Luso-Brazilian Digital Library (LBDL) stands out due to its historical significance and the importance it places on preserving and protecting the heritage of Brazil and Portugal. Initially, the project included the digital collections of the national digital libraries of Brazil and Portugal, the National Register of Digital Objects, and the Brazilian Virtual Memory Network. The LBDL can be accessed at https://bdlb.bn.gov.br/.

On the LBDL home page, Figure 1, a carousel of images from the digital collection is highlighted. At the top is a simple search engine, and just below are five navigation tabs: "About LBDL", "Special Projects", "Coluso", "Partners" and "Collection". At the bottom, just below the carousel, is another larger search engine, followed by highlights from the collection, also in carousel format.

Figure 1: LBDL Homepage.

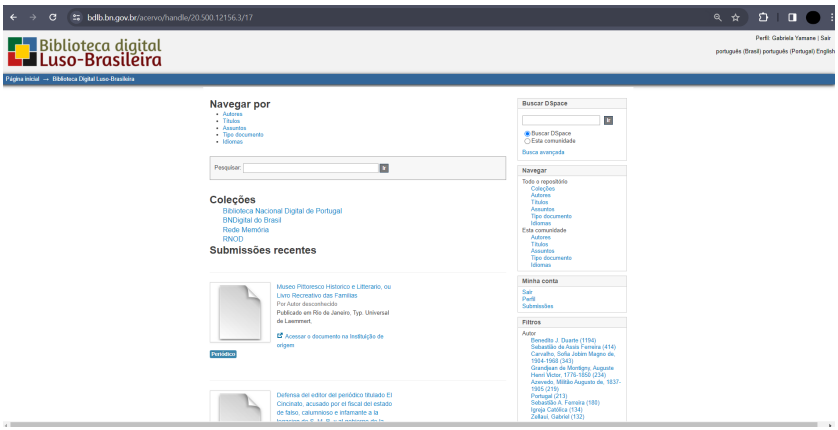


Source: Authors (2024).

The platform does not provide visual or auditory accessibility options, nor does it provide alternative language references on the home page. In addition, there is no FAQ section to help users with search and retrieval, nor is there a contact point for clarification.

Browsing and searching options are presented to researchers when they click the Collection tab, as shown in Figure 2. Three language options are available on this screen: Portuguese (Brazil), Portuguese (Portugal), and English. The items, called "Recent Submissions", are displayed in list format and include data such as title, author, material type, print, and access link to the originating institution. In the right-hand column, there are filter options for simple and advanced searches; you can search the entire repository or a specific collection. You can also subscribe to collections and receive updates.

Figure 2: Collection interface.

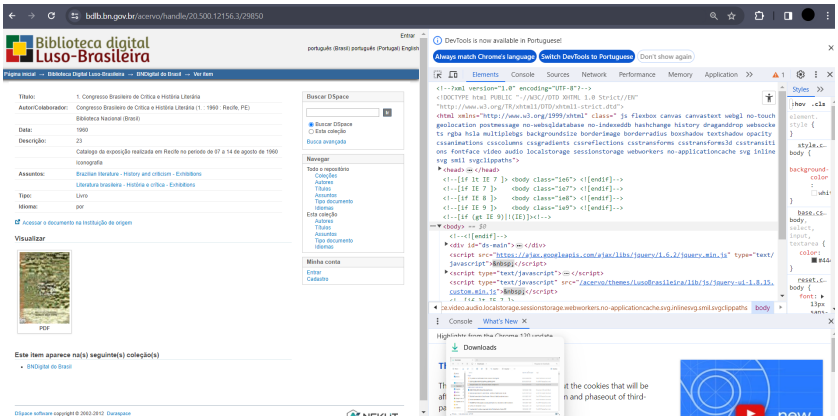


Source: Authors (2024).

The documents available through the LBDL are organized into collections or communities, encompassing for example Album, Article, Atlas, Drawing, Disk, Electronic document, Ephemera, Fascicle, Flyer, Photograph, Print, Image, Incunabula, Libretto, Book, Manuscript, Map, Printed and manuscript cartographic material, Two-dimensional graphic material, Printed and manuscript textual material, Matrix, Microform,

Multimedia, Printed and manuscript music, Rare work, Score, Musical scores, printed and manuscript, Periodical, Plant, Computer products, Electronic resource, SOUND, Polycopy text, Book tome.

Figure 3: Register presentation.



Source: Authors (2024).

Regarding the presentation of the records, as depicted in Figure 3, it was noted that only a few fields allow interconnection with other records, thus enabling the retrieval of additional information through hyperlinks, such as in the case of subjects. In this regard, other types of relationships are not established, which limits the search and discovery of new documents, for example, by date, document type, authorship, or imprint.

Each record displays a description of the item that generally includes the fields covered by the Dublin Core metadata standard (Title; Creator; Subject; Description; Publisher; Contributor; Date; Resource Type; Format; Identifier; Source; Language; Relationship; Coverage; Rights), but when viewing the data, researchers are not informed of the metadata standards or protocols used. Furthermore, there is no way to import metadata or select the visualization format. It was possible to identify the use of Dublin Core from the source code of the library page.

The LBDL's digital collection is not centralized in one of the project's repositories. Each item in the digital collection, known as a digital object,

is maintained and managed by each participating institution, which is responsible for making it available. When a researcher accesses a particular item, he or she is directed to the original institution's collection via a hyperlink. This approach reflects the decentralized nature of the LBDL, where the LBDL acts as a service aggregator, unifying and providing access to different digital collections held by partner institutions. This distributed structure emphasizes the collaboration among participating institutions, allowing researchers to explore and use the digital resources of different institutions in a single location.

In this context, by directing researchers to the original collections through hyperlinks, the LBDL takes a decentralized approach to preserving digital collections. In this way, each institution can apply its best digital preservation practices, promoting the sustainability of the collections over time.

Rockembach and Pavão (2024) state that the hyperlink serves as the link between the user and the online resource they seek, and studies in the field have shown that ensuring access to these hyperlinks is highly fragile, requiring digital preservation measures.

As described by Sayão (2010), digital preservation is a constant challenge, both in terms of technical and organizational activities, because the physical state of digital collections is highly dependent on the current technological context, which in turn can change rapidly.

According to Rockembach and Pavão (2024), institutions such as libraries and museums that make historical and cultural information available on the Web tend to have more comprehensive digital preservation policies, which may include specifics for digitizing and storing physical artifacts as well as procedures for curating and organizing this content in the Web environment. The authors note that tools for capturing, storing, and managing content on the Web are essential elements that must be considered to ensure that information is preserved and made available to users over time.

Therefore, it is essential for institutions collaborating on digital preservation projects to coordinate in order to apply policies and best practices for the preservation and sustainability of their cultural heritage.

Methodology

The methodology used to carry out the research is qualitative in nature.

With regard to the approach to the subject and the objectives, it is descriptive in nature, using bibliographical research to explore the existing scientific literature on the subject in order to identify the conceptual basis and the context in which it is inserted.

The study adopted bibliographical research as its data collection method.

The methodological procedures involved stages that included, first, a bibliographic survey at national and international levels, followed by the collection of data and metadata in the Luso-Brazilian Digital Library. After this stage, the structure of the LBDL was analyzed, highlighting the availability and organization of the records. This assessment included compliance with established metadata standards and the requirements of the Best Practices for the Dissemination of Data on the Web, as recommended by the World Wide Web Consortium (W3C), with special attention to ensuring the sustainability of the system.

Results

As mentioned in the introduction, the digitization of rare collections is a way to preserve historical and cultural collections in digital formats for sustainability and long-term availability. Nevertheless, the majority of institutions lack a well-defined digital preservation policy, with many policies being either vague or entirely absent.

Faced with this issue, this study was based on the identification of the metadata standards adopted by the LBDL and proposes a mapping of the best practices recommended by the W3C.

Table 2 shows a mapping between the thirty-five best practices recommended by the W3C and the way in which the digital resources of the Luso-Brazilian Digital Library (LBDL) are represented and made available, according to the dataset available at: https://bdlb.bn.gov.br/acervo/handle/20.500.12156.3/29850.

Table 2: Mapping between W3C and LBDL Best Practices.

BP	Requirements	Comply	Doesn't comply	Partially complies	BP analysis
1	Provide metadata	X			
2	Provide descriptive metadata			X	The descriptive metadata is available as a dataset in human-readable format. To access the metadata in a valid machine-readable format, it is necessary to access the developer options.
3	Provide structural metadata			X	Structural metadata is not available in datasets in human-readable format. To access the metadata in a valid machine-readable format, it is necessary to access the developer options.
4	Provide data license information			X	DSpace License Identified © 2002-2012 Duraspace.
5	Provide data provenance information		X		
6	Provide data quality information		X		It was not possible to identify.
7	Provide a version indicator		X		It was not possible to identify.
8	Provide version history		X		It was not possible to identify.

9	Use persistent URIs as identifiers of datasets	X			Handle
10	Use persistent URIs as identifiers within datasets	X			PURL, Bn of Portugal.
11	Assign URIs to dataset versions and series		X		
12	Use machine-readable standardized data formats	X			
13	Use locale-neutral data representations	X			
14	Provide data in multiple formats		X		
15	Reuse vocabularies, preferably standardized ones		X		
16	Choose the right formalization level	X			
17	Provide bulk download		X		
18	Provide Subsets for Large Datasets		X		
19	Use content negotiation for serving data available in multiple formats		X		
20	Provide real-time access	X			
21	Provide data up to date			X	It does not specify the update frequency.
22	Provide an explanation for data that is not available		X		
23	Make data available through an API	X			Google API's
24	Use Web Standards as the foundation of APIs	X			

25	Provide complete documentation for your API		X		
26	Avoid Breaking Changes to Your API		X		
27	Preserve identifiers			X	In some cases, it informs you that it has been removed but does not inform you of a possible contact.
28	Assess dataset coverage	X			DCterms.
29	Gather feedback from data consumers		X		
30	Make feedback available		X		
31	Enrich data by generating new data		X		
32	Provide Complementary Presentations		X		
33	Provide Feedback to the Original Publisher		X		
34	Follow Licensing Terms		X		
35	Cite the Original Publication	X			

Source: Authors (2024).

Table 2 shows that the LBDL fulfills the requirement in BP 1 - "Metadata provision" by employing Dublin Core terms, for example, to offer descriptive metadata. In BP 2 - "Providing descriptive metadata" there is partial compliance as the metadata is accessible in a human readable format, but to access it in a machine readable format it is necessary to use the developer options. As for structural metadata, according to BP 3 - "Provide structural metadata", there is also partial compliance, since this metadata is not available in a human readable format, and access to developer options is required to read it in a machine-readable format.

Regarding the requirement that deals with the license, BP 4 - "Provide information about the data license", the DSpace© 2002-2012 Duraspace license was identified, other licenses could not be identified because the information is not clearly expressed on the LBDL page and therefore could not be identified by human users.

There are also requirements that are not complied by the LBDL or are not clearly identified. For example, BP 5 - "Provide data provenance information" in the context of the LBDL is not clearly presented and it has not been possible to determine whether the requirement is met, so it has not been located by human users.

Best Practices 6, 7, and 8, related to "Provide Data Quality Information", "Provide Version Indicator", and "Provide Version History", were assigned a "Does Not Meet" rating because it was not possible to determine whether the requirement was met because the information is not explicitly available on the LBDL page and therefore would not be identified by human users.

Regarding URIs, the LBDL was found to comply with Best Practices BP 9 and 10, "Use persistent URIs as dataset identifiers" and "Use persistent URIs as identifiers within datasets". However, in relation to BP 11, "Assign URIs to versions of datasets and series", it was not possible to determine whether the requirement is met.

In BP 12 - "Use standardized machine-readable data formats", it should be noted that the XML data is embedded in the developer's code and is not available in a separate file for download or import.

It was not possible to determine whether the requirement was met for BP: 17, 18, 19, 22, 29, 30, 31, 32, 33, 34 because the information is not explicit on the LBDL page and therefore could not be located by human users.

Regarding updates, in the requirement "Provide updated data", BP 21, the frequency of data updates is not specified.

Concerning APIs in BPs 23 and 24, the LBDL meets the requirements to "Make data available through an API" and "Use web standards as a basis

for building APIs". However, it does not meet the requirements to "Provide complete documentation for APIs" and "Avoid changes that affect the functioning of your API", which correspond to BPs 25 and 26, respectively.

Regarding BP 27 - "Maintain identifiers", the LBDL it was observed that the LBDL partially meets the requirements, as in some cases, as in some cases it informs that the identifier has been removed or is not available, but it does not provide a contact to report the problem or try other forms of access.

Discussion

Through the mapping, Table 2, it was possible to observe whether or not the LBDL complies with the requirements established by the W3C Best Practices, allowing the identification of areas in which the Digital Library can improve its standards and align its work with the best practices for publishing data on the Web, with a view to preserving and providing access to cultural heritage.

The statistical analysis of the mapping between the W3C Best Practices and the Luso-Brazilian Digital Library (LBDL) shows the panorama of the LBDL's compliance with these guidelines. Out of the 35 requirements identified, 11 were identified as "Comply", suggesting alignment with the recommended practices. However, 19 requirements were defined as "Doesn't Comply" as they could not be identified by human users, areas that require attention and development to ensure greater compliance with the established requirements. In addition, 5 requirements were defined as "Partially Comply", indicating the presence of positive initiatives, but with a gap for specific improvements. This analysis offers a view of the LBDL scenario in relation to best practices, providing insights for improving the quality, access and interoperability of the digital resources provided.

With regard to the Best Practices that were defined as "Doesn't comply", it is possible to establish relationships and group them according to their specificity and in accordance with the established guidelines. The groups are organized as follows:

Group 1 - Metadata and Contextual Information. Providing provenance information (BP 5) and quality information (BP 6) are essential aspects for understanding the origin and reliability of data; version indicators (BP 7) and version history (BP 8) contribute to the traceability and temporal reliability of data; assigning URIs to versions of datasets and series (BP 11) and providing an explanation for unavailable data (BP 22) are also related to the contextualization and transparency of data.

Group 2 - Data formats and access. Providing data in multiple formats (BP 14) and offering bulk downloads (BP 17) relate to the diversity of formats and data access options; providing subsets for large datasets (BP 18) and using content negotiation (BP 19) aim to facilitate data retrieval and manipulation.

Group 3 - Interactivity and Enrichment. Collecting feedback from data consumers (BP 29) and sharing available feedback (BP 30) indicate the importance of interactivity and engagement with users; enriching data by generating new data (BP 31), providing complementary visualizations (BP 32), and providing feedback to the original publisher (BP 33) are related to continuous improvement and interactivity in the library environment.

Group 4 - Governance and Compliance. Providing complete documentation for APIs (BP 25) and avoiding changes that affect the functioning of the API (BP 26) are related to the governance and stability of the services provided; adherence to license terms (BP 34) is fundamental to ensuring legal and ethical compliance in the provision of data.

These relationships indicate that greater focus will be directed towards metadata, data formats, user interactivity, and governance issues, all of which can significantly contribute to aligning the LBDL with the best practices recommended by the W3C.

Conclusion

This research examined the Luso-Brazilian Digital Library (LBDL) through the lens of the W3C Best Practices, highlighting its significance for the preservation of the Luso-Brazilian cultural heritage. The analysis

revealed that although the LBDL meets some of the requirements of the Best Practices proposed by the W3C, there are aspects that require greater attention, particularly concerning metadata and data formats, as these are considered key elements for ensuring digital preservation. Therefore, the importance of preservation policies that guarantee the access, use and reuse of data among collaborating institutions over time is emphasized.

Bringing together the tangible and intangible layers of digital environments, as the BDL does in promoting interoperability, presents a challenge that will enable any type of user, whether human or machine, to recognize, process, and utilize data and metadata. The adoption and utilization of metadata standards linked to the W3C Best Practices are sine qua non for creating accessible digital collections that can be used and reused in the long term.

It is hoped that this study can contribute to a reflection on digital preservation, emphasizing the need for more precise guidelines and explicit policies. The LBDL serves as a prime example of the ongoing challenges in digital preservation and the importance of adhering to internationally recognized standards to ensure the interoperability and sustainability of digital cultural heritage. The LBDL's gradual embrace of W3C Best Practices can ensure its contribution to the preservation of cultural heritage for future generations.

References

Biblioteca Digital Luso-Brasileira. (2024). *Sobre a BDLB*. BDLB. https://bdlb.bn.gov.br/?page_id=193

Castro, F. F. D. (2012). *Elementos de interoperabilidade na catalogação descritiva*: configurações contemporâneas para a modelagem de ambientes informacionais digitais. [Tese de Doutorado, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista]. Repositório Institucional Unesp. http://hdl.handle.net/11449/103364

Comissão Mundial sobre Meio Ambiente e Desenvolvimento. (1991). *Nosso futuro comum*. FGV. https://edisciplinas.usp.br/pluginfile.php/4245128/mod_resource/content/3/Nosso%20Futuro%20Comum.pdf

Digital Library Federation. (1998). *A working definition of digital library*. DFL. https://old.diglib.org/about/dldefinition.htm

Garcia, R. M. (2021). Desenvolvimento da nova Biblioteca Digital da BBMUSP: relato de experiência. In Colmenero-Ruiz, M. J., Cuevas-Cerveró, A., Passarelli, B. & Paletta, F. C. (Org.). *Memória, Patrimônio Digital e Políticas de informação:* trilhas abertas no Brasil e Espanha (pp. 67-83). ECA-USP.

Hurley, B. J., Price-Wilkin, J., Proffitt, M., & Besser, H. (1999) *The Making of America II Testbed Project:* a digital library service model. Digital Library Federation. https://www.clir.org/pubs/reports/pub87/

Instituto Brasileiro de Informação em Ciência e Tecnologia. (2023). *DSpace*. https://www.gov.br/ibict/pt-br/assuntos/tecnologias-para-a-informacao/DSpace

Instituto Brasileiro de Informação em Ciência e Tecnologia. (2020) *Repositórios digitais*. https://antigo.ibict.br/informacao-para-a-pesquisa/repositorios-digitais

Lóscio, B. F., Burle, C., & Calegari, N. (2017). *Data on the Web best practices.* W3C Working Draft, World Wide Web Consortium (W3C). https://www.w3.org/TR/dwbp/

Oliveira, R. R., & Carvalho, C. L. (2009). *Implementação de Interoperabilidade entre Repositórios Digitais por meio do Protocolo OAI-PMH.* (Relatório Técnico). Universidade Federal de Goiás, Instituto de Informática, Goiânia.

Pinheiro, A. V. (2015) Livro raro e sustentável. *Revista Museu*. Retrieved January 04, 2024, from https://www.revistamuseu.com.br/site/br/artigos/18-de-maio/18-maio-2015/3098-livro-raro-e-sustentavel.html

Reitz, J. M. (2004). *Online Dictionary For Library And Information Science* - (ODLIS). https://odlis.abc-clio.com/odlis_d.html

Rockembach, M., & Pavão, C. G. (2024). *Arquivamento da web e preservação digital*. Pimenta Cultura. https://doi.org/10.31560/pimentacultural/2024.98119

Sayão, L. F. (2010). Uma outra face dos metadados: informações para a gestão da preservação digital. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 15(30), 1-31. https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1

W3C. World Wide Web Consortium. (2017). *Data on the web best practices*. W3C. https://www.w3.org/TR/dwbp/

Yee, M. M. (2009). Can bibliographic data be put directly onto the semantic web?. *Information Technology and Libraries*, *28*(2), 55–80. https://doi.org/10.6017/ital.v28i2.3175