

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

Maria Lígia Triques¹ , Paula Regina Ventura Amorim Gonzalez² 
Ana Cristina de Albuquerque³ 

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 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 eles enfrentam na busca pelo acesso integrado. **Metodologia:** A pesquisa se caracteriza como bibliográfica, qualitativa e exploratória, com análise baseada na discussão da literatura combinados com os resultados obtidos na análise das plataformas. **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 como 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.

| 1

PALAVRAS-CHAVE

Interoperabilidade. Repositórios digitais. *Digital Public Library of America*. DPLA.

Correspondência do autor

¹ Universidade Estadual de Londrina,
Londrina, PR, Brasil /
e-mail: mligia.triques@uel.br

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

³ Universidade Estadual de Londrina,
Londrina, PR, Brasil /
e-mail: albuati@uel.br

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

ABSTRACT

Introduction: Digital Public Library of America (DPLA) is an initiative that aims to expand the reach of its content and the possibilities for improving and using data to generate new knowledge from data integration. The search integration, however, results in a series of possibilities and challenges to which digital repositories, data providers and servers need to adapt. **Objective:** Present an overview of the technological solutions underlying the collection and availability of data from DPLA's partner

repositories, called Hubs, to point out aspects of the interoperability opportunities and challenges they face in the search for integrated access. **Methodology:** Qualitative, theoretical, and descriptive-exploratory research, using Content Analysis as a method. **Results:** Technological solutions adopted by the DPLA to collect and aggregate data are presented as results, as well as the Hubs, data providers, which use different solutions to aggregate data from the various contributing cultural institutions and supply them to the DPLA. **Conclusion:** It is concluded that despite the technical, semantic, and managerial challenges, there are great efforts to adopt solutions and best practices aimed mainly at engaging data providers. Both the Hubs and their contributing institutions have programs to encourage and aid digitization and documentation, as well as guidelines to guarantee the quality and legal rights of access to the content, which is essential to improve the users' experience and multi-dimension possibilities to transforming information into knowledge.

KEYWORDS

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

CRediT

- **Reconhecimentos:** Não aplicável.
- **Financiamento:** Este estudo foi financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).
- **Conflitos de interesse:** Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- **Aprovação ética:** Sim.
- **Disponibilidade de dados e material:** O conjunto de dados está disponível no apêndice do próprio artigo.
- **Contribuições dos autores:** Conceituação, Curadoria de Dados, Aquisição de Financiamento, Investigação, Metodologia, Recursos, Supervisão, Validação, Visualização, Escrita – rascunho original, Escrita – revisão & edição: TRIQUES, M. L.; Conceituação, Metodologia, Administração de Projetos, Recursos, Supervisão, Validação, Escrita – revisão & edição: GONÇALEZ, P. R. V. A.; Supervisão, Recursos, Validação, Escrita – revisão & edição: ALBUQUERQUE, A. C.

| 2



JITA: HS. Repositories



Artigo submetido ao sistema de similaridade

Submetido em: 13/09/2021 – Aceito em: 21/01/2022 – Publicado em: 06/03/2022

1 INTRODUÇÃO

Os acervos e as coleções culturais constituem-se como importantes registros do conhecimento e, como tais, são valiosas fontes de informação que representam os objetos ou testemunhos, materiais ou imateriais, da tradição ou da manifestação de um povo. Por meio da disponibilização em ambientes digitais, principalmente em repositórios, esses acervos e coleções têm sido cada vez mais encontrados em formatos digitais, a partir da criação e do compartilhamento de seus dados de forma aberta e distribuída na web (HYVÖNEN, 2012). Os dados de acervos e coleções culturais representam as várias formas de expressão dos povos, refletindo a diversidade de costumes, práticas cotidianas, crenças e linguagens.

Esses conjuntos de dados diversificados passam a representar uma parcela dos grandes conjuntos. Eles podem se beneficiar da interface web a fim de potencializar suas redes semânticas, a partir do que Marcondes (2018) denomina “links semânticos”, possíveis relacionamentos culturalmente relevantes que os recursos informacionais culturais estabelecem entre si, no processo de vinculação de seus dados.

Diferentes soluções tecnológicas criadas para as aplicações na web têm viabilizado e facilitado essa rede de relacionamentos semânticos, principalmente com o conjunto de tecnologias que propiciam o *Linked Data* e o *Linked Open Data*, no caso de dados conectados sob licenças gratuitas, abertas e públicas. Ambos consistem em boas práticas, fundamentadas em padrões internacionais recomendados pelo *World Wide Web Consortium* (W3C), como o *Uniform Resource Identifier* (URI) e o *Resource Description Framework* (RDF), entre outras, que permitem a criação de conjuntos de dados conectados (ISOTANI; BITTENCOURT, 2015). Mesmo com a constante evolução das soluções tecnológicas potencializando o uso e a abrangência dos acervos e das coleções digitais, a interoperabilidade ainda continua a ser um dos principais eixos de discussão em vários aspectos que envolvem os estudos e as implementações de conjuntos de dados, principalmente no caso da integração. Como destacam Santarem Segundo, Silva e Martins (2018), a informação e o conhecimento são requisitos-chave no contexto contemporâneo e nele a integração torna-se indispensável para transformar informação em conhecimento de maior alcance.

Nessa perspectiva, pode-se dizer que, quanto ao acesso, uso e reúso, a integração também possibilita transformar informação em conhecimento com melhor qualidade, ao permitir a ampliação e o enriquecer as fontes de dados, bem como as informações e as conexões semânticas entre eles, tendo em vista a comunidade e o público-alvo. Marcondes (2018) ao discorrer sobre a implementação de relacionamentos culturais relevantes como interoperabilidade entre coleções digitais, ampliação de sua visibilidade e usabilidade e, autonomia de instituições de memória, faz referência aos “links semânticos”.

A integração, exige que ocorra a troca de dados de forma interoperável, ou seja, sem perda de funcionalidade e sentido. A consolidação de algumas dessas estruturas permitiu, entre outras aplicações, o desenvolvimento de serviços e espaços de interação para os acervos e coleções culturais, possibilitando o desenvolvimento de processos distribuídos e interoperáveis de criação, bem como de compartilhamento de conteúdo, a exemplo da grande agregadora de dados Europeia¹ (HYVÖNEN, 2012; SAYÃO, 2016). No contexto norte-americano, uma outra iniciativa semelhante à Europeia ganhou destaque por seus esforços em unificar fontes de dados culturais em nível nacional. A *Digital Public Library of America*², conhecida pelo acrônimo DPLA, é uma plataforma aberta que permite a busca em uma única interface do conteúdo cultural norte-americano disponibilizado na web por meio de diferentes repositórios digitais integrados.

Tanto a DPLA como a Europeia adotam tecnologias e padrões para cumprir suas

¹ Disponível em: <https://europeana.eu/pt>.

² Disponível em: <http://dp.la/>.

missões e objetivos, fundamentando-se nas tecnologias da web Semântica e nos princípios do *Linked Open Data*. No caso da DPLA, sua funcionalidade se dá a partir de uma estrutura própria denominada *DPLA Metadata Application Profile* (DPLA MAP³), que permite integrar as experiências, bem como necessidades específicas de descrição de cada comunidade, a partir da coleta e da agregação dos metadados das instituições parceiras ou, como são chamados, os “*Hubs*”, membros dos quais a DPLA colhe dados (MATIENZO; RUDERSDORF, 2014).

Os *Hubs* são os fornecedores dos dados e os disponibilizam por meio de sistemas que coletam dados de uma variedade de repositórios digitais, mantidos tanto por grandes instituições nacionais, como também por instituições menores, que cooperam por região geográfica ou escopo para compilar seus registros de dados. Em ambos os casos, os *Hubs* comprometem-se a desenvolver e manter uma infraestrutura apropriada a fim de dar suporte à agregação em escala e complexidade nacional. Desse modo, a DPLA preocupa-se em estabelecer as diretrizes para que os membros sejam capazes de fornecer dados interoperáveis mesmo utilizando sistemas com particularidades em termos de interfaces, padrões, protocolos, normas de proteção intelectual e, principalmente, de interpretação de dados (SAYÃO; MARCONDES, 2008).

Portanto, é importante considerar que a integração de dados pode ser alcançada de várias formas e em diferentes níveis, assim como a interoperabilidade, o que resulta em uma série de possibilidades e desafios para os quais os provedores de dados precisam se adaptar. Assim, considerando-se a iniciativa norte-americana pela perspectiva do acesso integrado, questiona-se quais têm sido os caminhos percorridos pelos repositórios participantes para viabilizar o acesso integrado aos acervos e às coleções culturais. O objetivo norteador desse estudo é, apresentar um panorama das soluções tecnológicas subjacentes à coleta e disponibilização de dados dos repositórios parceiros da DPLA, de modo a pontuar alguns aspectos das oportunidades e dos desafios que enfrentam na busca pelo acesso integrado. | 4

2 METODOLOGIA

A metodologia utilizada neste trabalho é de caráter bibliográfico e qualitativo, configurando-se como uma pesquisa exploratória. Parte da análise da plataforma que promove o acesso integrado a acervos e coleções culturais disponibilizados em repositórios digitais. A DPLA apresenta ampla divulgação de seus sistemas de funcionamento, o que a torna apropriada para a análise, com uma vasta documentação e bibliografia disponíveis.

Em funcionamento desde abril de 2013, a DPLA oferece em sua plataforma na web o “DPLA Pro”, espaço on-line no qual os vários profissionais envolvidos se conectam e compartilham informações a respeito da plataforma. Nesse espaço, encontra-se a documentação que informa, orienta, capacita as instituições parceiras e demais interessados a contribuir com a iniciativa, bem como a trabalhar com a estruturação dos dados. Além disso, cada *Hub* provedor de dados da DPLA também apresenta documentação própria, o que permitiu coletar os dados necessários para este estudo.

Como a DPLA trabalha com quase 50 *Hubs* em sua lista total de parcerias e cada um desses *Hubs* com dezenas de instituições contribuintes, este estudo dividiu a análise dos provedores de dados em dois grupos: os *Hubs* que fazem parte da *DPLA’s Membership Program*⁴, programa de sócios que utilizam principalmente as ferramentas desenvolvidas pela DPLA; e os não sócios, que apresentam sistemas próprios diversificados sem utilizar as tecnologias fornecidas da DPLA e, mesmo assim, conseguem integrar seus dados de modo efetivo. Desse modo, buscou-se limitar a abrangência da análise, descrevendo individualmente

³ Disponível em: <http://dp.la/map>.

⁴ Disponível em: <https://pro.dp.la/Hubs/membership-program>.

apenas as soluções dos *Hubs* não sócios, uma vez que os *Hubs* sócios seguem as diretrizes da DPLA.

Em relação ao levantamento bibliográfico, foram pesquisadas as seguintes fontes: Portal de Periódicos CAPES; Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI); *Google Scholar* e *Scientific Electronic Library Online* (SciELO); sem delimitação de tempo.

A análise foi feita considerando-se o levantamento bibliográfico e a observância das interfaces disponibilizadas na web que gerou uma discussão acerca dos aspectos relacionados aos métodos e às tecnologias que permitem a DPLA manter-se em funcionamento em conjunto com seus parceiros, oportunizando, assim, o debate a respeito das possibilidades e dos desafios de se atingir a integração de dados de repositórios digitais culturais.

3 INTEGRAÇÃO E INTEROPERABILIDADE EM REPOSITÓRIOS DIGITAIS

A busca por ampliar as formas de acesso e uso às informações não é novidade para as instituições como bibliotecas, arquivos e museus que, continuamente, procuram aprimorar os meios disponíveis a fim de cumprir seus papéis sociais. Com o atual desenvolvimento da web, as possibilidades de acesso e uso às informações ampliaram-se em diferentes níveis, permitindo ir além da disponibilização por meio de sistemas de catálogos *on-line* (MARCONDES, 2016), publicando e disponibilizando não somente a descrição das informações em formatos digitais, mas também os próprios recursos ou objetos informacionais.

Nesse cenário, os repositórios digitais trouxeram novas perspectivas e possibilidades de acesso e compartilhamento das informações e dos recursos informacionais, configurando-se como “[...] uma forma de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar material por longos períodos e prover o acesso apropriado” (VIANA; MÁRDERO ARELLANO; SHINTAKU, 2005, p. 3). Aliados ao contínuo desenvolvimento das tecnologias, esses ambientes transformaram-se em importantes meios pelos quais a informação circula de forma aberta, é armazenada, gerenciada, preservada, compartilhada e recuperada, o que só se tornou viável a partir de sistemas capazes de gerar processos interoperáveis.

Nesse quesito, o ambiente web e suas soluções tecnológicas têm cumprido um papel essencial como meio de comunicação entre os diferentes sistemas que precisam interoperar para trabalharem em conjunto o seu potencial. No entanto, alcançar a interoperabilidade não é uma tarefa simples, envolvendo, sobretudo, processos, tecnologias e protocolos para que, quando os dados forem transferidos de um sistema para outro, garanta-se sua integridade (MARTÍNEZ; LARA, 2007).

Além disso, de acordo com Candela *et al.* (2007), a interoperabilidade é uma propriedade multidimensional que se aplica a estruturas dos ambientes que abrigam acervos e coleções digitais de diferentes domínios e afeta o que os autores elencam como os seis conceitos básicos dos ambientes: Conteúdo, Funcionalidade, Usuário, Qualidade, Política e Arquitetura. Por essa razão, a interoperabilidade tem sido bastante discutida na literatura científica ao longo dos anos sob diferentes perspectivas.

Uma dessas perspectivas, pertinente a este estudo, é a de que o conceito de interoperabilidade envolve mais do que a questão da interação entre componentes computacionais, ou seja, técnicos. Como explicam Sayão e Marcondes (2008, p. 136), no contexto das instituições culturais, “[...] o conceito de interoperabilidade é complexo e estratificado, refletindo a diversidade de visões, o número de variáveis envolvidas e a interdisciplinaridade que está subjacente e ele.”

Conforme Arms (2002), a interoperabilidade é um processo que visa, a partir de componentes que são tecnicamente diferentes e gerenciados por diferentes organizações, construir serviços que sejam coerentes para os usuários. Isso, segundo o autor, requer acordos de cooperação em três níveis: técnico, de conteúdo e organizacional. O nível técnico envolve

os componentes computacionais para que as informações possam ser trocadas, isto é, diz respeito às tecnologias como protocolos, estruturas e padrões. O nível de conteúdo diz respeito aos acordos semânticos sobre a interpretação das informações, ou seja, envolve os acordos compartilhados para interpretar as representações dos dados e metadados. Em relação ao nível organizacional, refere-se às regras para o acesso, a preservação e o planejamento de coleções e serviços, envolvendo principalmente processos de autenticação, direitos e licenças.

Tal perspectiva traz a possibilidade de discutir a interoperabilidade sob o ponto de vista também das comunidades envolvidas, uma vez que, para elas, os sistemas, ou melhor, as interfaces desses sistemas são melhor aproveitadas pelo público-alvo se apresentarem uma visão unificada e coerente de recursos informacionais heterogêneos, permitindo o acesso ao conteúdo de diversas fontes, bem como promovendo a navegação em ambientes integrados (SAYÃO; MARCONDES, 2008, HYVÖNEN, 2012, SANTAREM SEGUNDO; SILVA; MARTINS, 2019).

3.1 A interoperabilidade de dados heterogêneos

Para estabelecer a efetiva comunicação e a troca de informações entre sistemas, o principal desafio encontra-se em eliminar a heterogeneidade dos dados, que pode ser em nível estrutural, sintático e semântico. A heterogeneidade estrutural está relacionada às diferentes formas com que os dados são organizados em esquemas conceituais. Já a heterogeneidade sintática advém de sintaxes diferentes, ou seja, linguagens atribuídas a conceitos correspondentes. E a heterogeneidade semântica considera as diferenças de significado e interpretação atribuídos aos dados (SHETH; LARSON, 1990, CRUZ; XIAO, 2005).

A heterogeneidade estrutural e a sintática estão diretamente relacionadas a interoperabilidade técnica e organizacional, uma vez que diferentes fontes de dados podem armazenar seu conteúdo em diferentes formatos e locais ou “silos” de dados, que são bancos de dados distribuídos em rede, usando estruturas com padrões, normas e protocolos específicos, como é comum entre as instituições culturais (HYVÖNEN, 2012). Quanto à heterogeneidade semântica, ela incide principalmente sobre a interoperabilidade de conteúdo, que ocorre quando há uma discordância ou diferença em relação ao significado, interpretação ou uso pretendido dos mesmos dados ou dados relacionados (SHETH; LARSON, 1990), fato recorrente se forem considerados os diferentes domínios que lidam com o patrimônio cultural.

Isso faz com que o tratamento semântico seja um dos maiores desafios da integração, pois geralmente o significado preciso dos dados não está explícito ou não é bem comunicado devido às variações nas formas de interpretar, representar e estruturar os conhecimentos de domínio. Como destaca Hyvönen (2012), “[...] os problemas de interoperabilidade podem ser resolvidos efetivamente usando um único esquema. No entanto, esquemas diferentes são necessários e usados para diferentes tipos de dados em sistemas que lidam com conteúdo entre domínios”.

No caso do domínio do patrimônio cultural, essa diversidade de tipos de dados ocorre principalmente porque cada instituição desenvolve padrões de representação com base em suas necessidades específicas. Quando disponibilizados na web, esses dados em formatos específicos não conseguem interoperar com os demais sistemas, tornando seu uso comprometido ou inviável. Por isso, a adoção de estruturas capazes de trocar informações, superando as diferenças estruturais, sintáticas e semânticas ainda permanece uma das grandes discussões para áreas que buscam a integração.

Superar o desafio da integração se torna um meio de maximizar o valor e o potencial de reuso dos acervos e coleções, de modo que novos conhecimentos possam ser gerados, em especial, ao considerar os dados de patrimônios culturais, que por natureza são heterogêneos, multilíngues e derivados de uma variedade de fontes e eventos situacionais (HYVÖNEN, 2012). Logo, cresce a tendência entre as instituições culturais em reunir esforços e buscar meios para integrar acervos e coleções a partir dessas estruturas capazes de trocar informações, mesmo

utilizando dados heterogêneos. A literatura científica aponta para diversos momentos em que soluções foram desenvolvidas com esse fim, destacando as soluções tecnológicas da web Semântica que têm potencializado ainda mais esse objetivo.

3.2 Soluções tecnológicas para a integração e a interoperabilidade de dados

No ambiente web, os diferentes fatores relacionados à interoperabilidade têm sido resolvidos por meio dos padrões recomendados pelas agências internacionalmente reconhecidas, como a W3C (*World Wide Web Consortium*) e o DCMI (*Dublin Core Metadata Initiative*), que visam promover a representação e o compartilhamento de dados padronizados e estruturados a partir de soluções baseadas em linguagens, ontologias, modelos de dados, normas e protocolos desenvolvidos pensando no ambiente web.

A recomendação de formatos em linguagem XML (*eXtensible Markup Language*) tem sido muito utilizada no decorrer dos anos para os fatores relacionados à padronização sintática na codificação dos dados devido a sua funcionalidade flexível, aberta e independente de dispositivo (DOERR, 2003, SANTAREM SEGUNDO, 2004). Entretanto, outras linguagens mais modernas vêm sendo utilizadas e incorporadas nas diferentes aplicações da web, como Turtle (*Terse RDF Triple Language*) e JSON (*JavaScript Object Notation*). Em relação aos fatores semânticos, devido a sua maior complexidade, as soluções envolvem modelos conceituais e ontologias, construídos com base nos princípios semânticos de cada domínio (DOERR, 2003; HYVÖNEN, 2012).

Como os fatores de interoperabilidade são interdependentes, o *Resource Description Framework* (RDF) é uma das principais soluções tecnológicas recomendada pela web semântica, pois configura-se como um modelo de dados simples, que representa os recursos na web em termos de sintaxe e semântica de domínio de conhecimento, permitindo a interoperabilidade estrutural. Usando o *RDF-Schema* (RDFS), é possível fornecer os mecanismos necessários para a declaração e definição de propriedades e relacionamentos dos dados, definindo as características específicas de domínios e sua semântica subjacente (DIAS; SANTOS, 2003). O formato dessas declarações é constituído de três elementos: sujeito, predicado e objeto, denominado triplas (“recurso, propriedade e valor”, ou ainda “entidade, atributo, valor”), formando os chamados grafos (BIZER; HEATH; BERNERS-LEE, 2009). A função das triplas é expressar um relacionamento entre dois recursos digitais, sendo que um deles é o sujeito recurso/entidade e, o outro, o objeto/valor, entre eles está o predicado, também chamado de propriedade ou atributo, que representa a natureza do relacionamento entre os recursos (WORLD WIDE WEB CONSORTIUM, 2014).

Tanto os recursos quanto as propriedades são descritos usando um *Uniform Resource Identifier* (URI), que consiste em um identificador uniforme para identificar de modo unívoco um recurso digital na web, de tal modo que os recursos possam ser associados a partir de uma rede de relacionamentos (WORLD WIDE WEB CONSORTIUM, 2006). Além do RDF e dos URIs, que auxiliam nos vários fatores da interoperabilidade, as ontologias são importantes ferramentas para apoiar redes de dados, uma vez que permitem construir “[...] uma relação organizada entre termos dentro de um domínio, favorecendo a possibilidade de contextualizar os dados, tornando mais eficiente e facilitando o processo de interpretação [...]” (SANTAREM SEGUNDO, 2015, p. 226).

Em conjunto com os modelos conceituais próprios de cada domínio, as ontologias permitem comunicar as especificações de conteúdo de modo compartilhado a partir da modelagem dos aspectos semânticos desses domínios. Com isso, modelos e ontologias ampliam o uso dos padrões de metadados, permitindo o desenvolvimento dos Perfis de Aplicação, que reúnem um conjunto de metadados selecionados de diferentes esquemas a um esquema comum (ZENG; QIN, 2016). Nesse contexto, destacam-se os princípios do *Linked Open Data* (LOD), uma vez que a publicação de dados conectados de forma aberta viabiliza a reutilização e a

integração de várias fontes (HARPRING, 2016), otimizando a representação dos dados, minimizando as diferenças estruturais, ampliando as possibilidades de colaboração, as formas de uso e descoberta pelos usuários.

Outro importante aspecto tecnológico a ser considerado para a integração do conteúdo cultural que afeta os diferentes fatores de interoperabilidade é a descrição e visualização de imagens na web. Muitos objetos digitais imagéticos estão em “silos”, com acesso restrito a aplicativos personalizados e construídos localmente. Por isso, uma importante solução disponível é o *International Image Interoperability Framework* (IIIF), que define *Application Programming Interface* (APIs⁵) para padronizar a descrição e a visualização de imagens, de modo a fornecer metadados estruturados, possibilitando que qualquer aplicativo ou visualizador compatível com o padrão possa utilizar, compartilhar e exibir as imagens e seus metadados com qualidade (IIIF, [2021]).

Sendo assim, essas soluções tecnológicas recomendadas também afetam, em maior ou menor grau, os fatores voltados à interoperabilidade organizacional, uma vez que também influenciam nas decisões das políticas de acesso, uso, gestão e preservação dos acervos e coleções. Logo, a integração implica, além de encontrar meios de harmonizar formatos de metadados, também definir mecanismos que garantam que as medidas de qualidade do conteúdo sejam interoperáveis com as medidas de qualidade dos sistemas participantes (CANDELA *et al.*, 2007).

Os ambientes integrados, portanto, podem ser alcançados de diferentes formas, dependendo do grau de interoperabilidade e dos resultados a que se pretende alcançar, o que, por sua vez, dependem dos diferentes níveis de engajamento das fontes de dados, ou seja, dos participantes envolvidos. Assim, esses fatores fazem com que haja diversas alternativas para alcançar a integração. O estudo de Santarem Segundo, Silva e Martins (2018) permite verificar a integração de dados sob o ponto de vista das possibilidades técnicas de funcionamento dos modelos interoperáveis a partir dos protocolos, que são conjunto de regras que define a comunicação entre sistemas (SANTAREM SEGUNDO; SILVA; MARTINS, 2018). Para isso, os autores discutem os modos de operação dos protocolos em quatro categorias: agregação; sindicância (distribuição – servidor-cliente); protocolo de publicação; e busca distribuída. Todas as categorias apresentam oportunidades e desafios a serem considerados no que tange à disponibilização de acervos e coleções digitais.

Duas delas, a “agregação” e a “busca distribuída”, têm sido especialmente importantes para as discussões acerca da integração de repositórios digitais, permitindo recuperar e integrar metadados. Ambas as abordagens têm sido consideradas pela perspectiva da recuperação de informações como importantes para combinar dados de diferentes fontes e fornecê-los por meio de uma visão unificada na web (MARCONDES; SAYÃO, 2001, HYVÖNEN, 2012), porém usando estratégias diferentes que podem ser verificadas pelo uso das soluções empregadas para cumprir a tarefa. O quadro 2 compara esses dois tipos de abordagens que permitem a integração de dados.

Quadro 2. Descrição do processo de recuperação para integração de dados.

	AGREGAÇÃO	BUSCA DISTRIBUÍDA
DESCRIÇÃO	Ação ocorre previamente em um estágio separado de pré-processamento.	Ação ocorre durante o processamento de consultas.

⁵ Conjunto de funções e diretrizes usadas para interagir com um programa de computador (*software*), possibilitando que parte das funcionalidades de um serviço ou produto na web possam ser usadas em outras plataformas da maneira mais assertiva e conveniente para seus usuários (POMERANTZ, 2015; SANTAREM SEGUNDO; SILVA; MARTINS, 2018).

PROCESSO	Consolida-se primeiro as coleções locais em um <i>Data Warehouse</i> (armazém de dados) por meio da coleta (“ <i>harvesting</i> ”) de dados para, em seguida, possibilitar a consulta em um banco de dados global com interface única e integrada.	Uma consulta de pesquisa é enviada para bancos de dados locais distribuídos, nos quais um protocolo especifica estruturas de dados e regras de intercâmbio para que os registros que são identificados como pertinentes a essa busca sejam combinados em uma lista de ocorrências global.
CARACTERÍSTICA	Ocorrem processos separados de criação de conteúdo em cada bancos de dados locais, sendo esses conteúdos reunidos em um armazém de dados global, que os mantêm atualizados à medida que os bancos de dados locais participantes evoluem.	As organizações participantes podem continuar com sistemas de banco de dados locais independentes, mas devem concordar com o protocolo de consulta e em usar formatos de metadados compatíveis.
FUNÇÃO	Protocolos que facilitam o processo de exposição e coleta de metadados compartilháveis para que possam ser codificados e usados fora do seu ambiente local.	Protocolos que permitem servidores receberem solicitações de pesquisa federadas e responderem com resultados.
EXEMPLOS DE PROTOCOLOS	OAI-PMH e OAI-ORE.	Z39.50 e sucessores para web : SRU e SRW.

Fonte: Elaborado com base em Marcondes e Sayão (2001), Hyvönen (2012) e Santarem Segundo, Silva e Martins (2018).

A primeira abordagem envolve principalmente os protocolos *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) e *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-ORE), nos quais a comunicação se dá entre um provedor de serviços, que faz solicitações, e um provedor de dados, que responde enviando dados. O OAI-PMH permite que os provedores de dados exponham seus metadados estruturados de forma padrão, de modo que eles possam ser coletados pelos provedores de serviço que fazem solicitações no mesmo padrão (OPEN ARCHIVES INITIATIVE, [2021]a). Da mesma forma, o OAI-ORE define padrões, mas para a descrição e troca de agregações de objetos digitais, que combinam vários tipos recursos, como texto, imagens e vídeo (OPEN ARCHIVES INITIATIVE, [2021]b). Nesse caso, as consultas são realizadas em um repositório global e os usuários são redirecionados ao servidor específico quando solicitam o acesso ao conteúdo original (MARCONDES; SAYÃO, 2001).

A criação de armazéns de dados (*Data Warehouse*), ou seja, de uma base centralizada, prevê que os itens de um acervo ou uma coleção possam ser representados como recursos da web interligados, utilizando suas soluções tecnológicas para as associações semânticas (HYVÖNEN, 2012). Essa estratégia é adequada para a recuperação de grandes conjuntos de metadados dos repositórios digitais e, nesse caso, a colaboração e a coordenação são indispensáveis entre esses provedores de dados para manter os dados constantemente sincronizados e atualizados.

Em relação à “busca distribuída a diferentes servidores”, também chamada somente de “busca federada”, as consultas ocorreram de forma independente em cada banco de dados local, o que exige um protocolo padrão para que os resultados sejam consolidados em uma única interface, fornecendo respostas satisfatórias. Esse processo geralmente acontece devido a motivos legais, que impedem de integrar todas as coleções em um armazenamento central. Os protocolos que mais têm sido usados no contexto da web para esse propósito são o *Search and Retrieve URL* (SRU) e o *Search/Retrieve Web Service* (SRW), sendo ambos os protocolos resultado do esforço colaborativo internacional a fim de desenvolver uma interface padrão de pesquisa para a web, baseando-se na funcionalidade do Z39.50 (OCLC, 2021). Tal processo se fundamenta na estratégia cliente-servidor, na qual o cliente emite uma consulta ao servidor (provedor de dados), que é processada e, então, o cliente recebe os dados provenientes do

servidor.

As duas abordagens são, portanto, baseadas em modelos funcionais de comunicação que usam as soluções tecnológicas para permitir que as fontes de dados atendam a solicitações recebidas, fornecendo esses dados de forma funcional. Para isso, os serviços de gerenciamento de metadados são bastante usados, como o *software* REPOX, que possui uma interface gráfica para todas as funcionalidades de um processo de integração, incluindo vários canais, como OAI-PMH, HTTP, Z39.50, a fim de importar ou recuperar dados de provedores de dados, serviços para transformar dados entre esquemas de acordo com regras especificadas e serviços para compartilhar dados (EUROPEANA PRO, 2015).

Nesse sentido, a iniciativa da *Digital Public Library of America* (DPLA) é considerada um dos grandes exemplos de como diversas fontes de dados podem colaborar para formar uma rede funcional de dados interoperáveis.

4 A INICIATIVA DA DIGITAL PUBLIC LIBRARY OF AMERICA (DPLA)

A *Digital Public Library of America* (DPLA) teve o seu desenvolvimento impulsionado pelo ideal de um ambiente digital nacional aberto, gratuito e de qualidade ao acesso público. Contando com os esforços de diferentes profissionais, em outubro de 2010, iniciou-se o processo de planejamento da DPLA. A partir de uma reunião com representantes de bibliotecas, fundações, universidades, entre outros parceiros na cidade de Cambridge nos Estados Unidos, iniciou-se o projeto conjunto de criar uma rede aberta e distribuída de recursos on-line. No ano de 2011, começou um processo de intensa organização para definir, projetar e construir a DPLA, procedimento que levaria dois anos. Nesse tempo, com sede no Centro *Berkman Klein*, foram criados fluxos de trabalho com vários profissionais como bibliotecários, inovadores, humanistas digitais e outros profissionais voluntários, liderados por um comitê de direção. Assim, em abril de 2013, a DPLA foi lançada e tornou-se uma plataforma digital nacional livre, aberta e acessível ao público (DPLA PRO, [2021]a).

Durante o período de desenvolvimento pré-lançamento da DPLA, foi criado o *DPLA Metadata Application Profile* (DPLA MAP), uma parte fundamental da infraestrutura que suporta o funcionamento da plataforma. O DPLA MAP configura-se em um modelo de dados, desenvolvido com base no *Europeana Data Model* (EDM), usando apenas uma parte deste para a descrição de seus dados e demais soluções tecnológicas existentes, como o *Dublin Core* (DC), *Open Archives Initiative Object Resue & Exchange* (OAI-ORE) e *Resource Description Framework* (RDF) (MATIENZO; RUDERSDORF, 2014). Desde sua primeira versão pública, o DPLA MAP vem sendo atualizado, aprimorado e encontra-se em sua quinta versão, publicada em 2017. O modelo integra as experiências e necessidades específicas de descrição de dados de cada comunidade a partir da coleta, bem como agregação dos metadados dos conteúdos fornecidos pelas instituições parceiras. O DPLA MAP é a base para os dados estruturados na *Application Programming Interface* (API) da DPLA (DPLA MAP WORKING GROUP, 2017a), que permitem o amplo reúso dos dados pelos usuários da plataforma.

Assim sendo, o DPLA MAP é entendido como: “[...] um perfil de aplicativo ou um conjunto de elementos de metadados, obtidos de vários esquemas para um uso local específico. É também um modelo de metadados semântico, ou uma estrutura abstrata, que descreve as relações entre diferentes tipos de dados sobre um mesmo item” (DPLA MAP WORKING GROUP, 2017a, p. 1, tradução nossa).

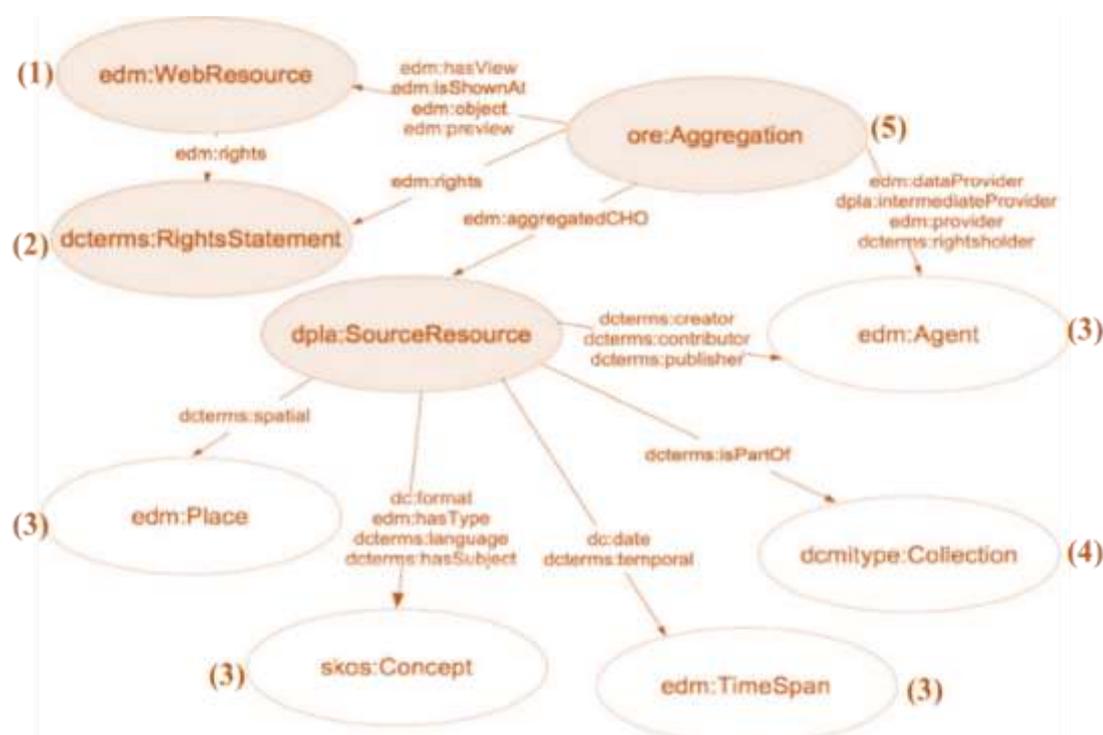
Como tal, o DPLA MAP é mais robusto e abstrato do que um esquema, ou padrão de metadados como o *Dublin Core*, ou demais padrões descritivos específicos de domínio. Ele

consiste em um Perfil de Aplicação⁶ que descreve os elementos a partir das entidades e dos relacionamentos usando o RDF, o que permite combinar esquemas para adaptarem-se às necessidades específicas (ZENG; QIN, 2016; DPLA MAP WORKING GROUP, 2017a). Por ser um modelo abstrato de metadados, o DPLA MAP pode ser expresso em qualquer codificação padrão, sendo que, no caso da DPLA API, é usada a linguagem *Java Script Object Notation* (JSON) para *Linked Data*, denominada JSON-LD⁷ (DPLA MAP WORKING GROUP, 2017a). Nessa perspectiva, o objetivo do DPLA MAP, tal como um Perfil de Aplicação baseado em RDF, é estabelecer o relacionamento entre as entidades que caracterizam um conteúdo, como a autoria e a data de criação, de tal modo a apresentar uma representação rica e bem estruturada. Para isso, o DPLA MAP faz uso de entidades e classes, com o intuito de representar os dados, e de *namespaces*⁸ a fim de denominar os campos de metadados.

O processo de representação ocorre com base na estrutura do RDF, tal como apresentado na figura 1, em formato de grafos de triplas, indicando o sujeito, o predicado e o objeto (recurso, propriedade e valor). O DPLA MAP apresenta em seu modelo abstrato um conjunto de propriedades específicas para cada recurso, chamado de classe (POMERANTZ, 2015). Cada classe contém uma lista de possíveis propriedades com base em conjuntos de elementos de metadados existentes, como *Dublin Core* e EDM, além dos elementos definidos pela própria DPLA, indicados respectivamente pelos *namespaces*: ‘*dc*’, ‘*edm*’, ‘*dpla*’ (DPLA MAP WORKING GROUP, 2017b).

Na classe ‘*dpla:SourceResource*’, estão as propriedades que contêm os metadados descritivos fornecido pelos provedores, como título, data, formato, que se referem ao “recurso de origem”, os objetos digitais.

Figura 1. Classes do DPLA MAP v.5.



Fonte: DPLA MAP Working Group (2017b).

⁶ Conjunto definido de propriedades de metadados que combina elementos selecionados de vários esquemas padronizados junto com os definidos localmente. Políticas e diretrizes também são definidas para um perfil específico (ZENG; QIN, 2016, DPLA MAP WORKING GROUP, 2017b).

⁷ Formato que possibilita construir mapeamentos adicionais ao JSON baseado no modelo RDF (SANTAREM SEGUNDO; SILVA; MARTINS, 2018).

⁸ Prefixo que precede o nome do elemento ou do atributo de metadados indicando sua origem (ARAKAKI, 2016).

A maioria das propriedades desta classe é baseada no *Dublin Core* (DC) e vincula-se a outras classes no DPLA MAP que, como é possível visualizar na figura 1: (1) armazenam informações sobre a versão digital do recurso original (*edm:WebResource*); (2) armazenam informações sobre direitos e reutilização de materiais (*dcterms:RightsStatement*); (3) permitem a descrição aprimorada de determinados campos (*Agent, Concept, Place e TimeSpan*); (4) reúnem informações sobre conjuntos ou coleções definidos localmente pelos provedores (*demyope:Collection*); e (5) empacotam todas essas informações juntas na classe '*ore:Aggregation*'. Na classe '*ore:Aggregation*', ocorre o armazenamento de propriedades importantes sobre os *Hubs* fornecendo informações acerca do local, da miniatura (visualização do conteúdo) e do registro de metadados original dos *Hubs*.

Isso permite não só a agregação de metadados, mas também a reutilização de elementos de diferentes esquemas já publicados, reutilizando vocabulários, facilitando a interoperabilidade e não duplicando informações. Desse modo, essa importante funcionalidade faz com que a DPLA consiga reunir os diferentes provedores de dados que, por sua vez, colaboram para tornar seu conteúdo pronto a fim de ser utilizado ao aderir aos padrões recomendados, permitindo que ocorra o processo de integração.

4.1 Os Hubs da DPLA

Um dos grandes destaques de iniciativas como a da DPLA é a sua função de integrar informações para potencializar a geração de conhecimento. Por essa razão, diferentes instituições concordam em empreender esforços e recursos para colaborar em benefício de uma interface de integração. Contudo, cada instituição apresenta especificidades e tecnologias próprias para gerenciar seus dados, o que faz com esse nível de interoperabilidade e agregação de dados proposto pela DPLA seja bastante diverso e complexo. Por isso, a DPLA trabalha com dois tipos de *Hubs*, que são os parceiros provedores. Eles são categorizados como os *Hubs* de Conteúdo e os *Hubs* de Serviço.

Os provedores denominados *Hubs* de Conteúdo ('*Content Hubs*') compreendem as grandes instituições culturais, como bibliotecas, arquivos e museus, que se caracterizam por compartilhar vastos conjuntos de registros de dados, mantendo-os e aprimorando-os conforme necessário. De acordo com a documentação da DPLA e dos *sites* de cada *Hub* (que podem ser acessados via *hyperlink* em cada nome). Os atuais *Hubs* de Conteúdo são:

1. **ARTstor:** fornece conteúdo para estudo, ensino e aprendizagem nas artes e campos associados, incluindo imagens e mídia de alta qualidade de acervos de museus, bibliotecas, acadêmicos e de artistas, incluindo materiais raros em coleções de acesso aberto disponíveis gratuitamente.
2. **Biodiversity Heritage Library (BHL):** atua como um consórcio mundial de bibliotecas de história natural, botânica e pesquisa em biodiversidade.
3. **David Rumsey:** materiais cartográficos, que incluem coleções de atlas, mapas de parede, globos, geografias escolares, mapas de bolso de parede, infantis, manuscritos, livros de exploração, cartas marítimas e uma variedade de outros itens sobre cartografia.
4. **J. Paul Getty Trust:** dedicada à apresentação, conservação e interpretação do legado artístico mundial relacionado a artes visuais, fornecendo acesso por meio de seus programas: *Getty Conservation Institute, Getty Foundation, J. Paul Getty Museum e Getty Research Institute*.
5. **Harvard Library:** disponibiliza em acesso público gratuito parte do conteúdo digital da biblioteca de Harvard, incluindo suas coleções raras e especiais como arte antiga, manuscritos e audiovisual.
6. **Hathi Trust:** repositório colaborativo que fornece acesso ao conteúdo digital do acervo de diversas bibliotecas acadêmicas e de pesquisa a partir de serviços e programas que fornecem: o acesso temporário de emergência durante interrupções de serviço; o uso do conteúdo para análise por meio de pesquisa de texto e mineração de dados; impressão com retenção coletiva e dá acesso às publicações federais.
7. **Internet Archive:** reúne o conteúdo publicado na *Internet* e outros artefatos em formato digital,

documentando e preservando a história da web. Sua coleção inclui principalmente os materiais efêmeros como sites, gravações de áudio, shows ao vivo, vídeos, programas de notícias de televisão, imagens e softwares.

8. **Library of Congress:** fornece em sua plataforma digital acesso a livros, gravações, fotografias, jornais, mapas e manuscritos de seu acervo, bem como materiais jurídicos confiáveis que apoiam as ações do Congresso dos EUA.
9. **National Archives and Records Administration (NARA):** agência federal independente que preserva e compartilha registros da história, do governo e dos cidadãos dos EUA, disponibilizando acesso digital a documentos e materiais criados pela administração pública.
10. **New York Public Library:** provedor gratuito que reúne bibliotecas e filiais na região da cidade de Nova York, oferecendo acesso digital gratuito a uma parte do seu acervo que conta com materiais como livros, vídeos, mapas, manuscritos, ilustrações, fotos para usuários, desde crianças a profissionais acadêmicos.
11. **Smithsonian Institution:** complexo composto por museus e galerias, um parque zoológico nacional, instalações de pesquisa e bibliotecas que dá acesso digital a suas coleções e seus acervos, incluindo livros, periódicos, objetos de museu, manuscritos, imagens e vídeos.
12. **United States Government Publishing Office (GPO):** agência federal encarregada de produzir e distribuir produtos e serviços de informação para todos os três ramos do Governo Federal, fornecendo acesso público permanente sem custos às informações governamentais, como projetos de lei, leis, regulamentos, documentos presidenciais, estudos e outros documentos federais por meio do *Federal Depository Library Program (FDLP)* e *GovInfo*.
13. **University of Southern California Digital Library (USCDL):** disponibiliza o conteúdo digital fornecidos pelas 23 bibliotecas e centros de informação que fazem parte da USC, apoiando o acesso, a descoberta, a criação e a preservação do conhecimento produzido e relacionado à universidade.
14. **University of Washington Libraries:** disponibiliza o conteúdo digital como fotografias, mapas, jornais, pôsteres, relatórios e outras mídias referentes aos três campi (*Seattle, Tacoma e Bothell*) e do *Friday Harbor Laboratories* da universidade.

13

Os *Hubs* de Conteúdo são assim chamados, pois caracterizam-se principalmente por serem grandes agregadores de dados de diferentes tipos de repositórios digitais, mantendo um relacionamento direto com a DPLA, muitas vezes, fornecendo os registros de dados em diferentes Perfis de Aplicação, que são então transformados para o DPLA MAP (DPLA MAP WORKING GROUP, 2017b).

Já os denominados *Hubs* de Serviço (*'Service Hubs'*) são formados por instituições contribuintes estaduais ou regionais, que se organizaram em rede e colaboram para enviar os registros dos dados dos seus repositórios à DPLA. Da mesma forma que os *Hubs* de Conteúdo, os *Hubs* de Serviço são agregadores, mas reúnem o conteúdo por região geográfica ou escopo. Eles compartilham recursos, funções e responsabilidades, não sendo necessário, portanto, que cada instituição contribuinte preste todos os serviços. Assim, instituições contribuintes de menor porte, que não têm a capacidade de comprar ou hospedar seu próprio sistema de repositório e serviços de digitalização, podem trabalhar com seu conteúdo digital colaborativamente, sendo capazes de atender aos vários requisitos de agregação (DPLA PRO, [2021]b). Essas instituições contribuintes compartilham seus dados com os *Hubs* de Serviço, que são subsequentemente compartilhados com DPLA.

Desse modo, os *Hubs* de Serviço cumprem um importante papel na proposta da DPLA de reunir a riqueza cultural de todo o país e disponibilizá-la digitalmente, uma vez que conectam parceiros de diferentes tamanhos e origens, dando visibilidade nacional e mundial a esse conteúdo cultural.

Eles reúnem metadados que caracterizam os objetos digitais (textos on-line, fotografias, material manuscrito, arte etc.) de bibliotecas, sociedades históricas, arquivos, museus e outras instituições de patrimônio cultural que participam de sua rede, muitas vezes, hospedando esses recursos localmente, além de compartilhar metadados e visualizações de conteúdo (miniaturas, etc.) por meio do DPLA. (DPLA PRO, [2021]b, não paginado, tradução nossa)

Com isso, os *Hubs* de Serviço também desempenham muitas vezes o papel de repositórios para essas instituições, assumindo a responsabilidade na digitalização, preservação e armazenamento a longo prazo dos objetos digitais, bem como fornecendo assistência no que diz respeito aos metadados, visando a garantia de qualidade, normalização, padronização e aprimoramento. Ainda, envolvem-se com o engajamento das comunidades, auxiliando no desenvolvimento de tecnologia, ferramentas e no aperfeiçoamento profissional (DPLA PRO, [2021b]). Com isso, cada *Hub* de Serviços se compromete com as seguintes funções:

- Representar sua comunidade (estado, região etc.) como o ponto de contato para a DPLA e obter a adesão da comunidade em questões significativas que afetam seus parceiros.
- Agregar os metadados de seus parceiros em um único padrão e compartilhá-los com o DPLA por meio de uma fonte de dados coletável.
- Lidar ativamente com as questões dos metadados (incluindo direitos autorais e rotulagem de licença) e trabalhar com parceiros na correção oportuna.
- Oferecer divulgação aos seus parceiros e, com a equipe da DPLA, desenvolver a capacidade dos profissionais locais em tópicos, como dados abertos, qualidade e padrões de dados, direitos autorais e licenças e demais assuntos relevantes.
- Realizar manutenção das tecnologias (como OAI-PMH, API, *ResourceSync* etc.) que permitem que metadados padronizados sejam compartilhados com DPLA de forma regular e consistente.
- Envolver-se com a comunidade mais ampla de criadores, provedores e usuários de dados, local e nacionalmente. (DPLA PRO, [2021b], não paginado, tradução nossa)

A documentação da DPLA indica os atuais *Hubs* de Serviço que fornecem dados a plataforma:

1. **Big Sky Country Digital Network:** reúne e dá acesso ao conteúdo digital das instituições culturais no domínio geográfico de Montana, na Dakota do Norte.
2. **California Digital Library (DCL):** reúne e dá acesso ao conteúdo digital das instituições parcerias da Universidade da Califórnia, tanto do campus como por meio de colaborações externas.
3. **Connecticut Digital Archive (CTDA):** faz parte do Programa de Repositório de Preservação Digital da Universidade de Connecticut, reunindo e dando acesso a uma ampla gama de recursos digitais para instituições educacionais e culturais, bem como a agências estaduais em Connecticut.
4. **Digital Commonwealth:** fornece recursos e serviços para apoiar a criação, gerenciamento e disseminação de materiais de herança cultural mantidos por instituições culturais de Massachusetts.
5. **Digital Library of Georgia:** colabora com instituições de educação e cultura da Geórgia para fornecer acesso a recursos digitais e serviços micrográficos sobre a história, cultura e vida do estado, apoia o ensino, a pesquisa e as missões de serviço do *Georgia Library Learning Online* (GALILEO) e do Sistema Universitário da Geórgia.
6. **Digital Library of Tennessee (TEL):** reúne e dá acesso a recursos digitais, incluindo revistas, periódicos acadêmicos, podcasts, vídeos, e-books, materiais de preparação para testes, registros do censo federal e demais materiais de fonte primária do Tennessee.
7. **Digital Maine:** parceria da *Maine State Library*, *Maine State Archive* e demais instituições comunitárias para fornecer acesso à história em todo o estado do Maine, garantindo transparência no governo e compartilhando as histórias das pessoas e lugares. Conta com coleções de mapas, registros da igreja, histórias locais, pesquisa de genealogia, relatórios de imagens da comunidade e outros relatórios, publicações e pesquisas das agências estaduais.
8. **Digital Maryland:** programa colaborativo de preservação digital para todo o estado de Maryland, envolvendo o *University System of Maryland & Affiliated Institutions* (USMAI), um consórcio de 17 faculdades e universidades públicas, dando acesso a documentos históricos e culturais, imagens, áudio e vídeos que registram a história do estado.
9. **Digital Virginias:** colaboração entre as organizações que originalmente formava um Hub de Conteúdo da DPLA, envolvendo apenas a *University of Virginia* e, atualmente, incorpora os parceiros regionais *George Mason University*, *William & Mary*, *Virginia Commonwealth*

University, Virginia Tech e West Virginia University a fim de criar um conjunto combinado de materiais históricos e culturais para toda a região da Virgínia e da Virgínia Ocidental.

10. **District Digital:** colaboração entre a *DC Public Library* e o *Washington Research Library Consortium* para ajudar a reunir as coleções digitais de instituições culturais dentro e ao redor do Distrito de Columbia.
11. **Empire State Digital Network (ESDN):** rede administrada pelo *Metropolitan New York Library Council (METRO)* em colaboração com oito conselhos regionais de bibliotecas aliadas que colaboram para agregar conteúdo de projetos existentes de coleções digitais estadual e regional das organizações culturais em todo o estado de Nova York.
12. **Green Mountain Digital Archive (GMDA):** colaboração entre *Middlebury College, Vermont State Archives & Records Administration, Vermont Historical Society, Rockingham Free Public Library, Norwich University, St. Michael's College, University of Vermont* e *Vermont Department of Libraries*, que reúne fotografias, documentos, mapas, gravações e outros recursos digitais relacionados ao estado de Vermont.
13. **Illinois Digital Heritage Hub (IDHH):** formada por quatro instituições, a *Chicago Public Library*, a *Consortium of Academic and Research Libraries* in Illinois, a *Illinois State Library*, e a *University of Illinois at Urbana-Champaign Library*, que reúnem os recursos do patrimônio cultural digital de todo o estado de Illinois, incluindo o conteúdo do *Illinois Digital Archives* e do *CARLI Digital Collections*.
14. **Indiana Memory:** colaboração entre as instituições culturais do estado de Indiana para fornecer acesso à riqueza de suas fontes primárias que são disponibilizadas digitalmente.
15. **Kentucky Digital Library (KDL):** iniciativa colaborativa da *Kentucky Virtual Library (KYVL)* para dar acesso a coleções de arquivos digitais relacionadas à história e cultura compartilhados na comunidade de Kentucky.
16. **Michigan Service Hub:** colaboração entre a *Library of Michigan (LOM)*, a *Midwest Collaborative for Library Services (MCLS)*, a *University of Michigan (UM)*, *Wayne State University (WSU)*, *Michigan State (MSU)* e *Western Michigan University (WMU)* para agregar as coleções digitais de diversas instituições na região de Michigan.
17. **Minnesota Digital Library:** iniciativa que reúne e oferece acesso a coleções digitais exclusivas, compartilhadas por organizações de patrimônio cultural de todo o estado de Minnesota, incluindo cartões postais, mapas, cartas e registros de história oral.
18. **Mississippi Digital Library:** iniciativa colaborativa do estado do Mississippi que oferece um espaço on-line para pesquisar e explorar os conteúdos cultural e histórico mantidos por instituições e repositórios em todo o estado de Mississippi.
19. **Missouri Hub (MOHub):** afiliação de instituições que buscam dar visibilidade e relevância às coleções digitais oferecidas on-line no Missouri, agregando informações sobre objetos digitais.
20. **Mountain West Digital Library (MWDL):** iniciativa de colaboração entre instituições culturais e de ensino criada pela *Utah Academic Library Consortium*.
21. **NJ/DE Digital Collective:** coletivo que agrega dados de bibliotecas, museus, organizações de patrimônio cultural e outras instituições vinculadas ao estado de New Jersey e Delaware.
22. **North Carolina Digital Heritage Center (Digital NC):** programa estadual de digitalização e publicação digital que trabalha com instituições de patrimônio cultural de todos os tamanhos em todo o estado da Carolina do Norte.
23. **Ohio Digital Network (ODN):** projeto desenvolvido no estado de Ohio para coordenar as coleções digitalizadas e disponibilizá-las on-line, a partir do programa *Digitization Hubs (DigiHubs)* que envolve os parceiros *Columbus Metropolitan Library Public Library of Cincinnati and Hamilton County, Toledo Lucas County Public Library* e o *Cleveland Public Library*, auxiliando desde o uso de equipamentos até a criação de metadados.
24. **Oklahoma Hub (OK Hub):** parceria entre as instituições *Oklahoma Department of Libraries, Oklahoma Historical Society, Oklahoma State University Library* e a *University of Oklahoma Libraries*, oferecendo recursos exclusivos, particularmente nas áreas de história e cultura dos nativos americanos, ciências ambientais e agrícolas, bem como sobre as vidas e experiências de gerações do estado de Oklahoma.
25. **Orbis Cascade Alliance:** serviço de coleções digitais que realiza limpeza, treinamento e suporte de metadados para suas coleções, incluindo o desenvolvimento e a implementação da documentação, além de ser agregador de dados.

26. **PA Digital:** parceria envolvendo bibliotecas, sociedades históricas, museus, universidades e demais instituições pelo estado da Pensilvânia que reúnem patrimônios culturais de coleções e recursos históricos.
27. **Plains to Peaks Collective:** parceria entre as *Colorado & Wyoming State Libraries*, apoiado pelo *Institute of Museum and Library Services (IMLS)* para dar acesso a coleções digitais da história cultural das regiões.
28. **Portal to Texas History:** portal com materiais raros, históricos e de fonte primária sobre o estado do Texas, criado e mantido pelas bibliotecas da *University of North Texas*. Oferece coleções etnicamente diversas, como a “Sociedade de Preservação do Patrimônio Dinamarquês” e os “Programas funerários afro-americanos” da Biblioteca Pública de *San Antonio*, para serem usados por acadêmicos e pelo público em geral. Além disso, oferece oportunidade para que pequenas comunidades rurais preservem e acessem sua própria história.
29. **Recollection Wisconsin:** projeto que reúne recursos de patrimônio cultural digital, incluindo fotografias, mapas, cartas, diários, registros de histórias orais, artefatos e outros recursos históricos de mais de 200 bibliotecas, museus e outras instituições de patrimônio cultural do estado de Wisconsin.
30. **South Carolina Digital Library (SCDL):** projeto que coordena a distribuição de recursos necessários para incentivar esforços de digitalização e fornece acesso gratuito e com licença legal a coleções on-line de mais de 40 instituições no estado da Carolina do Sul.
31. **Sunshine State Digital Network (SSDN):** rede colaborativa de coleções digitais envolvendo organizações de patrimônio cultural em todo o estado da Flórida, apoiada pela *Library Services and Technology Act administered* da *Florida Department of State, Division of Library and Information Services*.

Além dos *Hubs* supracitados, a DPLA informa que quatro projetos estão em progresso. São eles o *Northwest Heritage Hub*, o *Orbis Cascade Alliance* (Washington/Oregon), o *NJ/DE Digital Collective* e o *New Hampshire Digital Library*.

Com dados bastante heterogêneos, cada um dos *Hubs* apresenta especificidades, tecnologias e níveis de governança próprios que melhor se adequam à realidade de suas comunidades, cabendo à DPLA auxiliá-los para que possam adotar fluxos de trabalho contínuos de agregação, normalização e provimento de metadados, de modo que permaneçam engajados no esforço colaborativo de integração de dados.

4.2 O trabalho de agregação desempenhado pelos Hubs da DPLA

Como objetivo deste estudo, foi proposto apresentar um panorama das soluções tecnológicas subjacentes à coleta e disponibilização de dados dos parceiros da DPLA. Sendo assim, para delimitar a análise, optou-se por dividir os *Hubs* em dois grupos.

O primeiro grupo é o dos sócios da DPLA, que são os membros que pagam uma taxa anual pela parceria, obtendo em retorno benefícios, principalmente, em relação às ferramentas e aos serviços da DPLA, como o desenvolvimento de APIs para a manutenção e o acesso a dados aprimorados, que permite a reutilização por desenvolvedores, pesquisadores e outros interessados dentro e fora da rede. Além disso, os sócios têm acesso a registros de metadados mapeados e enriquecidos pela DPLA, incluindo geolocalização aprimorada e limpeza de dados para atender aos padrões.

A DPLA ainda proporciona aos sócios: um cronograma de ingestão regular de metadados com base na frequência de alterações dados do *Hub*; acesso a dados analíticos; participação em grupos de trabalho e forças-tarefa e em projetos de curadoria do conteúdo, como exposições e listas de navegação; participação em rede para resolver problemas comuns de forma colaborativa; participação no site da *Rede Hub Wiki* para apoiar o trabalho individual e coletivo, facilitando a comunicação; fornecimento de treinamento, documentação, teste de garantia de qualidade e consultoria para apoiar a ingestão inicial; disponibilização de recursos para criar um site local DPLA. Isso faz com que a maioria desses *Hubs* apresentem soluções

similares para a coleta, agregação e provimento de seus dados, pois seguem as recomendações da DPLA.

O segundo grupo é o dos não sócios, provedores de dados que não fazem parte do programa de afiliação. Esses *Hubs*, em geral, utilizam ferramentas e serviços definidos localmente para o gerenciamento e agregação, o que confere a seus dados formatos variados. Apesar disso, esses dados ainda podem ser ingeridos e integrados pela DPLA, pois, independente do grupo de provedores, todos os metadados, tanto os criados localmente quanto os providos por parceiros, são entregues à DPLA em um único fluxo, no qual, independente do esquema original, os dados serão mapeados para um único esquema (ex. *Dublin Core*) e estruturados de acordo com o padrão OAI-PMH a fim de serem compatíveis para a coleta. Então, depois de coletados pela DPLA, os metadados podem ser transformados na estrutura do DPLA *Metadata Application Profile* (MAP), sendo armazenados e publicados via JSON-LD (DPLA, [2021]).

Embora o OAI-PMH ainda seja o tipo de alimentador mais utilizado pelos *Hubs*, permitindo-os fornecer dados no padrão *Dublin Core* simples ou qualificado ou, ainda, no *Metadata Object Description Schema* (MODS), cada vez mais, as instituições buscam a especificação *ResourceSync*⁹, uma aprimoração do OAI-PMH que permite a sincronização não apenas de metadados, mas também dos próprios objetos digitais (SOMPEL, 2014).

Para o processo de provimento dos dados, alguns fatores são destacados pela DPLA: a) a importância de que os *Hubs* e as instituições contribuintes compartilhem e mantenham os *links* de localização - URL (*Uniform Resource Locator*) para os itens e registros originais, bem como miniaturas (imagens) que representam o conteúdo em suas coleções locais; b) atentar-se para que os campos obrigatórios, como o “*direitos*” e o “*nome da instituição*” que contribuiu com o registro, sejam sempre devidamente preenchidos; c) manter a consistência dos metadados em todos os conjuntos de dados de um único *Hub* para que sejam sempre estruturados e interpretados da mesma maneira em todas as suas coleções; e d) todos os registros compartilhados com a DPLA devem estar disponíveis sem restrições, sob declarações de direitos padronizados e recomendações para o uso de declarações ou sob licença *Creative Commons* (DPLA, 2017).

Mesmo seguindo essas diretrizes para o provimento dos dados, o segundo grupo exibe alguns fatores particulares devido a seus sistemas independentes. Esses fatores foram sistematizados e apresentados no quadro 3, no qual cada *Hub* está descrito em função de suas principais funcionalidades de agregação de acordo com as informações disponíveis em seus *sites*.

Quadro 3. Principais funcionalidades dos *Hubs* não sócios.

PROVEDOR DE DADOS		FUNCIONALIDADE
Hubs de Conteúdo	ARTstor	Usa principalmente a ferramenta JSTOR, para catalogar e compartilhar as coleções públicas, e o visualizador de imagens IIF, para a disponibilização das imagens. Fornece o programa <i>Data for Research</i> (DfR) que permite a análise de texto e dados de pesquisa em humanidades digitais.
	Biodiversity Heritage Library (BHL)	Utiliza ferramentas de inteligência taxonômica da <i>Global Names Architecture</i> para os nomes científicos, produz URLs estáveis e atribui DOIs (identificadores de objetos digitais), oferece exportações de dados via APIs e nos formatos MODS, BibTex e RIS, disponibilizando seu conteúdo digital em acesso aberto sob licença <i>Creative Commons</i> .

⁹ Especificação que descreve uma estrutura de sincronização para a web, permitindo que sistemas de terceiros permaneçam sincronizados com os recursos em evolução de um servidor (OPEN ARCHIVES INITIATIVE, 2017).

	David Rumsey	Utiliza o <i>software</i> LUNA da <i>Luna Imaging</i> para disponibilização do conteúdo digitalizado em acesso aberto sob licença <i>Creative Commons</i> , que fornece uma API de apresentação para a manipulação e suporte em IIF dos resultados de pesquisa de coleção cruzadas, grupos de mídia e formatos para exibição.
	J. Paul Getty Trust	Faz uso do software de código aberto <i>Arches</i> , desenvolvido para gerenciamento, descoberta, visualização de dados e também para projetos e fluxo de atividades. Utiliza as ferramentas <i>Quire</i> , para publicação de texto em multiformato, e o IIF, para acesso a imagens.
	Hathi Trust	Faz uso da própria API denominada <i>HathiTrust Data API</i> para recuperar objetos digitais, incluindo o conteúdo digitalizado ou nato digitais, compostos de imagens e texto OCR (<i>Optical Character Recognition</i>) e os metadados correspondentes. Utiliza a licença de código-fonte do <i>software</i> Kakadu para a criação, manutenção e entrega de imagens.
	Internet Archive	Utiliza a funcionalidade do serviço <i>Archive-it</i> , programa próprio que permite o arquivamento do conteúdo da <i>web</i> , fornecendo ferramentas, treinamento e suporte técnico para capturar e preservar materiais dinâmicos na <i>Internet</i> . Utiliza APIs para tornar os metadados acessíveis por meio do portal <i>Internet Archive Wayback Machine</i> .
	Library of Congress	Utiliza o MODS como esquema de descrição de objetos de metadados e oferece o cruzamento/mapeamento entre MODS e DC para viabilizar a agregação de dados.
	University of Southern California Digital Library (USCDL)	Utiliza os serviços do <i>CONTENTdm</i> , software de coleções digitais da OCLC, que oferece suporte ao DC qualificado e à API de imagem e de apresentação IIF. Todos os itens das coleções incluem metadados sobre direitos autorais, licenças e uso dos materiais.
	University of Washington Libraries	Utiliza os serviços da OCLC, incluindo o software <i>CONTENTdm</i> , e o padrão de <i>Dublin Core</i> para os metadados.
Hubs de Serviço	Digital Maine	Utiliza os serviços do <i>Digital Commons</i> , <i>software</i> de repositório para publicar, gerenciar e disponibilizar abertamente todo o espectro de acervos digitais que oferece geolocalização, cronogramas, metadados personalizáveis e galerias de imagens que permitem visualização panorâmicas e zoom.
	Digital Maryland	Fornece pouca documentação e um conjunto de diretrizes próprias em um documento denominado <i>Metadata Style Guide</i> , disponível no site, que definem o formato dos metadados que são aceitos pela plataforma.
	Kentucky Digital Library (KDL)	Utiliza as ferramentas integradas do <i>CONTENTdm</i> da OCLC, que permitem suporte para visualização em dispositivos móveis, oferecem suporte ao padrão IIF e a demais interfaces personalizadas integrando coleções via API.
	South Carolina Digital Library (SCDL)	Usa o <i>SCDL Metadata Schema</i> criado com base na <i>Dublin Core Metadata Initiative</i> (DCMI) e nas melhores práticas da OCLC para <i>CONTENTdm</i> , além de outros repositórios compatíveis com OAI-PMH. Baseada também em algumas das diretrizes da DPLA.

Fonte: Elaborado pelos(as) autores(as).

A análise do quadro 3 permite delinear um panorama das soluções tecnológicas utilizadas na coleta e disponibilização de dados dos parceiros da DPLA, dando a possibilidade de pontuar alguns aspectos das oportunidades e dos desafios que os provedores de dados enfrentam na busca pelo acesso integrado.

Em relação às oportunidades, há os benefícios que as instituições passam a ter com as parcerias, não só no plano mais alto, que é a relação com a DPLA, mas no plano das próprias instituições que, ao colaborarem, onde desenvolvem um importante trabalho de dar visibilidade ao conteúdo cultural de diversas fontes, ampliando o acesso e a preservação da história, bem como da memória da sua comunidade. Já os desafios mostram-se em diferentes níveis. Como

visto, existem as questões relacionadas aos problemas universalmente conhecidos para a interoperabilidade de dados envolvendo a parte técnica, que podem ser solucionados pela escolha das tecnologias disponíveis e adequadas para cada caso. No tocante ao conteúdo, há principalmente conflitos na forma de interpretar cada diferente tipo de objeto digital, sejam eles digitalizados, ou nato digitais.

Por último, influenciado todas essas questões, encontram-se os desafios relacionados à governança. Cada instituição faz suas próprias escolhas de gerenciamento e interoperabilidade, o que muitas vezes gera conflitos quando buscam integrar seus esforços. Portanto, reafirma-se que uma das principais recomendações para superar esses desafios é a de que, desde o início de um projeto de acervos e coleções digitais, as instituições preocupem-se com os aspectos ligados à interoperabilidade, evitando o isolamento em sistemas únicos. Além disso, fica evidente que as boas práticas na criação, no uso ou reúso de padrões e esquemas de metadados impactam diretamente a capacidade de integração.

Logo, destaca-se a importância da implementação de abordagens de curadoria digital que orientem os diversos níveis de gerenciamento, desde o planejamento até o uso e o reúso, visando garantir não só a capacidade de intercâmbio dos dados, mas, sobretudo, que esses dados sejam válidos e úteis para diferentes contextos de usos.

5 CONSIDERAÇÕES FINAIS

A proposta deste estudo, apesar de não exaustiva, possibilitou a construção de inferências acerca das possibilidades e dos desafios de integração de dados, por meio das informações fornecidas, tanto pela DPLA quanto por seus *Hubs*. Essas informações foram validadas junto a fundamentação teórica, possibilitando alcançar os resultados e as discussões. Diante disso, a constatação mais explícita é a de que existem grandes esforços para a integração de dados. O que é comprovado pelo número de soluções disponíveis e a constante preocupação em aprimorá-las para garantir não só a interoperabilidade funcional, mas também a consistência dos fluxos de trabalho junto à evolução da web, a qualidade dos dados, incluindo metadados e objetos digitais como um todo. Isso permite refletir que existem diferentes dimensões possíveis para se discutir o trabalho de integração de dados. O estudo da DPLA forneceu um panorama dessas possibilidades, com foco nos processos inerentes à agregação de dados de repositórios digitais de acervos e coleções culturais.

Verificou-se que esses processos envolvem, além das soluções tecnológicas necessárias para as funcionalidades computacionais, também o engajamento de provedores de dados. Todos os *Hubs* da DPLA apresentam programas de incentivo e auxílio à digitalização e à documentação do conteúdo, bem como diretrizes para garantir a qualidade e os direitos legais de acesso. Oportunamente, o estudo permitiu destacar o importante papel da DPLA em relação à melhoria da experiência dos usuários com os acervos e as coleções culturais, dando a possibilidade de multidimensionar as formas de transformar informação em conhecimento.

Assim, destaca-se a importância de iniciativas como a da DPLA e da Europeia, entre outras, por promoverem e disponibilizarem suas diretrizes e tecnologias em acesso aberto gratuito, de modo que novas iniciativas possam se fundamentar para criar seus próprios sistemas, bem como criar sistemas interoperáveis a partir desses padrões. O Brasil é um exemplo de país que tem buscado desenvolver-se nessa direção, por meio do “Programa Acervo em Rede”, que promove a digitalização e a documentação dos acervos museológicos do Instituto Brasileiro de Museus (Ibram). Com o denominado “Projeto Tainacan”, os primeiros resultados da iniciativa já podem ser verificados a partir dos repositórios dos museus da rede Ibram¹⁰, que têm publicado suas coleções com o *software* livre *Tainacan*, solução baseada em *WordPress*. Os próximos passos da iniciativa seguem na direção de disponibilizar um portal de

¹⁰ Disponível em: <https://tainacan.org/blog/2020/12/31/tainacan-acoes-2020/>.

busca integrada, reunindo acervos museológicos e permitindo o intercâmbio dos dados na dimensão da cultura brasileira.

REFERÊNCIAS

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. Disponível em: <http://hdl.handle.net/11449/147979>. Acesso em: 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. Disponível em: <http://www.dlib.org/dlib/january02/arms/01arms.html>. Acesso em: 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. Disponível em: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 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. Disponível em: http://www.delos.info/index.php?option=com_content&task=view&id=345. Acesso em: 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. Disponível em: <https://www.cs.uic.edu/~advis/publications/dataint/eis05j.pdf>. Acesso em: 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. Disponível em: <https://www.e-publicacoes.uerj.br/index.php/cadinf/article/view/6619/4734>. Acesso em: 02 jun. 2021.

DOERR, M. **DCC Digital Curation Manual**: instalment on ontologies. Edimburgo, Escócia: Digital Curation Manual. 2008. Disponível em: <http://hdl.handle.net/1842/3341>. Acesso em: 02 jun. 2021.

DPLA. **DPLA Standardized Rights Statements Implementation Guidelines**, 2017. Disponível em: <https://bit.ly/3KijSsP>. Acesso em: 02 jun. 2021.

DPLA. **Partnering with DPLA**, [2021]. Disponível em: https://docs.google.com/document/d/1gshwQ0Oj84I5q-_JxHo7wjyLeXln9lCg6SUtIMkPu_g/edit#. Acesso em: 02 jun. 2021.

DPLA PRO. **History**, [2021a]. Disponível em: <https://pro.dp.la/about-dpla-pro/history>. Acesso em: 02 jun. 2021.

DPLA PRO. **Becoming a Service Hub**: service Hub models, roles, and responsibilities, [2021b]. Disponível em: <https://pro.dp.la/prospective-Hubs/becoming-a-service-Hub>. Acesso em: 02 jun. 2021.

EUROPEANA PRO. **Apresentando REPOX**: uma ferramenta para gerenciar espaços de metadados, 2015. Disponível em: <https://pro.europeana.eu/post/introducing-repo-x-a-tool-to-manage-metadata-spaces>. Acesso em: 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]. Disponível em: <https://iiif.io/about/>. Acesso em: 02 jun. 2021.

ISOTANI, S.; BITTENCOURT, I. I. **Dados Abertos Conectados**: em busca da web do conhecimento. Novatec Editora, 2015. Disponível em: http://pgcl.uenf.br/arquivos/dadosabertosconectados_011120181613.pdf. Acesso em: 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. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2735/1748>. Acesso em: 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. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/102416>. Acesso em: 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. Disponível em: <http://revista.ibict.br/ciinf/article/view/909>. Acesso em: 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. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3700>. Acesso em: 02 jun. 2021.

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

OCLC. **SRW/U**, 2021. Disponível em: <https://www.oclc.org/research/areas/data-science/srw.html>. Acesso em: 02 jun. 2021.

OPEN ARCHIVES INITIATIVE. **ResourceSync Framework Specification (ANSI/NISO Z39.99-2017)**, 2017. Disponível em: <http://www.openarchives.org/rs/1.1/resourcesync>

OPEN ARCHIVES INITIATIVE. **Open Archives Initiative Protocol for Metadata Harvesting**, [2021a]. Disponível em: <https://www.openarchives.org/pmh/>. Acesso em: 02 jun. 2021.

OPEN ARCHIVES INITIATIVE. **Open Archives Initiative Protocol for Metadata Harvesting**, [2021b]. Disponível em: <https://www.openarchives.org/ore/>. Acesso em: 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) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2004. Disponível em: <http://repositorio.unesp.br/handle/11449/93618>. Acesso em: 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. Disponível em: <http://revistas.ancib.org/index.php/tpbci/article/view/359/359>. Acesso em: 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. Disponível em: <https://doi.org/10.22478/UFPB.1809-4783.2019V29N2.38107>. Acesso em: 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. Disponível em: https://www.researchgate.net/publication/319403030_Digitalizacao_de_acervos_culturais_reuso_curadoria_e_preservacao. Acesso em: 02 jun. 2021.

| 22

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. Disponível em: <https://www.scielo.br/j/tinf/a/LSxTfhK6NfX54t4ypBK87kM/?lang=pt#>. Acesso em: 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. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/96602.96604>. Acesso em: 02 jun. 2021.

SOMPEL, H. V. **ResourceSync**. St. Louis, MO: CNI 2014 Spring Membership Meeting. jun. 2014. Disponível em: <https://www.niso.org/standards-committees/resourcesync>. Acesso em: 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. Disponível em: <http://hdl.handle.net/10760/8168>. Acesso em: 02 jun. 2021.

WORLD WIDE WEB CONSORTIUM. **RDF 1.1 Primer**, 2014. Disponível em: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>. Acesso em: 02 jun. 2021.

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