Article

Integration of cultural data from digital repositories: an overview of the DPLA Hubs

Maria Lígia Triques ¹ (iD), Paula Regina Ventura Amorim Gonçalez ² (iD)
Ana Cristina de Albuquerque ³ (iD)

ABSTRACT

Introduction: The Digital Public Library of America (DPLA) is an initiative aimed at expanding the reach of its content and the possibilities of improving and using data to generate new knowledge from data integration. The search for integration, however, results in a series of possibilities and challenges for digital repositories, providers and data servers. **Objective:** To present an overview of the technological solutions underlying data collection and availability from the DPLA partner repositories, named Hubs, to point out aspects of the interoperability opportunities and challenges they face in the search for integrated access. **Methodology:** This is a bibliographical, qualitative and exploratory research, with analysis based on the discussion found in the literature and the results obtained in the analysis of the platforms. **Results:** The technological solutions adopted both by the DPLA to collect and aggregate data and by the Hubs, data providers, which use different solutions to aggregate data from the various contributing cultural institutions and provide them to the DPLA, are presented as results. **Conclusion:** The study concluded that despite the technical, semantic and managerial challenges, great efforts have been undertaken to adopt solutions and best practices aimed at the engagement of data providers. Both the Hubs and their contributing institutions present programs to encourage and support digitization and documentation, as well as guidelines to guarantee the quality and legal rights of access to content, which is essential to improve users' experience and multi-dimension the possibilities to transform information into knowledge.

KEYWORDS

Interoperability. Digital repositories. Digital Public Library of America. DPLA.

| 1

A integração de dados culturais de repositórios digitais: um panorama dos Hubs da DPLA

RESUMO

Introdução: A *Digital Public Library of America* (DPLA) é uma iniciativa que visa ampliar o alcance de seu conteúdo e as possibilidades de aprimoramento e utilização dos dados para a geração de novos conhecimentos a partir da a integração de dados. A busca pela integração, contudo, resulta em uma série de possibilidades e desafios para os quais os repositórios digitais, provedores e servidores de dados, precisam se adaptar. **Objetivo**: Apresentar um panorama das soluções tecnológicas subjacentes à coleta e disponibilização de dados dos repositórios parceiros da DPLA, denominados *Hubs*, de modo a pontuar aspectos das oportunidades e dos desafios de interoperabilidade que

Author's correspondence

¹ State University of Londrina, Londrina, PR, Brazil / e-mail: mligia.triques@uel.br

² Universidade Federal of Espírito Santo, Vitória, ES, Brasil / e-mail: paulaventuramorim@gmail.com

³ State University of Londrina, Londrina, PR, Brazil / e-mail: albuanati@uel.br

eles enfrentam na busca pelo acesso integrado. **Metodologia:** A pesquisa qualitativa e teórica, descritiva-exploratória, utilizando como método a Análise de Conteúdo. **Resultados:** Apresenta-se como resultados as soluções tecnológicas adotadas tanto pela DPLA, para coletar e agregar dados, como dos *Hubs*, provedores de dados, que utilizam diferentes soluções para agregar dados das diversas instituições culturais contribuintes e fornecê-los à DPLA. **Conclusão**: Conclui-se que apesar dos desafios técnicos, semânticos e gerenciais, existem grandes esforços para adotar soluções e melhores práticas visando principalmente o engajamento dos provedores de dados. Tanto os *Hubs* quanto suas instituições contribuintes apresentam programas de incentivo e auxílio à digitalização e à documentação, bem com diretrizes para garantir a qualidade e os direitos legais de acesso do conteúdo, o que é fundamental para melhorar a experiência dos usuários e multidimensionar as possibilidades de transformar informação em conhecimento.

PALAVRAS-CHAVE
Interoperabilidade. Repositórios digitais. *Digital Public Library of America.* DPLA.

CRediT

JITA: HS. Repositories.

Article submitted in the similarity system

1 INTRODUCTION

Cultural Heritage and collections constitute important knowledge records and, as such, are valuable information sources that represent objects or testimonies, tangible or intangible, of the tradition or manifestation of a people. Through the availability in digital environments, especially in repositories, these collections have been increasingly found in digital formats, from the creation and sharing of their data in an open and distributed way on the web (HYVÖNEN, 2012). Data from cultural heritage and collections represent the various forms of expression of peoples, reflecting the diversity of customs, daily practices, beliefs and languages.

These diversified data sets come to represent a part of larger sets. They can benefit from the web interface to enhance their semantic networks, from what Marcondes (2018) calls "semantic links", possible culturally relevant relationships that cultural information resources establish with each other, in the process of linking their data.

Different technological solutions created for web applications have enabled and facilitated this network of semantic relationships, especially with the set of technologies that provide Linked Data and Linked Open Data, in the case of data connected under free, open and public licenses. Both consist of good practices, based on international standards recommended by the World Wide Web Consortium (W3C), such as the Uniform Resource Identifier (URI) and the Resource Description Framework (RDF), among others, that allow the creation of connected data sets (ISOTANI; BITTENCOURT, 2015).

Even with the constant evolution of technological solutions enhancing the use and scope of digital collections, interoperability is still one of the main areas of discussion in various aspects involving the studies and implementations of data sets, especially in the integration case. As highlighted by Santarem Segundo, Silva and Martins (2018), information and knowledge are key requirements in the contemporary context, and in this context, integration becomes indispensable to transform information into more far-reaching knowledge.

From this perspective, it can be said that, regarding access, use and reuse, integration also allows to transform information into knowledge with better quality, by allowing the expansion and enrichment of data sources, as well as the information and semantic connections between them, considering the community and target audience. Marcondes (2018) refers to "semantic links" when discussing the implementation of relevant cultural relationships such as interoperability between digital collections, expanding their visibility and usability, and autonomy of memory institutions.

Integration requires the interoperable data exchange, that is, without loss of functionality and meaning. The consolidation of some of these structures allowed, among other applications, the development of services and interaction spaces for cultural heritage and collections, enabling the development of distributed and interoperable processes of creation, as well as content sharing, such as Europeana, the great data aggregator[1] (HYVÖNEN, 2012; SAYÃO, 2016). In the North American context, another initiative similar to Europeana gained prominence for its efforts to unify cultural data sources at the national level. The Digital Public Library of America[2], known by the acronym DPLA, is an open platform that allows the search in a single interface of North American cultural content available on the web through different integrated digital repositories.

Both DPLA and Europeana adopt technologies and standards to fulfill their missions and objectives, based on Semantic web technologies and the principles of Linked Open Data. In the DPLA case, its functionality is based on its own structure called DPLA Metadata Application Profile (DPLA MAP[3]), which allows for the integration of experiences, as well as specific description needs of each community, from the collection and aggregation of metadata

---

[1] Available on: https://europeana.eu/pt.
[2] Available on: http://dp.la/.
[3] Available on: http://dp.la/map.

of partner institutions or, as they are called, the "Hubs", members from which the DPLA collects data (MATIENZO; RUDERSDORF, 2014).

Hubs are data providers and make them available through systems that collect data from a variety of digital repositories, maintained by both large national institutions and smaller institutions that cooperate by geographic region or scope to compile their data records. In either case, the Hubs commit to developing and maintaining an appropriate infrastructure to support aggregation at national scale and complexity. In this way, the DPLA is concerned with establishing guidelines for members to be able to provide interoperable data even when using systems with particularities in terms of interfaces, standards, protocols, intellectual protection norms and, mainly, data interpretation (SAYÃO; MARCONDES, 2008).

Therefore, data integration can be achieved in many ways and at different levels, as well as interoperability, which results in a number of possibilities and challenges that data providers need to adapt to. Thus, considering the North American initiative from the perspective of integrated access, this study questions which paths have been taken by the participating repositories to enable integrated access to collections. The guiding objective of this study is to present an overview of the technological solutions underlying the data collection and availability from the DPLA partner repositories to point out some aspects of the opportunities and challenges they face in the search for integrated access.

2 METHODOLOGY

The methodology used in this work is bibliographic and qualitative, configuring an exploratory investigation. Part of the analysis of the platform that promotes integrated access to cultural heritage and collections made available in digital repositories was performed. The DPLA widely disseminates its operating systems, which makes it suitable for analysis, with a vast amount of available documentation and bibliography.

In operation since April 2013, the DPLA offers the "DPLA Pro" on its web platform. It is an online space in which the various professionals connect and share information about the platform. In this space, documentation that informs, guides, trains partner institutions and other interested parties to contribute to the initiative are found, as well as how to work with data structuring. In addition, each DPLA data provider Hub also presents its own documentation, which allowed us to collect the data necessary for this study.

As the DPLA works with almost 50 Hubs in its total list of partnerships, and each of these Hubs with dozens of contributing institutions, this study divided the analysis of data providers into two groups: the Hubs part of the DPLA's Membership Program[4], partners who mainly use the tools developed by the DPLA; and non-members, who have their own diversified systems without using the technologies provided by the DPLA and, even so, are able to integrate their data effectively. In this way, we sought to limit the scope of the analysis, individually describing only the solutions of the non-partner Hubs, once the partner Hubs follow the DPLA guidelines.

Regarding the bibliographic survey, the following sources were researched: Portal de Periódicos CAPES; Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); Google Scholar and Scientific Electronic Library Online (SciELO); with no time limit.

The analysis was carried out considering the bibliographic survey and the observation of the interfaces available on the web, which generated a discussion about aspects related to the methods and technologies that allow the DPLA to remain in operation together with its partners, thus providing opportunities for the debate about the possibilities and challenges of achieving data integration from digital cultural repositories.

| 4

---

[4] Available on: https://pro.dp.la/Hubs/membership-program.

3 INTEGRATION AND INTEROPERABILITY IN DIGITAL REPOSITORIES

The quest to expand the forms of information access and use is not a novelty for institutions such as libraries, archives and museums that continually seek to improve the available means to fulfill their social roles. With the development of the web, the possibilities of accessing and using information have expanded at different levels, allowing to go beyond availability through online catalog systems (MARCONDES, 2016), publishing and making available not only the description of the information in digital formats, but also the information resources or objects themselves.

In this scenario, digital repositories have brought new perspectives and possibilities for accessing and sharing information and information resources, configuring themselves as "[...] a form of digital object storage that has the ability to maintain and manage material over long periods and provide appropriate access" (VIANA; MÁRDERO ARELLANO; SHINTAKU, 2005, p. 3, our translation). Allied to the continuous development of technologies, these environments have become important means by which information circulates in an open way, is stored, managed, preserved, shared and retrieved, which only became viable from systems capable of generating interoperable processes.

In this regard, the web environment and its technological solutions have played an essential role as a means of communication across different systems that need to interoperate to work their potential together. However, achieving interoperability is not a simple task, involving, above all, processes, technologies and protocols so that, when data are transferred from one system to another, their integrity is guaranteed (MARTÍNEZ; LARA, 2007).

Furthermore, according to Candela et al. (2007), interoperability is a multidimensional property that applies to structures of environments that house digital collections from different domains and affects what the authors list as the six basic concepts of environments: Content, Functionality, User, Quality, Policy and Architecture. For this reason, interoperability has been widely discussed in the scientific literature over the years from different perspectives.

One of these perspectives, relevant to this study, is that the concept of interoperability involves more than the issue of interaction between computational components, that is, technical components. As Sayão and Marcondes (2008, p. 136, our translation) explain, in the context of cultural institutions, "[...] the concept of interoperability is complex and stratified, reflecting the diversity of views, the number of involved variables and the interdisciplinarity underlying it."

According to Arms (2002), interoperability is a process that aims, from components that are technically different and managed by different organizations, to build coherent services for users. This, according to the author, requires cooperation agreements at three levels: technical, content and organizational. The technical level involves the computational components so that information can be exchanged, that is, it concerns technologies such as protocols, structures and standards. The content level concerns semantic agreements on information interpretation, that is, it involves shared agreements to interpret representations of data and metadata. Regarding the organizational level, it refers to the rules for access, preservation and planning of collections and services, mainly involving authentication processes, rights and licenses.

This perspective provides discussion on interoperability from the point of view of the involved communities, as for them, the systems, or rather, the interfaces of these systems are better used by the target audience if they present a unified and coherent vision of heterogeneous information resources, allowing access to content from different sources, as well as promoting navigation in integrated environments (SAYÃO; MARCONDES, 2008, HYVÖNEN, 2012, SANTAREM SEGUNDO; SILVA; MARTINS, 2019).

| 5

3.1 Interoperability of heterogeneous data

To establish effective communication and information exchange across systems, the main challenge lies in eliminating data heterogeneity, which can be at a structural, syntactic and semantic level. Structural heterogeneity is related to the different ways data are organized into conceptual schemes. On the other hand, syntactic heterogeneity comes from different syntaxes, that is, languages assigned to corresponding concepts. And semantic heterogeneity considers the differences in meaning and interpretation attributed to the data (SHETH; LARSON, 1990, CRUZ; XIAO, 2005).

Structural and syntactic heterogeneity are directly related to technical and organizational interoperability, since different data sources can store their content in different formats and data locations or silos, which are distributed network databases, using structures with different characteristics, specific standards, norms and protocols, as is common among cultural institutions (HYVÖNEN, 2012). As for semantic heterogeneity, it mainly focuses on content interoperability, which occurs when there is a disagreement or difference in relation to the meaning, interpretation or intended use of the same or related data (SHETH; LARSON, 1990), a recurring fact if the different domains dealing with cultural heritage are considered.

These challenges turns the semantic treatment as one of the biggest challenges of integration, once the precise meaning of the data is often not explicit or is not well communicated due to variations in the ways of interpreting, representing and structuring domain knowledge. As Hyvönen (2012, p. 42) highlights, "[...] interoperability problems can be tackled effectively by using a single schema. However, different schemas are needed and used for different kinds of objects in portal applications dealing with cross-domain contents."

In the cultural heritage domain, this diversity of data types occurs mainly because each institution develops representation standards based on its specific needs. When on the web, these data in specific formats cannot interoperate with other systems, compromising or preventing their use. Therefore, the adoption of structures capable of exchanging information, overcoming structural, syntactic and semantic differences still remains one of the great discussions for areas that seek integration.

| 6

Overcoming the challenge of integration becomes a means of maximizing the value and the potential reuse of collections so that new knowledge can be generated, especially when considering cultural heritage data, which, by nature are heterogeneous, multilingual and derived from a variety of sources and situational events (HYVÖNEN, 2012). Therefore, there is a growing trend among cultural institutions to join efforts and seek ways to integrate collections from these structures capable of exchanging information, even using heterogeneous data. The scientific literature points to several moments when solutions were developed for this purpose, highlighting the technological solutions of the Semantic web that have further enhanced this objective.

3.2 Technological solutions for data integration and interoperability

In the web environment, the different factors related to interoperability have been resolved through standards recommended by internationally recognized agencies, such as the W3C (World Wide Web Consortium) and the DCMI (Dublin Core Metadata Initiative), which aim to promote representation and sharing of standardized and structured data from solutions based on languages, ontologies, data models, standards and protocols developed with the web environment in mind.

The recommendation of formats in XML language (eXtensible Markup Language) has been widely used over the years for factors related to syntactic standardization in data encoding due to its flexible, open and device-independent functionality (DOERR, 2003, SANTAREM SEGUNDO, 2004). However, other more modern languages have been used and incorporated in different web applications, such as Turtle (Terse RDF Triple Language) and JSON

(JavaScript Object Notation). Regarding the semantic factors, due to their greater complexity, the solutions involve conceptual models and ontologies, built based on the semantic principles of each domain (DOERR, 2003; HYVÖNEN, 2012).

As the interoperability factors are interdependent, the Resource Description Framework (RDF) is one of the main technological solutions recommended by the semantic web, as it is configured as a simple data model, which represents the web resources in terms of syntax and semantics of knowledge domain, enabling structural interoperability. Using the RDF-Schema (RDFS), it is possible to provide the necessary mechanisms for the declaration and definition of data properties and relationships, defining the specific characteristics of domains and their underlying semantics (DIAS; SANTOS, 2003). The format of these declarations consists of three elements: subject, predicate and object, called triples ("resource, property and value", or even "entity, attribute, value"), forming the so-called graphs (BIZER; HEATH; BENERS-LEE, 2009). The function of triples is to express a relationship between two digital resources, one of which is the resource/entity subject and the other the object/value, among them is the predicate, also called a property or attribute, which represents the nature of the relationship between resources (WORLD WIDE WEB CONSORTIUM, 2014).

Both resources and properties are described using a Uniform Resource Identifier (URI), which is a uniform identifier to uniquely identify a digital resource on the web, in such a way that resources can be associated through a network of relationships (WORLD WIDE WEB CONSORTIUM, 2006). In addition to RDF and URIs, which help in the various interoperability factors, ontologies are important tools to support data networks, since they allow building "[...] an organized relationship between terms within a domain, favoring the possibility of contextualizing the data, making the interpretation process more efficient and easier [...]" (SANTAREM SEGUNDO, 2015, p. 226, our translation).

Together with the conceptual models of each domain, ontologies allow the communication of content specifications in a shared way from the modeling of the semantic aspects of these domains. As a result, models and ontologies expand the use of metadata patterns, allowing the development of Application Profiles, which bring together a set of metadata selected from different schemas to a common schema (ZENG; QIN, 2016). In this context, the principles of Linked Open Data (LOD) stand out, since the publication of openly connected data enables the reuse and integration of various sources (HARPRING, 2016), optimizing data representation, minimizing structural differences, expanding collaboration possibilities, the forms of use and discovery by users.

Another important technological aspect to be considered for the integration of cultural content which affects the different interoperability factors is the description and visualization of images on the web. Many image digital objects are in "silos" with restricted access to customized applications and are locally built. Therefore, an important available solution is the International Image Interoperability Framework (IIIF), which defines Application Programming Interfaces (APIs[5]) to standardize image description and visualization to provide structured metadata, enabling any application or viewer compatible with the standard to use, share and display the images and their metadata with quality (IIIF, [2021]).

Therefore, these recommended technological solutions also affect, to a greater or lesser extent, the factors aimed at organizational interoperability, since they also influence the decisions of policies for access, use, management and preservation of cultural heritage and collections. Therefore, integration implies, in addition to finding ways to harmonize metadata formats, also defining mechanisms that guarantee that the content quality measures are interoperable with the quality measures of the participating systems (CANDELA et al., 2007).

Integrated environments, therefore, can be achieved in different ways, depending on the degree of interoperability and the results intended to be achieved, which, in turn, depend on

| 7

---

[5] Set of functions and guidelines used to interact with a computer program (software), enabling part of the functionality of a service or product on the web to be used on other platforms in the most assertive and convenient way for its users (POMERANTZ, 2015; SANTAREM SEGUNDO; SILVA; MARTINS, 2018).

the different levels of data source engagement, that is, of the involved participants. Thus, these factors provide several alternatives to achieve integration. The study by Santarem Segundo, Silva and Martins (2018) allows verifying data integration from the point of view of the technical possibilities of operation of the interoperable models from the protocols, which are a set of rules that define the communication across systems (SANTAREM SEGUNDO; SILVA; MARTINS, 2018). For this, the authors discuss the protocols' operation modes in four categories: aggregation; syndication (distribution – server-client); publication protocol; and distributed search. All categories present opportunities and challenges to be considered regarding the availability of cultural heritage and digital collections.

Two of them, "aggregation" and "distributed search", have been especially important for discussions about the integration of digital repositories, allowing the metadata retrieval and integration. Both approaches have been considered from the information retrieval perspective as important to combine data from different sources and provide them through a unified view on the web (MARCONDES; SAYÃO, 2001, HYVÖNEN, 2012), however using different strategies that can be verified by using the solutions employed to accomplish the task. Chart 1 compares these two types of approaches that enable data integration.

Chart 1. Description of the retrieval process for data integration.

	AGGREGATION	DISTRIBUTED SEARCH
DESCRIPTION	Action takes place beforehand in a separate preprocessing stage.	Action occurs during query processing.
PROCESS	First, the local collections are consolidated in a Data Warehouse through data harvesting, and then the query in a global database with a single and integrated interface is enabled.	A search query is sent to distributed local databases, where a protocol specifies data structures and interchange rules so that records identified as pertinent to that search are combined into a global hit list.
FEATURE	Separate content creation processes take place in each local database, and that content is gathered into a global data warehouse, which keeps them up to date as participating local databases evolve.	Participating organizations may continue to have independent local database systems, but must agree to the query protocol and to use compatible metadata formats.
FUNCTION	Protocols that facilitate the process of shareable metadata exposing and collecting so they can be encoded and used outside of their local environment.	Protocols that allow servers to receive federated search requests and respond with results.
EXEMPLES OF PROTOCOLS	OAI-PMH and OAI-ORE.	Z39.50 and web successors: SRU and SRW.

Source: Based on Marcondes and Sayão (2001), Hyvönen (2012) and Santarem Segundo, Silva and Martins (2018).

| 8

The first approach especially involves the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and Open Archives Initiative Protocol for Metadata Harvesting (OAI-ORE) protocols, in which communication takes place between a service provider, which makes requests, and a data provider, which responds by sending data. The OAI-PMH allows data providers to expose their structured metadata in a standard way, so that they can be collected by service providers making requests in the same standard (OPEN ARCHIVES INITIATIVE, [2021] a). Similarly, the OAI-ORE defines standards, but for the description and exchange of digital object aggregations, which combine various types of resources, such as text, images and video (OPEN ARCHIVES INITIATIVE, [2021b]). In this case, queries are performed in a global repository and users are redirected to the specific server when they request access to the original content (MARCONDES; SAYÃO, 2001).

The creation of data warehouses, that is, from a centralized base, presupposes that the items of a collection can be represented as interconnected web resources, using its technological solutions for semantic associations (HYVÖNEN, 2012). This strategy is suitable for retrieving large sets of metadata from digital repositories, and in this case collaboration and coordination are indispensable between these data providers to keep the data constantly synchronized and up-to-date.

Regarding "distributed search to different servers", also called "federated search", the queries occurred independently in each local database, which requires a standard protocol so that the results are consolidated in a single interface, providing satisfactory answers. This process usually happens due to legal reasons, which avoid integrating all collections into a central warehouse. The protocols that have been used the most on the web for this purpose are the Search and Retrieve URL (SRU) and the Search/Retrieve Web Service (SRW), both protocols are the result of an international collaborative effort to develop a standard interface search engine for the web, building on the functionality of Z39.50 (OCLC, 2021). This process is based on the client-server strategy, in which the client issues a query to the server (data provider), which is processed and then the client receives the data from the server.

Both approaches are therefore based on functional communication models that use technological solutions to enable data sources to respond to incoming requests, providing that data in a functional way. For this, metadata management services are widely used, such as REPOX software, which has a graphical interface for all the functionalities of an integration process, including several channels, such as OAI-PMH, HTTP, Z39.50, in order to of importing or retrieving data from data providers, services for transforming data between schemas according to specified rules, and services for sharing data (EUROPEANA PRO, 2015).

In this sense, the Digital Public Library of America (DPLA) initiative is considered one of the great examples of how diverse data sources can collaborate to form a functional network of interoperable data.

4 THE DIGITAL PUBLIC LIBRARY OF AMERICA (DPLA) INITIATIVE | 9

The Digital Public Library of America (DPLA) had its development driven by the ideal of an open free high quality national digital environment for public access. Counting on the efforts of different professionals, in October 2010, the DPLA planning process began. From a meeting with representatives of libraries, foundations, universities, among other partners in the city of Cambridge in the United States, the joint project to create an open and distributed network of online resources originated. In 2011, a process of intense organization began to define, design, and build the DPLA, a procedure that took two years. In that time, based on the Berkman Klein Center, workflows were created with various professionals such as librarians, innovators, digital humanists, and other volunteer professionals, led by a leading committee. Thus, in April 2013, the DPLA was launched and became a free, open and publicly accessible national digital platform (DPLA PRO, [2021a]).

During the DPLA pre-launch development period, the DPLA Metadata Application Profile (DPLA MAP) was created, a fundamental part of the infrastructure that supports the functioning of the platform. The DPLA MAP is configured in a data model, developed from the Europeana Data Model (EDM), using only part of it for the description of data and other existing technological solutions, such as Dublin Core (DC), Open Archives Initiative Object Resue & Exchange (OAI-ORE) and Resource Description Framework (RDF) (MATIENZO; RUDERSDORF, 2014). Since its first public version, the DPLA MAP has been updated, improved, and is now in its fifth version, published in 2017. The model integrates the specific experiences and data description needs of each community from the collection as well as aggregation. content metadata provided by partner institutions. The DPLA MAP is the basis for

the structured data in the DPLA Application Programming Interface (API) (DPLA MAP WORKING GROUP, 2017a), which allows broad data reuse by the platform users.

Therefore, the DPLA MAP is understood as: "[...] an application profile, or a set of metadata elements, taken from multiple schemas for a particular local use. It is also a semantic metadata model, or abstract structure that describes the relationships between different types of data about the same thing" (DPLA MAP WORKING GROUP, 2017a, p. 1).

As such, the DPLA MAP is more robust and abstract than a schema, or metadata standard such as the Dublin Core or other domain-specific descriptive standards. It consists of an Application Profile[6] that describes elements from entities and relationships using RDF, which allows the combination of schemas to adapt to specific needs (ZENG; QIN, 2016; DPLA MAP WORKING GROUP, 2017a). As it is an abstract metadata model, the DPLA MAP can be expressed in any standard encoding, and in the case of the DPLA API, the Java Script Object Notation (JSON) language for Linked Data, called JSON-LD[7] (DPLA MAP) is used. WORKING GROUP, 2017a). From this perspective, the objective of the DPLA MAP, as an Application Profile based on RDF, is to establish the relationship between the entities that characterize a content, such as authorship and creation date, in such a way as to present a rich and well-defined structured representation. For this, the DPLA MAP uses entities and classes to represent the data, and namespaces[8] to name the metadata fields.

The representation process is based on the RDF structure, as shown in Figure 1, in the form of triple graphs, indicating the subject, the predicate and the object (resource, property and value). The DPLA MAP presents model a set of specific properties for each resource in its abstract, called a class (POMERANTZ, 2015). Each class contains a list of possible properties based on sets of existing metadata elements such as Dublin Core and EDM, in addition to elements defined by the DPLA itself, indicated respectively by the namespaces: 'dc', 'edm', 'dpla' (DPLA MAP WORKING GROUP, 2017b).
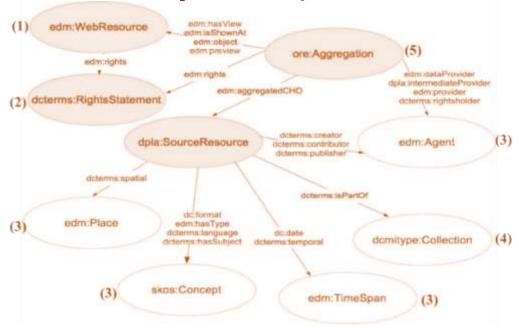
In the 'dpla:SourceResource' class are the properties that contain the descriptive metadata provided by the providers, such as title, date, format, which refer to the "origin resource", the digital objects.

| 10

---

[6] Defined set of metadata properties that combine selected elements from various standardized schemas together with locally defined ones. Policies and guidelines are also defined for a specific profile (ZENG; QIN, 2016, DPLA MAP WORKING GROUP, 2017b)

[7] Format that permits building additional mappings to JSON based on the RDF model (SANTAREM SEGUNDO; SILVA; MARTINS, 2018).

[8] Prefix that precedes the name of the metadata element or attribute indicating its origin (ARAKAKI, 2016).

Source: DPLA MAP Working Group (2017b).

Most of this class's properties are based on Dublin Core (DC) and link to other classes in DPLA MAP which, as observed in Figure 1: (1) store information about the digital version of the original resource (edm:WebResource); (2) store information about material rights and reuse (dcterms:RightsStatement); (3) allow improved description of certain fields (Agent, Concept, Place and TimeSpan); (4) gather information about locally defined sets or collections by providers (demtyope:Collection); and (5) pack all this information together in the 'ore:Aggregation' class. In the 'ore:Aggregation' class, important properties about the Hubs are stored, providing information about the location, the thumbnail (content view) and the Hubs' original metadata record.

This allows both metadata aggregation and the reuse of elements from different schemas previously published, reusing vocabularies, facilitating interoperability, and not duplicating information. In this way, this important functionality enables the DPLA to bring together the different data providers which, in turn, collaborate to make their content ready to use by adhering to the recommended standards, allowing the integration process to take place.

4.1 The DPLA Hubs

One of the great highlights of initiatives such as the DPLA's is its function of integrating information to enhance knowledge generation. For this reason, different institutions agree to make efforts and join resources to collaborate for the benefit of an integration interface. However, each institution has its own specificities and technologies to manage its data, which makes this level of interoperability and data aggregation proposed by the DPLA quite diverse and complex. Therefore, DPLA works with two types of Hubs or provider partners. They are categorized as Content Hubs and Service Hubs.

The Content Hubs providers comprise large cultural institutions, such as libraries, archives, and museums, characterized by sharing vast sets of data records, maintaining, and improving them as necessary. According to the DPLA documentation and Hub websites (which can be accessed via a hyperlink in each name). The current Content Hubs are:

1. **ARTstor:** Provides content for study, teaching and learning in the arts and associated fields, including high-quality images and media from museum, library, scholarly and artist collections, including rare materials in freely available open access collections.

| 11

RDBCI: Rev. Dig. Bibliotec e Ci. Info. / RDBCI: Dig. J. of Lib. and Info. Sci.| Campinas, SP | v.20| e022007 | 2022

2. **Biodiversity Heritage Library (BHL):** acts as a worldwide consortium of natural history, botanical and biodiversity research libraries.
3. **David Rumsey:** Cartographic materials, which include collections of atlases, wall maps, globes, school geographies, pocket wall maps, children's maps, manuscripts, exploration books, marine charts, and a variety of other cartography items.
4. **J. Paul Getty Trust:** dedicated to the presentation, conservation and interpretation of the world's artistic legacy related to the visual arts, providing access through its programs: Getty Conservation Institute, Getty Foundation, J. Paul Getty Museum and Getty Research Institute.
5. **Harvard Library**: provides free public access to part of the Harvard library's digital content, including its rare and special collections such as ancient art, manuscripts and audiovisual.
6. **Hathi Trust:** collaborative repository that provides access to digital content from the various academic and research library collections from services and programs that provide: temporary emergency access during service interruptions; the use of content for analysis through text search and data mining; collective retention printing and gives access to federal publications.
7. **Internet Archive:** gathers content published on the Internet and other artifacts in digital format, documenting and preserving the history of the web. Its collection mainly includes ephemeral materials such as websites, audio recordings, live shows, videos, television news programs, images and software.
8. **Library of Congress:** on its digital platform, provides access to books, recordings, photographs, newspapers, maps, and manuscripts from its collection, as well as reliable legal materials that support the actions of the US Congress.
9. **National Archives and Records Administration (NARA):** independent federal agency that preserves and shares records of US history, government and citizens, providing digital access to documents and materials created by the public administration.
10. **New York Public Library:** free provider that brings together libraries and branches in the New York City region, offering free digital access to a part of its collection that has materials such as books, videos, maps, manuscripts, illustrations, photos for users, from children to academic professionals.
11. **Smithsonian Institution:** comprises museums and galleries, a national zoo, research facilities and libraries that provides digital access to its collections and holdings, including books, periodicals, museum objects, manuscripts, images and videos.
12. **United States Government Publishing Office (GPO**): federal agency responsible for producing and distributing information products and services for all three branches of the Federal Government, providing free permanent public access to government information such as bills, laws, regulations, presidential documents, studies, and other federal documents through the Federal Depository Library Program (FDLP) and GovInfo.
13. **University of Southern California Digital Library (USCDL):** provides digital content provided by the 23 libraries and information centers part of the USC, supporting knowledge access, discovery, creation, and preservation produced and related to the university.
14. **University of Washington Libraries:** provides digital content such as photographs, maps, newspapers, posters, reports, and other media for the university's three campuses (Seattle, Tacoma, and Bothell) and Friday Harbor Laboratories.

Content Hubs are so called because they are mainly characterized by being large data aggregators from different types of digital repositories, maintaining a direct relationship with the DPLA, often providing data records in different Application Profiles, which are then transformed into the DPLA MAP (DPLA MAP WORKING GROUP, 2017b).

On the other hand, the Service Hubs are formed by state or regional contributing institutions organized in a network which collaborate to send the data records from their repositories to the DPLA. Similarly to Content Hubs, Service Hubs are aggregators, but they gather content by geographic region or scope. They share resources, roles, and responsibilities, so the contributing institution does not need to provide all services. Thus, smaller contributing institutions, which cannot afford to purchase or host their own repository system and digitization services, can work with their digital content collaboratively, and still meet the various aggregation requirements (DPLA PRO, [2021b]). These contributing institutions share their data with Service Hubs, which is subsequently shared with DPLA.

In this way, the Service Hubs play an important role in the DPLA's proposal to bring together the cultural heritage across the country and expose it digitally, as they connect partners of different sizes and origins, giving national and global visibility to this cultural content.

They bring together metadata that resolves to digital objects (online texts, photographs, manuscript material, artwork, etc.) from libraries, historical societies, archives, museums, and other cultural heritage institutions participating in their network, often hosting these resources locally, as well as sharing metadata and content previews (thumbnails, etc.) through DPLA. (DPLA PRO, [2021Bb], unpaged)

With this, Service Hubs also often play the role of repositories for these institutions, taking responsibility for the digitization, preservation, and long-term storage of their digital objects, as well as assisting for metadata, aiming to warrant quality, normalization, standardization, and improvement. Also, they get involved with the engagement of communities, helping in the development of technology, tools and professional development (DPLA PRO, [2021b]). With this, each Service Hub undertakes the following:

> - Representing their community (state, region, etc.) as the contact point for the DPLA and obtaining community buy-in on significant issues affecting their partners.
> - Aggregating their partners' metadata into a single standard and sharing it with DPLA through one harvestable data source.
> - Actively addressing metadata concerns (including copyright and licensing labeling) and working with partners on timely remediation.
> - Providing outreach to their partners and, with DPLA staff, developing local practitioners' capacity on topics such as open data, data quality and standards, copyright and licensing, and other relevant subjects.
> - Maintaining technologies (such as OAI-PMH, API, ResourceSync, etc.) that allow for standardized metadata to be shared with DPLA on a regular, consistent basis.
> - Engaging with the broader community of data creators, providers, and users, locally and nationally. (DPLA PRO, [2021b], unpaged).

The DPLA documentation indicates the current Service Hubs that provide data to the platform: | **13**

1. **Big Sky Country Digital Network:** brings together and provides access to digital content from cultural institutions in the geographic domain of Montana, North Dakota.
2. **California Digital Library (DCL):** brings together and provides access to digital content from the University of California partner institutions, both on campus and through external collaborations.
3. **Connecticut Digital Archive (CTDA):** is part of the University of Connecticut's Digital Preservation Repository Program, gathering and providing access to a wide range of digital resources for educational and cultural institutions as well as state agencies in Connecticut.
4. **Digital Commonwealth:** Provides resources and services to support the creation, management and dissemination of cultural heritage materials maintained by Massachusetts cultural institutions.
5. **Digital Library of Georgia:** Collaborates with Georgian educational and cultural institutions to provide access to digital resources and micrographic services on the history, culture and life of the state, supports the teaching, research and service missions of Georgia Library Learning Online (GALILEO) and the University System of Georgia.
6. **Digital Library of Tennessee (TEL):** Gathers and provides access to digital resources, including magazines, academic journals, podcasts, videos, e-books, test preparation materials, federal census records, and other Tennessee primary source materials.
7. **Digital Maine:** Maine State Library, Maine State Archive and other community institutions partnered to provide access to history across the state of Maine, ensuring transparency in government and sharing the stories of people and places. It has collections of maps, church records, local histories, genealogy research, community image reports, and other state agency reports, publications, and research.
8. **Digital Maryland:** collaborative digital preservation program for the state of Maryland, involving the University System of Maryland & Affiliated Institutions (USMAI), a consortium

of 17 public colleges and universities, providing access to historical and cultural documents, images, audio and videos that record the history of the state.

9. **Digital Virginias:** collaboration among organizations that originally formed a DPLA Content Hub, involving only the University of Virginia and currently incorporates regional partners George Mason University, William & Mary, Virginia Commonwealth University, Virginia Tech and West Virginia University to create a combined set of historical and cultural materials for the entire region of Virginia and West Virginia.

10. **District Digital:** Collaboration between the DC Public Library and the Washington Research Library Consortium to help bring together the digital collections of cultural institutions in and around the District of Columbia.

11. **Empire State Digital Network (ESDN):** network administered by the Metropolitan New York Library Council (METRO) in collaboration with eight regional councils of allied libraries that collaborate to aggregate content from existing projects of state and regional digital collections of cultural organizations across New York state.

12. **Green Mountain Digital Archive (GMDA):** collaboration between Middlebury College, Vermont State Archives & Records Administration, Vermont Historical Society, Rockingham Free Public Library, Norwich University, St. Michael's College, University of Vermont and Vermont Department of Libraries, which brings together photographs, documents, maps, recordings, and other digital resources relating to the state of Vermont.

13. **Illinois Digital Heritage Hub (IDHH):** composed of four institutions, the Chicago Public Library, the Consortium of Academic and Research Libraries in Illinois, the Illinois State Library, and the University of Illinois at Urbana-Champaign Library, which bring together the resources of the state of Illinois, including content from the Illinois Digital Archives and CARLI Digital Collections.

14. **Indiana Memory:** Collaboration among cultural institutions in the state of Indiana to provide access to the wealth of their primary sources available digitally.

15. **Kentucky Digital Library (KDL):** Collaborative initiative by the Kentucky Virtual Library (KYVL) to provide access to digital archive collections related to shared history and culture within the Kentucky community.

16. **Michigan Service Hub:** Collaboration among the Library of Michigan (LOM), the Midwest Collaborative for Library Services (MCLS), the University of Michigan (UM), Wayne State University (WSU), Michigan State (MSU) and Western Michigan University (WMU) to aggregate the digital collections of various institutions in the Michigan region.

17. **Minnesota Digital Library:** an initiative that brings together and offers access to unique digital collections shared by cultural heritage organizations across the state of Minnesota, including postcards, maps, letters and oral history records.

18. **Mississippi Digital Library:** A collaborative initiative by the state of Mississippi that provides an online space to research and explore cultural and historical content held by institutions and repositories in the state of Mississippi.

19. **Missouri Hub (MOHub):** affiliation of institutions that seek to give visibility and relevance to digital collections offered online in Missouri, aggregating information about digital objects.

20. **Mountain West Digital Library (MWDL):** collaborative initiative among cultural and educational institutions created by the Utah Academic Library Consortium.

21. **NJ/DE Digital Collective:** collective that aggregates data from libraries, museums, cultural heritage organizations and other institutions linked to the state of New Jersey and Delaware.

22. **North Carolina Digital Heritage Center (Digital NC):** State digital digitization and publishing program that works with cultural heritage institutions of all sizes in the state of North Carolina.

23. **Ohio Digital Network (ODN):** project developed in the state of Ohio to coordinate digitized collections and publish them online, from the Digitization Hubs (DigiHubs) program that involves the partners Columbus Metropolitan Library Public Library of Cincinnati and Hamilton County, Toledo Lucas County Public Library, and the Cleveland Public Library, supporting them from equipment usage to metadata creation.

24. **Oklahoma Hub (OK Hub):** partnership among the Oklahoma Department of Libraries, Oklahoma Historical Society, Oklahoma State University Library, and the University of Oklahoma Libraries, offering unique resources, particularly in the areas of Native American

history and culture, environmental sciences and agriculture, as well as the lives and experiences of generations in the state of Oklahoma.

25. **Orbis Cascade Alliance:** digital collection service that performs metadata cleansing, training, and support for their collections, including documentation development and implementation, in addition to being a data aggregator.

26. **Digital PA:** partnership involving libraries, historical societies, museums, universities, and other institutions across the state of Pennsylvania that brings together cultural heritage from collections and historical resources.

27. **Plains to Peaks Collective:** Partnership among Colorado & Wyoming State Libraries, supported by the Institute of Museum and Library Services (IMLS) to provide access to digital collections of the regions' cultural history.

28. **Portal to Texas History:** portal with rare, historical and primary source materials about the state of Texas, created and maintained by the libraries of the University of North Texas. It offers ethnically diverse collections such as the "Danish Heritage Preservation Society" and the San Antonio Public Library's "African American Funeral Programs" for use by scholars and the general public. In addition, it offers opportunities for small rural communities to preserve and access their history.

29. **Recollection Wisconsin:** project that brings together digital cultural heritage resources, including photographs, maps, letters, diaries, oral history records, artifacts, and other historical resources from more than 200 libraries, museums, and other cultural heritage institutions in the state of Wisconsin.

30. **South Carolina Digital Library (SCDL):** project that coordinates the distribution of resources needed to encourage digitization efforts and provides free and legally licensed access to online collections from more than 40 institutions in the state of South Carolina.

31. **Sunshine State Digital Network (SSDN):** collaborative network of digital collections involving cultural heritage organizations across the state of Florida, supported by the Library Services and Technology Act administered by the Florida Department of State, Division of Library and Information Services.

In addition to the aforementioned Hubs, DPLA reports that four projects are in progress. The Northwest Heritage Hub, the Orbis Cascade Alliance (Washington/Oregon), the NJ/DE Digital Collective and the New Hampshire Digital Library.

With highly heterogeneous data, each of the Hubs has its own specificities, technologies and levels of governance that best suit the reality of their communities, and the DPLA assists them to adopt continuous workflows of aggregation, normalization, and provision of metadata so that they remain engaged in the collaborative effort of data integration.

4.2 The aggregation effort performed by the DPLA Hubs

As an objective, this study proposed to present an overview of the technological solutions underlying the collection and data availability from the DPLA partners. Therefore, to delimit the analysis, the Hubs were divided in two groups.

The first group is the DPLA partners, members who pay an annual fee for the partnership, obtaining benefits in return, mainly in relation to the DPLA tools and services, such as the development of APIs for maintenance and access to enhanced data, which allows reuse by developers, researchers, and other stakeholders inside and outside the network. In addition, partners have access to the DPLA-mapped and enriched metadata records, including improved geolocation and data cleansing to meet standards.

DPLA also provides partners with: a regular metadata ingestion schedule based on the frequency of changes to the Hub's data; access to analytical data; participation in working groups and task forces and in content curation projects such as exhibits and navigation lists; networking to solve common problems collaboratively; participation in the Hub Wiki Network website to support individual and collective work, facilitating communication; providing training, documentation, quality assurance testing, and consulting to support initial ingestion; availability of resources to create a local DPLA site. The services make most of these Hubs

| **15**

present similar solutions for data collection, aggregation, and provision as they follow the DPLA recommendations.

The second group is non-members, data providers that do not participate of the affiliation program. These Hubs generally use locally defined tools and services for management and aggregation, which gives their data different formats.

Nevertheless, these data can still be ingested and integrated by the DPLA, as, regardless of the group of providers, all metadata, both those created locally and those provided by partners, are delivered to the DPLA in a single flow, in which, regardless of the original schema, the data will be mapped to a single schema (e.g., Dublin Core) and structured according to the OAI-PMH standard to be compatible for collection. Then, after being collected by the DPLA, the metadata can be transformed into DPLA Metadata Application Profile (MAP) structure and stored and published via JSON-LD (DPLA, [2021]).

Although the OAI-PMH is still the Hubs' most used type of feeder, allowing them to provide data in the simple or qualified Dublin Core standard, or even in the Metadata Object Description Schema (MODS), institutions are increasingly looking for the specification ResourceSync[9], an OAI-PMH enhancement that allows synchronization of both metadata and digital objects (SOMPEL, 2014).

For the data provision process, some factors are highlighted by the DPLA: a) the importance that the Hubs and contributing institutions share and maintain the location links - URL (Uniform Resource Locator) for the original items and records, as well as thumbnails (images) that represent the content in their local collections; b) make sure that the mandatory fields, such as "rights" and "name of the institution" that contributed to the registration, are always duly filled in; c) maintain metadata consistency across all datasets in a single Hub so that it is always structured and interpreted in the same way across all of its collections; and d) all records shared with the DPLA must be available without restriction, under standardized rights statements and recommendations for the use of statements or under a Creative Commons license (DPLA, 2017).

Even following these guidelines for data provision, the second group exhibits some factors due to their independent systems. These factors were systematized and presented in Table 2, in which each Hub is described according to its main aggregation functionalities according to the information available on its websites.

| 16

Table 2. Main features of non-member Hubs.

DATA PROVIDER		FUNCTIONALITY
Content Hubs	ARTstor	Mainly uses the JSTOR tool to catalog and share public collections, and the IIIF image viewer to make the images available. Provides the Data for Research (DfR) program that enables text analysis and research data in digital humanities.
	Biodiversity Heritage Library (BHL)	Uses Global Names Architecture taxonomic intelligence tools for scientific names, produces stable URLs and assigns DOIs (digital object identifiers), offers data exports via APIs and in MODS, BibTex and RIS formats, making its digital content available in open access under Creative Commons license.
	David Rumsey	Uses Luna Imaging's LUNA software to make digitized content available in open access under a Creative Commons license, which provides a presentation API for handling and IIIF support of cross-collection search results, media groups, and formats for display.

---

[9] Specification that describes a synchronization framework for the web, allowing third-party systems to stay in sync with a server's evolving resources (OPEN ARCHIVES INITIATIVE, 2017).

Data Providers	J. Paul Getty Trust	Makes use of Arches open-source software, developed for data management, discovery, visualization and also for projects and activity flow. It uses the Quire tools, for publishing multi-format text, and IIIF for accessing images.
	Hathi Trust	Uses the API called HathiTrust Data API to retrieve digital objects, including scanned or native digital content, composed of images and OCR (Optical Character Recognition) text and the corresponding metadata. Uses the Kakadu software source code license for image creation, maintenance, and delivery.
	Internet Archive	Uses the functionality of the Archive-it service, its own program that allows archiving web content, providing tools, training, and technical support to capture and preserve dynamic materials on the Internet. Uses APIs to make metadata accessible through the Internet Archive Wayback Machine portal.
	Library of Congress	Uses MODS as a metadata object description schema and offers cross-mapping between MODS and DC to enable data aggregation.
	University of Southern California Digital Library (USCDL	Uses the services of CONTENTdm, OCLC's digital collections software, which supports Qualified DC and the IIIF Image and Presentation API. All items in the collections include metadata about copyright, licenses, and material usage.
	University of Washington Libraries	Uses OCLC services, including CONTENTdm software, and the Dublin Core standard for metadata.
Service Hubs	Digital Maine	Uses the Digital Commons services, repository software to publish, manage and openly make available the full spectrum of digital collections that offers geolocation, timelines, customizable metadata, and image galleries that allow for panning and zooming.
	Digital Maryland	Provides little documentation and a set of its own guidelines in a document named Metadata Style Guide, available on the website, which define the metadata format accepted by the platform.
	Kentucky Digital Library (KDL)	Uses OCLC's CONTENTdm integrated tools, which support mobile viewing, support the IIIF standard and other custom interfaces by integrating collections via API.
	South Carolina Digital Library (SCDL)	Uses the SCDL Metadata Schema built upon the Dublin Core Metadata Initiative (DCMI) and OCLC best practices for CONTENTdm, in addition to other OAI-PMH compliant repositories. Also based on some of the DPLA guidelines.

Source: The authors.

| 17

The analysis of Table 2 allows for an overview of the technological solutions used in the collection and availability of data from DPLA partners, giving the possibility to point out some aspects of the opportunities and challenges that data providers face in the search for integrated access.

In terms of opportunities, partner institutions gain benefits from partnerships, not only at the highest level, which is the relationship with the DPLA, but at the level of the partner institutions themselves which, by collaborating, they develop an important work of giving visibility to cultural content from different sources, expanding access to and preservation of history and their community memory. The challenges, on the other hand, are shown at different levels. As seen, issues related to universally known problems for data interoperability involving the technical part are present, which can be solved by choosing the available and appropriate

technologies for each case. As for the content, the conflicts in interpreting each different type of digital object, whether they are digitized or born digital are observed.

Finally, influencing all these issues are the challenges related to governance. Each institution makes its own management and interoperability choices, which often creates conflicts when seeking to integrate their efforts. Therefore, it is reaffirmed that one of the main recommendations to overcome these challenges is that, from the beginning of a digital collection project, institutions must be concerned with aspects related to interoperability, avoiding isolation in single systems. Furthermore, it is evident that good practices in the creation, use or reuse of metadata patterns and schemas directly impact the integration capability.

Therefore, the importance of implementing digital curation approaches that guide the different levels of management, from planning to use and reuse, is highlighted, aiming to warrant both the ability to exchange data, and, above all, to ensure that these data are valid and useful for different contexts of use.

5 FINAL CONSIDERATIONS

The proposal of this study, although not exhaustive, allowed the construction of inferences about the possibilities and challenges of data integration, through the information provided, both by the DPLA and by its Hubs. This information was validated along with the theoretical references, leading to the results and the discussions.

In this context, the most explicit finding is that great efforts are taken to integrate data. This is evidenced by the number of available solutions and the constant concern to improve them to ensure functional interoperability, consistency of workflows with the evolution of the web, data quality, including metadata and digital objects. This allows us to reflect that there are different possible dimensions to discuss data integration work. The DPLA study provided an overview of these possibilities, focusing on the processes inherent to the data aggregation from digital repositories of cultural heritage and collections.

| 18

The study found that these processes involve the necessary technological solutions for the computational functionalities and the engagement of data providers. All the DPLA Hubs feature programs to encourage and support the digitization and documentation of content, as well as guidelines to ensure quality and legal access rights. In due course, the study made it possible to highlight the important role of the DPLA in terms of improving users' experience with cultural heritage and collections, disclosing the possibility of multi-dimensioning the ways of transforming information into knowledge.

Thus, the importance of initiatives such as those of the DPLA and Europeana, among others, is highlighted. They promote and give access to their guidelines and technologies in free open access, so that new initiatives can be based to create their own systems, as well as to create interoperable systems based on these standards. Brazil is an example that has sought to develop in this direction, through the "Programa Acervo em Rede", which promotes the digitization and documentation of the museum collections of the Brazilian Institute of Museums (Ibram). With the "Tainacan Project", the first results of the initiative can already be verified from the repositories of the museums of the Ibram network, which publish their collections with the free software Tainacan, a solution based on WordPress. The next steps of the initiative are towards the provision of an integrated search portal, gathering museum collections and allowing data exchange in the Brazilian culture dimension.

REFERENCES

ARAKAKI, F. A. Linked data: ligação de dados bibliográficos. 2016. 144 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016. Available at: http://hdl.handle.net/11449/147979. Access on: 02 jun. 2021.

ARMS, W. A. et al. A Spectrum of interoperability, the site for science prototype for the NSDL. D-Lib magazine, n.8 v.1, 2002. Available at: http://www.dlib.org/dlib/january02/arms/01arms.html. Access on: 02 jun. 2021.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data: the story so far. In: SHETH, A. P. (Ed.). Semantic services, interoperability and web applications: emerging concepts. Hershey: Information Science Reference, 2009. Available at: http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf. Access on: 02 jun. 2021.

CANDELA, L., et al. The DELOS Digital Library Reference Model. Foundations for Digital Libraries (version 0.98). DELOS Network of Excellence on Digital Librarie, nov. 2007. Available at: http://www.delos.info/index.php?option=com_content&task=view&id=345. Access on: 02 jun. 2021.

CRUZ, I. F.; XIAO, H. The role of ontologies in data integration. Engineering intelligent systems for electrical engineering and communications, Bentley, WA, v. 13, n. 4, p. 245, 2005. Available at: https://www.cs.uic.edu/~advis/publications/dataint/eis05j.pdf. Access on: 02 jun. 2021.

DIAS, T. D.; SANTOS, N. Web semântica: conceitos básicos e tecnologias associadas. Cadernos do IME-Série Informática, v. 14, p. 80-92, 2003. Available at: https://www.e-publicacoes.uerj.br/index.php/cadinf/article/view/6619/4734. Access on: 02 jun. 2021.

DOERR, M. DCC Digital Curation Manual: instalment on ontologies. Edimburgo, Escócia: Digital Curation Manual. 2008. Available at: http://hdl.handle.net/1842/3341. Access on: 02 jun. 2021.

DPLA. DPLA Standardized Rights Statements Implementation Guidelines, 2017. Available at: https://bit.ly/3KijSsP. Access on: 02 jun. 2021.

DPLA. Partnering with DPLA, [2021]. Available at: https://docs.google.com/document/d/1gshwQ0Oj84l5q-_JxHo7wjyLeXln9lCg6SUtIMkPu_g/edit#. Access on: 02 jun. 2021.

DPLA PRO. History, [2021a]. Available at: https://pro.dp.la/about-dpla-pro/history. Access on: 02 jun. 2021.

DPLA PRO. Becoming a Service Hub: service Hub models, roles, and responsibilities, [2021b]. Available at: https://pro.dp.la/prospective-Hubs/becoming-a-service-Hub. Access on: 02 jun. 2021.

EUROPEANA PRO. Apresentando REPOX: uma ferramenta para gerenciar espaços de metadados, 2015. Available at: https://pro.europeana.eu/post/introducing-repox-a-tool-to-manage-metadata-spaces. Access on: 02 jun. 2021.

HYVÖNEN, E. Publishing and using cultural heritage linked data on the semantic web. EUA: Morgan & Claypool Publishers, 2012.

IIIF. About IIIF, [2021]. Available at: https://iiif.io/about/. Access on: 02 jun. 2021.

ISOTANI, S.; BITTENCOURT, I. I. Dados Abertos Conectados: em busca da web do conhecimento. Novatec Editora, 2015. Available at: http://pgcl.uenf.br/arquivos/dadosabertosconectados_011120181613.pdf. Access on: 02 jun. 2021.

MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. Perspect. Ciênc. Inf. 2016, v.21, n.2, pp.61-83. Available at: http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2735/1748. Access on: 02 jun. 2021.

MARCONDES, C. H. Relacionamentos culturalmente relevantes para interligar objetos do patrimônio digital na web usando tecnologias de dados interligados. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19, 2018, Londrina, PR. Anais [...]. Londrina, PR: Universidade Estadual de Londrina, 2018. Available at: http://hdl.handle.net/20.500.11959/brapci/102416. Access on: 02 jun. 2021.

MARCONDES, C. H.; SAYÃO L. F. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. Ciência da Informação, Brasília, v. 30, n. 3, p. 24-33, 2001. Available at: http://revista.ibict.br/ciinf/article/view/909. Access on: 02 jun. 2021.

MATIENZO, M. A.; RUDERSDORF, A. The Digital Public Library of America ingestion ecosystem. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE & METADATA APPLICATIONS, Austin, Texas, USA, 2014. Proceedings […]. Austin, Texas: DCMI, 2014. Available at: https://dcpapers.dublincore.org/pubs/article/view/3700. Access on: 02 jun. 2021.

MARTÍNEZ, J.; LARA, P. La interoperabilidad de la información. Barcelona: Editorial UOC, 2007.

OCLC. SRW/U, 2021. Available at: https://www.oclc.org/research/areas/data-science/srw.html. Access on: 02 jun. 2021.

OPEN ARCHIVES INITIATIVE. ResourceSync Framework Specification (ANSI/NISO Z39.99-2017), 2017. Available at: http://www.openarchives.org/rs/1.1/resourcesync. Access on: 02 jun. 2021

OPEN ARCHIVES INITIATIVE. Open Archives Initiative Protocol for Metadata Harvesting, [2021a]. Available at: https://www.openarchives.org/pmh/. Access on: 02 jun. 2021.

OPEN ARCHIVES INITIATIVE. Open Archives Initiative Protocol for Metadata Harvesting, [2021b]. Available at: https://www.openarchives.org/ore/. Access on: 02 jun. 2021.

POMERANTZ, J. Metadata. Cambridge, Mass: MIT Press, 2015.

SANTAREM SEGUNDO, J. E. Recursos tecno-metodológicos para descrição e recuperação de informações na web. 2004. 157 f. Dissertação (Mestrado em Ciência da Informação) –

| 20

Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2004. Available at: http://repositorio.unesp.br/handle/11449/93618. Access on: 02 jun. 2021.

SANTAREM SEGUNDO, J. E. Web semântica, dados ligados e dados abertos. Tendências da Pesquisa Brasileira em Ciência da Informação, v. 8, n. 2, 2015. Available at: http://revistas.ancib.org/index.php/tpbci/article/view/359/359. Access on: 02 jun. 2021.

SANTAREM SEGUNDO, J. E. S.; SILVA, M. F.; MARTINS, D. L. Revisitando a interoperabilidade no contexto dos acervos digitais. Informação & Sociedade, v. 29, n. 2, 2019. Available at: https://doi.org/10.22478/UFPB.1809-4783.2019V29N2.38107. Access on: 02 jun. 2021.

SAYÃO, L. F. Digitalização de acervos culturais: reuso, curadoria e preservação. In: SEMINÁRIO SERVIÇOS DE INFORMAÇÃO EM MUSEUS, 4, 2016, São Paulo. Anais [...]. São Paulo: [s.n.], 2016, p. 47-61. Available at: https://www.researchgate.net/publication/319403030_Digitalizacao_de_acervos_culturais_reuso_curadoria_e_preservacao. Access on: 02 jun. 2021.

SAYÃO, L. F.; MARCONDES, C. H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. Transinformação, v. 20, p. 133-148, 2008. Available at: https://www.scielo.br/j/tinf/a/LSxTfhK6NfX54t4ypBK87kM/?lang=pt#. Access on: 02 jun. 2021.

SHETH, A. P.; LARSON, J. A. Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys (CSUR), v. 22, n. 3, p. 183-236, 1990. Available at: https://dl.acm.org/doi/pdf/10.1145/96602.96604. Access on: 02 jun. 2021.

SOMPEL, H. V. ResourceSync. St. Louis, MO: CNI 2014 Spring Membership Meeting. jun. 2014. Available at: https://www.niso.org/standards-committees/resourcesync. Access on: 02 jun. 2021.

VIANA, C. L. M; MÁRDERO ARELLANO, M A.; SHINTAKU, M. Repositórios Institucionais em Ciência e Tecnologia: uma experiência de customização do Dspace. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 3, 2013. Anais[...] São Paulo, Brasil, 2005. Available at: http://hdl.handle.net/10760/8168. Access on: 02 jun. 2021.

WORLD WIDE WEB CONSORTIUM. RDF 1.1 Primer, 2014. Available at: https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/. Access on: 02 jun. 2021.

ZENG, M. L.; QIN, J. Metadata. 2.ed. Chicago, IL: ALA Neal-Schuman, 2016.

| 21

RDBCI: Rev. Dig. Bibliotec e Ci. Info. / RDBCI: Dig. J. of Lib. and Info. Sci.| Campinas, SP | v.20| e022007 | 2022