

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”
FACULDADE DE FILOSOFIA E CIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

BRUNA MARIA CAMPOS DA CUNHA

**INTEROPERABILIDADE EM PROVEDORES DE DADOS E
PROVEDORES DE SERVIÇOS:
uma análise dos metadados e protocolos OAI-PMH e OAI-ORE**

BRUNA MARIA CAMPOS DA CUNHA

**INTEROPERABILIDADE EM PROVEDORES DE DADOS E
PROVEDORES DE SERVIÇOS:
uma análise dos metadados e protocolos OAI-PMH e OAI-ORE**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências - Universidade Estadual Paulista “Júlio de Mesquita Filho”, como requisito para obtenção do título de Mestre em Ciência da Informação.

Orientadora: Prof.^a Dr.^a Rachel Cristina Vesu Alves
Linha de pesquisa: Informação e Tecnologia

**Marília
2021**

C972i

Cunha, Bruna Maria Campos da

Interoperabilidade em provedores de dados e provedores de serviços : uma análise dos metadados e protocolos OAI-PMH e OAI-ORE / Bruna Maria Campos da Cunha. -- Marília, 2021
68 p.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília

Orientadora: Rachel Cristina Vesu Alves

1. Interconexão em rede (Telecomunicações). 2. Redes de computadores Protocolos. 3. Metadados. I. Título.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Marília

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: INTEROPERABILIDADE EM PROVEDORES DE DADOS E PROVEDORES DE SERVIÇOS: uma análise dos metadados e protocolos OAI-PMH e OAI-ORE

AUTORA: BRUNA MARIA CAMPOS DA CUNHA

ORIENTADORA: RACHEL CRISTINA VESU ALVES

Aprovada como parte das exigências para obtenção do Título de Mestra em CIÊNCIA DA INFORMAÇÃO, área: Informação, Tecnologia e Conhecimento pela Comissão Examinadora:

Prof(a). Dr(a). RACHEL CRISTINA VESU ALVES (Participação Virtual)
Departamento de Ciência da Informação / Unesp, Faculdade de Filosofia e Ciências, Marília

Prof(a). Dr(a). EDBERTO FERNEDA (Participação Virtual)
Departamento de Ciência da Informação / Unesp, Faculdade de Filosofia e Ciências, Marília

Prof(a). Dr(a). FELIPE AUGUSTO ARAKAKI (Participação Virtual)
Curso de Biblioteconomia / Universidade de Brasília

Marília, 15 de abril de 2021

Profa. Dra. Marta Lígia Pomim Valentim
Coordenadora do PPG em Ciência da Informação

AGRADECIMENTOS

Agradeço primeiramente a Deus, por proporcionar tudo nessa vida.

Agradeço à minha orientadora Rachel Cristina Vesu Alves, pela oportunidade, confiança, paciência e pelas diversas trocas de muito conhecimento.

Aos professores Edberto Ferneda e Felipe Augusto Arakaki, por disponibilizarem seu tempo, conhecimento e paciência para contribuírem de forma significativa com esta pesquisa.

Agradeço também a todos os professores do Programa de Pós-Graduação em Ciência da Informação, com os quais, mesmo indiretamente, pude presenciar trocas de conhecimento.

Aos funcionários da biblioteca da FFC, com os quais pude contar em vários momentos durante o desenvolvimento desta pesquisa.

Aos velhos amigos que pude reencontrar nesta fase da vida e que me ajudaram a passar por momentos de dúvidas, e aos novos amigos que pude fazer, os quais fizeram parecer que as dúvidas não existiam.

Aos familiares e amigos que acreditaram em mim e me apoiaram, mesmo que de longe, nessa nova caminhada.

Ao amado Luiz Galeffi que pacientemente normalizou e revisou todo esse trabalho.

Por fim, a todos aquelas que direta ou indiretamente contribuíram com toda essa fase, para que eu conseguisse chegar até aqui e que porventura tenha me esquecido de mencionar.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de financiamento 001.

A todos, o meu Muito obrigada!

*Anywhere you'll ever go
Everyone you'll ever know
You may never find it
Everyone you'll ever see
Anyone you choose to be
How you look will decide
What you see.*

A-HA

RESUMO

A interoperabilidade é algo crucial para os sistemas informacionais digitais, uma vez que esta permite a troca de dados entre diferentes sistemas. A pesquisa tem como foco o estudo dos protocolos para promover a interoperabilidade na web entre repositórios de dados e de serviços. O problema se dá em como promover a interoperabilidade, o uso e o reuso dos dados nestes provedores, por meio do uso de protocolos e padrões de metadados bibliográficos. Para tanto, tem como objetivo geral analisar os metadados bibliográficos, com foco nos provedores de dados e provedores de serviços para a web e nos protocolos OAI-PMH e OAI-ORE, e a justificativa se dá ao observar poucos estudos desenvolvidos em língua portuguesa sobre o tema. É possível verificar que a utilização do protocolo OAI-PMH é muito maior, porém a ampla utilização do OAI-ORE traria vantagens ainda não exploradas pela maioria dos provedores de dados e de serviços. Conclui-se que o protocolo OAI-PMH é mais eficaz em expor e coletar metadados, enquanto o OAI-ORE proporciona uma descrição melhor em termos de semântica, o que leva a crer que uma junção dos dois teria uma importância significativa para o uso e reuso dos dados e para a recuperação de recursos informacionais.

Palavras-chave: Interoperabilidade. Protocolos. OAI-PMH. OAI-ORE.

ABSTRACT

Interoperability is crucial for digital information systems, since it allows the exchange of data between different systems. The research focuses on the study of protocols to promote interoperability on the web between data and service repositories. The problem is how to promote interoperability, use and reuse of data at these providers, through the use of bibliographic metadata protocols and standards. To this end, its general objective is to analyze bibliographic metadata, focusing on data providers and service providers for the web and on the OAI-PMH and OAI-ORE protocols, and the justification is given by observing few studies developed in Portuguese on this subject. It is possible to verify that the use of the OAI-PMH protocol is much greater, however the widespread use of OAI-ORE would bring advantages not yet explored by most data and service providers. It is concluded that the OAI-PMH protocol is more effective in exposing and collecting metadata, while OAI-ORE provides a better description in terms of semantics, which suggests that a combination of the two would be of significant importance for the use and reuse of data and the recovery of information resources.

Keywords: Interoperability. Protocols. OAI-PMH. OAI-ORE.

LISTA DE FIGURAS E QUADROS

Figura 1 - Esquema geral da pesquisa	17
Figura 2 - Exemplo de interoperabilidade técnica	22
Figura 3 - Exemplo de interoperabilidade semântica	23
Figura 4 - Exemplo de interoperabilidade política/humana.....	24
Figura 5 - Exemplo de interoperabilidade intercomunitária.....	24
Figura 6 - Exemplo de interoperabilidade legal	25
Figura 7 - Exemplo de interoperabilidade internacional	26
Figura 8 - Exemplo de federação.....	28
Figura 9 - Exemplo de <i>harvesting</i>	30
Figura 10 - Exemplo de agregação.....	31
Figura 11 - Grafo RDF com recurso-propriedade-valor.....	42
Figura 12 - Exemplo de grafo RDF aplicado a esta pesquisa.....	43
Figura 13 - Página de jornal descrita em grafos RDF	45
Figura 14 - Exemplo de agregação do OAI-ORE.....	50
Figura 15 - Modelo completo do OAI-ORE.....	51
Figura 16 - Exemplo de descrição do OAI-ORE.....	52
Figura 17 - Exemplo de descrição em RDF e OAI-ORE	56
Figura 18 - Estrutura básica do EDM e identificação do <i>proxy</i>	58
Figura 19 - Exemplo de um sistema informacional.....	59
Quadro 1 - Diferentes aspectos da interoperabilidade tratados na literatura	21

LISTA DE ABREVIACÕES E SIGLAS

@archivesSIC	<i>Archive Ouverte em Sciences de l'Information et de la Communication</i>
AACR2r	<i>Anglo-American Cataloguing Rules, segunda edição</i>
BDTD	<i>Biblioteca Digital Brasileira de Teses de Dissertações</i>
CDD	<i>Classificação Decimal de Dewey</i>
CNRI	<i>Corporation for National Research Initiative</i>
DICI	<i>Diálogo Científico</i>
dLIST	<i>Digital Library of Information Science and Technology</i>
E-LIS	<i>E-prints in Library and Information Science</i>
HTTP	<i>HyperText Transfer Protocol</i>
IETF	<i>Internet Engineering Task Force</i>
IBICT	<i>Instituto Brasileiro de Informação em Ciência e Tecnologia</i>
Intercom	<i>Revista Brasileira de Ciências da Comunicação</i>
ISO	<i>International Organization for Standardization</i>
JISC	<i>Joint Information Systems Committee</i>
LDL	<i>Librarian's Digital Library</i>
MARC21	<i>Machine-Readable Cataloging</i>
MémSIC	<i>Mémoires de 3^o cycle en Science de l'Information et de la Communication</i>
NCSA	<i>National Center for Supercomputing Applications</i>
OAI	<i>Open Archives Initiative</i>
OAI-ORE	<i>Open Archives Initiative Object Reuse and Exchange</i>
OAI-PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
OCLC	<i>Online Computer Library Center</i>
OWL	<i>Web Ontology Language</i>
PICS	<i>Platform for Internet Content Selection</i>
RBH	<i>Revista Brasileira de Herbicidas</i>
RDF	<i>Resource Description Framework</i>
SABIIA	<i>Sistema Aberto e Integrado de Informação em Agricultura</i>
TIC	<i>Tecnologias de Informação e Comunicação</i>
UFPE	<i>Universidade Federal de Pernambuco</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>eXtensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Problema de pesquisa e hipótese	12
1.2	Objetivos	12
1.2.1	Objetivo geral.....	12
1.2.2	Objetivos específicos	13
1.3	Justificativa	13
1.4	Metodologia e estrutura do trabalho	14
2	INTEROPERABILIDADE	19
2.1	Aspectos da interoperabilidade	20
2.1.1	Interoperabilidade técnica	22
2.1.2	Interoperabilidade semântica	22
2.1.3	Interoperabilidade política/humana.....	23
2.1.4	Interoperabilidade intercomunitária	24
2.1.5	Interoperabilidade legal.....	24
2.1.6	Interoperabilidade internacional.....	25
2.2	Níveis de interoperabilidade	26
2.2.1	Federação	26
2.2.2	<i>Harvesting</i>	28
2.2.3	Agregação	30
3	SERVIÇOS DE METADADOS	32
3.1	Provedores de dados e provedores de serviços	36
3.1.1	OAI-PMH.....	37
3.2	<i>Resource Description Framework (RDF)</i>	39
4	O <i>LINKED DATA</i> E O OAI-ORE	46
5	CONSIDERAÇÕES FINAIS	60
	REFERÊNCIAS	63

1 INTRODUÇÃO

O estabelecimento da Web Semântica e do *Linked Data* tem causado inúmeras discussões em todas as áreas do conhecimento, incluindo o domínio bibliográfico. O *Linked Data* amplia o acesso aos recursos informacionais, disponibilizados em diferentes ambientes e domínios, mas que podem estar relacionados de algum modo. Um texto pode estar ligado a uma imagem que, por sua vez, pode estar ligada a um vídeo e por essas associações os usuários acessam uma variedade maior de tipos de materiais e assuntos relacionados, levando a todos os tipos de *links* possíveis (WITT, 2010).

Considerando o número crescente de repositórios e seus variados tipos, institucionais, temáticos, científicos, de pesquisa, entre outros; e considerando o contexto da Web Semântica e *Linked Data*, os profissionais que trabalham com a representação nesses ambientes passam a ter o desafio de promover a interoperabilidade e a integração de acervos heterogêneos. Assim, além dos repositórios provedores de dados, os profissionais passam a atuar em um novo ambiente informacional, os provedores de serviços, que disponibilizam a busca integrada em acervos de diferentes tipos de repositórios e, além disso, podem ter seus acervos disponibilizados no contexto da Web Semântica e *Linked Data*.

Assim, o desafio atual em relação à integração de acervos heterogêneos refere-se ao estabelecimento da interoperabilidade entre repositórios que atuam como provedores de dados e provedores de serviços e ainda o uso e reuso desses dados na *web* e no *Linked data*.

Para que o acesso aos recursos informacionais aconteça, assim como o uso e o reuso dos dados, torna-se necessária a utilização de padrões de metadados e protocolos para estabelecer certos níveis de interoperabilidade como, por exemplo, os níveis de *harvesting*, agregação e federação (SAYÃO, 2007).

Para Garcia e Sunye (2003), os participantes da *Open Archives Initiative* (OAI) podem ser divididos em provedores de dados e provedores de serviços. Os provedores de dados são os repositórios que mantém seus metadados expostos a partir de um protocolo, já os provedores de serviços oferecem as buscas a esses metadados ou outros serviços, agregando valor à iniciativa.

Os provedores de dados (um repositório digital institucional, temático ou de pesquisa, por exemplo) e os provedores de serviços (um repositório que congrega vários outros repositórios digitais por meio de uma plataforma comum de acesso) utilizam metadados e padrões de metadados de diferentes tipos e níveis, devido a heterogeneidade de acervos. A

interoperabilidade é um fator importante de ser estabelecido nesses ambientes, assim como o uso e reuso desses dados na *web* e no *Linked Data*.

A interoperabilidade pode ser entendida como a capacidade de comunicação e intercâmbio de dados entre sistemas diferentes, independentemente do tipo de *hardware* e *software*, além de ser considerada como a peça-chave para o uso e reuso de informações nos ambientes digitais (SAYÃO; MARCONDES, 2008).

Do ponto de vista bibliográfico, mais especificamente no âmbito das bibliotecas e repositórios digitais, o conceito de interoperabilidade é complexo e carrega uma diversidade de visões. Podemos citar o exemplo trazido pela colaboração da *Digital Library Research Group* e *Corporation for National Research Initiative* (CNRI), de que a interoperabilidade é a habilidade dos componentes e serviços serem funcionais e logicamente interoperáveis, por terem sido implantados de acordo com um modelo bem definido e publicações conhecidas (PAYETTE et al., 1999).

A interoperabilidade deve ser estabelecida nos seguintes âmbitos: técnico, semântico, político e humano, intercomunitário, legal e internacional. Assim, observa-se que o estabelecimento da interoperabilidade é algo complexo. Os metadados, padrões de metadados e protocolos são instrumentos que estarão presentes no estabelecimento da interoperabilidade, principalmente, nos âmbitos técnicos e semânticos.

O padrão de metadados *Dublin Core* é reconhecido internacionalmente como um padrão interoperável e está presente na maioria dos repositórios digitais. Porém, podemos encontrar outros tipos de padrões sendo utilizados. O padrão foi idealizado pela *Online Computer Library Center* (OCLC) e pela *National Center for Supercomputing Applications* (NCSA), e vem sendo amplamente utilizado a partir de 1995 (ARAKAKI; SANTOS; ALVES, 2015). O *Dublin Core* apresenta 15 elementos (metadados básicos) para a identificação e descrição mínima dos recursos no ambiente *web*. Além disso, apresenta também um nível qualificado, com elementos de refinamento e elementos adicionais para representação e a possibilidade de extensibilidade dos metadados em seu esquema.

O padrão apresenta como característica ser um padrão simples, extensível e interoperável. Além da sintaxe *eXtensible Markup Language* (XML) o padrão também disponibiliza seus metadados na sintaxe do *Resource Description Framework* (RDF) (GRÁCIO, 2002; ALVES; SOUZA, 2007).

O RDF tem por objetivo “[...] definir um mecanismo de representação de metadados para descrever recursos não vinculados a um domínio específico de aplicação” (DZIEKANIAK; KIRINUS, 2004, p. 27).

1.1 Problema de pesquisa e hipótese

A presente pesquisa tem como tema o estudo dos metadados e da interoperabilidade, sendo que o foco consiste no estudo dos protocolos para promover a interoperabilidade na *web* entre repositórios de dados e de serviços.

Deste modo, o problema desta pesquisa está em como promover a interoperabilidade, o uso e reuso dos dados em provedores de dados e provedores de serviços, utilizando protocolos e padrões de metadados bibliográficos.

Para atingir o nível mais alto de interoperabilidade, o de federação, é necessário atingir os níveis mais básicos e intermediários, por meio do *harvesting* e da agregação. O *harvesting* é responsável pela coleta de metadados em provedores de dados e serviços; e a agregação possibilita a criação de pacotes com recursos, suas descrições, sua origem e sua situação dos provedores de dados e serviços. Assim, a hipótese desta pesquisa é de que o protocolo OAI-PMH e OAI-ORE são tecnologias relevantes para proporcionar o acesso, o uso e reuso dos dados de provedores de dados e serviços na *web* e no *Linked Data*. Torna-se necessário neste cenário heterogêneo criar níveis diferentes de interoperabilidade, portanto, acredita-se que essas tecnologias poderão contribuir para o estabelecimento de níveis de interoperabilidade.

Deste modo, a proposta desta pesquisa é estudar o estabelecimento da interoperabilidade, do uso e reuso dos dados em provedores de dados e serviços, com base nos protocolos OAI-PMH e OAI-ORE.

1.2 Objetivos

1.2.1 Objetivo geral

O objetivo geral da pesquisa é analisar os metadados bibliográficos, com foco nos provedores de dados e provedores de serviços para a *web*, além dos protocolos OAI-PMH (para o *harvesting* de metadados) e OAI-ORE (para níveis de agregação), para o estabelecimento da interoperabilidade, do uso e reuso dos dados em provedores de dados e provedores de serviços.

1.2.2 Objetivos específicos

- Estudar as principais características e níveis de interoperabilidade estabelecidos na literatura;
- Estudar os aspectos atuais para o uso e reuso dos dados disponibilizados na *web* e no contexto do *Linked Data*;
- Analisar as características dos protocolos OAI-PMH e OAI-ORE e sua aplicação em provedores de dados e serviços;
- Analisar as contribuições dos protocolos citados para os ambientes informacionais digitais na *web* e no *Linked Data*.

1.3 Justificativa

O interesse para a realização deste projeto está na experiência anteriormente adquirida no trabalho em um repositório digital e o estudo anteriormente desenvolvido na graduação sobre os protocolos de interoperabilidade, no qual foi possível observar a existência de vários protocolos para a interoperabilidade e poucos estudos desenvolvidos em língua portuguesa sobre o tema.

Sayão (2007, p. 27) aponta que “O sucesso de um sistema de informações distribuídas - tal como é caracterizada a própria World Wide Web - depende fortemente de vinculação constante entre os recursos que estão disponíveis on-line”. A partir deste ponto de vista, entende-se que os metadados bibliográficos, metadados *web* e protocolos desempenham um papel importante no que se refere ao uso, reuso e interoperabilidade dos dados.

A interoperabilidade possibilita maior economia de tempo, exposição e interpretação de padrões de metadados, maior precisão e facilidade na realização de uma busca (GUAJARDO SALINAS, 2010). Assim, com crescente disponibilização dos recursos informacionais na *web*, as bibliotecas e repositórios digitais necessitam estabelecer ainda mais a interoperabilidade entre seus sistemas, para possibilitarem um acesso mais amplo aos recursos informacionais e serviços de busca com maior qualidade.

Do ponto de vista acadêmico, a pesquisa se justifica por trazer um tema pouco explorado, portanto, que necessita de mais análises e discussões. A realização desta pesquisa contribui para a área da Ciência da Informação, pois expande os horizontes na disponibilização

de resultados informacionais para diferentes usuários, evitando o retrabalho dos bibliotecários, uma vez que se faz possível o compartilhamento dos metadados com as descrições dos recursos.

As contribuições sociais desta pesquisa está em contribuir para a criação de uma grande rede de informações estruturadas, de sistemas que possibilitam navegar para fora dele através de materiais relacionados possibilitando resultados muito mais ricos e precisos.

1.4 Metodologia e estrutura do trabalho

A presente pesquisa tem natureza qualitativa, de cunho teórico. Consiste na análise exploratória e descritiva da literatura já existente e disponível sobre o tema (MARCONI; LAKATOS, 2006).

A pesquisa qualitativa é aquela que envolve diversas referências epistemológicas para se fundamentar, e pode ser realizada a partir da exploração de documentos para levantar informações sobre um determinado objeto e campo de trabalho (SEVERINO, 2007).

De acordo com Bardin (2011), uma análise qualitativa possui “características peculiares” e está especificada em elaboração de deduções e inferências. Assim, segundo a autora, a compreensão do sentido do que está escrito é importante, e se tem um maior risco de erro em relação à análise quantitativa, pois se trabalha com elementos isolados, ou seja, não há a necessidade da existência de dados quantitativos para auxiliar na compreensão, sendo muito importante o contexto desses elementos.

Para a realização do levantamento dos materiais relevantes para o tema, foram utilizadas palavras-chave em base de dados como *Web of Science*, BDTD, BRAPCI, *Scopus* e *SciELO* e fontes de informação em Ciência da Informação, repositórios, bibliotecas, catálogos de bibliotecas, em especial os da área de Ciência da Informação, procurando identificar trabalhos publicados, teses, dissertações, artigos de periódicos, entre outros materiais que abordem o tema.

Um dos métodos utilizados é o método de Análise de Conteúdo, o qual surge como um instrumento com o intuito de analisar “o que essa mensagem significa exatamente” (BARDIN, 2011; CAMPOS, 2004).

O desejo de entender o significado das mensagens surgiu a partir da decodificação minuciosa dos símbolos e mensagens encontradas nos textos bíblicos. No ano de 1640, na Suécia, uma considerada prematura Análise de Conteúdo é referenciada sob a perspectiva de pesquisa sobre a autenticidade de textos religiosos e sua possível influência sobre Luteranos (CAMPOS, 2004).

Assim, a Análise de Conteúdo constitui-se em interpretar o conteúdo de uma classe de documentos. Essa análise permite a reinterpretção das mensagens contidas no conjunto de documentos analisados e uma compreensão maior dos significados dessas mensagens do que uma leitura comum (MORAES, 1999).

De cunho científico, a Análise de Conteúdo desenvolveu-se no início do século nos Estados Unidos.

Posteriormente à Primeira Guerra Mundial, inicia-se o estudo e análises da imprensa e da propaganda desde 1915 com Harold Lasswell, que publica, em 1927 as análises da imprensa e propaganda desta época: *Technique in the World War* (BARDIN, 2011; CAMPOS, 2004).

Na década de 40, os acontecimentos da Segunda Guerra Mundial e os interesses políticos nos Estados Unidos fizeram com que a Análise de Conteúdo fosse amplamente aplicada nas propagandas dos jornais e revistas da época. Lasswell, da Universidade de Chicago continuava seus estudos sobre as análises de símbolos e se projetavam também Berelson e Lazarsfeld, publicando o livro *The analysis of communication contents*, em 1948 (CAMPOS, 2004).

A partir desses três principais autores, a Análise de Conteúdo vem sendo aperfeiçoada e utilizada em diferentes domínios, uma vez que pode ser considerada como um conjunto de técnicas de análise de comunicações (CAMPOS, 2004).

Constituída de um leque de ferramentas aplicável a um vasto campo, a Análise de Conteúdo também pode ser aplicada a documentos. Pode-se dizer que esta análise, aplicada ao ambiente de documentos, corresponde aos objetivos de:

- “a ultrapassagem da incerteza”: o que se vê na mensagem, é o que está efetivamente contido nela?
- “o enriquecimento da leitura”: pode uma leitura atenta aumentar a pertinência e a produtividade? (BARDIN, 2011).

Assim, a análise de conteúdo de mensagens, que deveria ser aplicada a qualquer forma de comunicação, possui duas funções:

- função heurística: enriquece a tentativa exploratória e aumenta a propensão à descoberta;
- função de administração de provas: a Análise de Conteúdo pode ser aplicada numa análise sistemática para se ter informação e/ou confirmação de hipóteses (BARDIN, 2011).

A utilização da Análise de Conteúdo em documentos, como apontado por Campos (2004, p. 613), inicia-se por:

Selecionado o *corpus* a ser analisado procede-se às leituras flutuantes de todo o material, com o intuito de apreender e organizar de forma não estruturada aspectos importantes para as próximas fases da análise. Na leitura flutuante toma-se contato com os documentos a serem analisados, conhece-se o contexto e deixa-se fluir impressões e orientações.

A partir disso, pode-se dividir a Análise de Conteúdo, com foco em documentos, em: pré-análise, exploração do material e o tratamento, inferência e interpretação dos resultados (BARDIN, 2011).

A pré-análise tem por objetivo a organização pela exposição à exploração dos documentos. Possui a missão da escolha dos documentos a serem analisados, a formulação das hipóteses e dos objetivos e a elaboração dos indicadores que fomentarão a interpretação final (BARDIN, 2011).

Dentro da pré-análise, existem algumas regras a serem seguidas para a seleção do documento, como por exemplo, regras de exaustividade, representatividade e pertinência do documento que será analisado. Essa primeira parte é flexível, pois deve permitir a formulação e reformulação das estratégias para a seleção dos materiais (BARDIN, 2011).

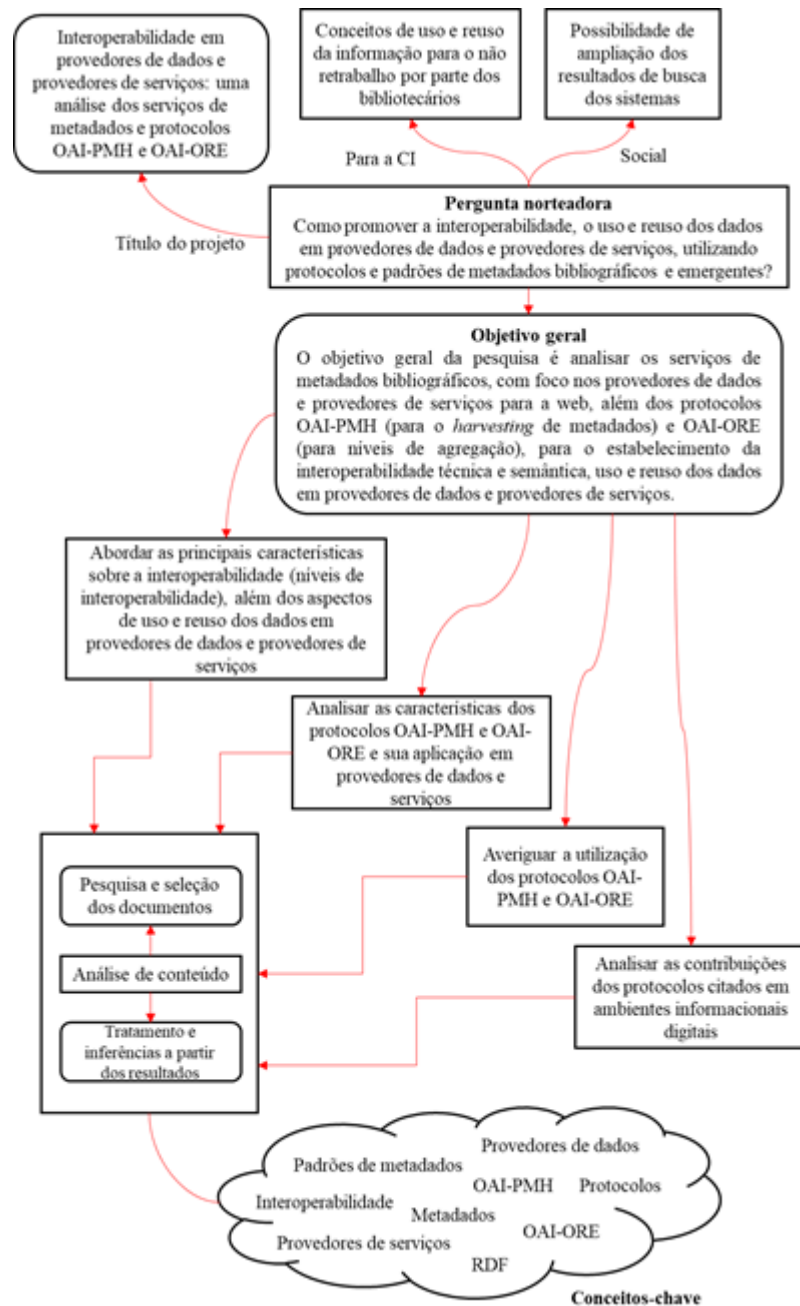
A exploração do material, para Bardin (2011), é a fase mais longa, pois consiste em operações de codificação ou enumeração do material em função das regras previamente formuladas.

A fase do tratamento, inferência e interpretação dos resultados se trata da fase em que os resultados ‘brutos’ obtidos se tornam válidos e “falantes”. Com esses resultados à disposição, se faz as interpretações e inferências à propósito dos objetivos previstos (BARDIN, 2011).

Por suas ferramentas e capacidade de permitir inferências, a Análise de Conteúdo possibilita uma interpretação dos documentos com temas relevantes ao tema, permitindo uma análise minuciosa das mensagens contidas nesses documentos e a codificação dos resultados teóricos almejados.

A Figura 1 apresenta o esquema geral desta pesquisa, com base nos temas de estudos aqui abordados:

Figura 1 - Esquema geral da pesquisa



Fonte: Elaborado pela autora, com base em Camperos Reyes (2018)

Para tanto, a presente pesquisa apresenta a seguinte metodologia e método:

- Etapa 1: realização das pesquisas nas bases de dados e outras fontes de informação da área da Ciência da Informação, através das palavras-chave: metadados; interoperabilidade; protocolo; provedores de dados; provedores de serviços; OAI-PMH; OAI-ORE;
- Etapa 2: a seguir, inicia-se a utilização do método da Análise de Conteúdo com a pré-análise dos documentos encontrados para a seleção do corpus da pesquisa. É

nessa fase que se verifica a relevância e a pertinência do material para a pesquisa, além da delimitação, à qual, para a presente pesquisa, serão considerados documentos do ano de 2008, de quando o protocolo OAI-ORE foi criado, até os mais atuais (2020);

- Etapa 3: a exploração do material, segunda etapa da Análise de Conteúdo, permite codificar e elencar os elementos mais importantes para a pesquisa, através de uma análise minuciosa dos documentos;
- Etapa 4: a terceira e última fase da Análise de Conteúdo é a fase do tratamento dos resultados coletados nas etapas anteriores. Nessa fase, permite que se realizem inferências e interpretações daquilo que foi analisado. A partir disso, o resultado é apresentado e discutido ao final da pesquisa.

Assim, através da Análise de Conteúdo de documentos, pretende selecionar os documentos a serem analisados a partir da pertinência desses com relação ao tema principal da pesquisa; fazer uma análise das mensagens contidas nesses documentos a fim de fazer inferências dessas mensagens e responder à questão inicial e abordar os objetivos da pesquisa.

Tomando a introdução como o primeiro capítulo, tem-se um panorama geral dos assuntos que serão abordados na pesquisa, quais são os objetivos que esta pretende alcançar e com quais métodos pretende realizar.

No segundo capítulo é abordada a interoperabilidade, esmiuçando seu surgimento, definições e diferentes aspectos com qual é tratada dentro da literatura. Nesse capítulo também são abordados diferentes tipos de interoperabilidade e como são entendidos.

Os serviços de metadados assim como provedores de dados e provedores de serviços são abordados no terceiro capítulo. Pretende-se aqui, ressaltar a importância dos metadados bem estruturados, além de como são utilizados dentro de provedores de dados e provedores de serviços, de modo que os sistemas se beneficiem da troca e reuso destes.

No quarto capítulo discorre-se acerca do conceito de *Linked Data* e do protocolo OAI-ORE, identificando a proximidade e o que tal protocolo trás em consonância com o conceito.

Assim, no capítulo cinco são apresentados os resultados obtidos a partir das discussões realizadas nos capítulos anteriores, consolidando os principais pontos de cada um.

A conclusão é apresentada no último capítulo, nela, são apresentadas as conclusões desta pesquisa, levando em consideração o problema inicial e os objetivos que se pretendia alcançar.

2 INTEROPERABILIDADE

Nascido em meio à explosão informacional e consolidado já no meio digital, o termo interoperabilidade se encontra no contexto da comunicação entre os ambientes informacionais digitais.

Levantados inicialmente em 1998 na *An international research agenda for digital libraries* (Uma agenda de pesquisa dos Estados Unidos e União Europeia), hoje os conceitos acerca da interoperabilidade aparecem de diferentes formas na literatura, porém, dizem respeito à possibilidade da realização de uma única busca, através de uma interface, em sistemas heterogêneos, com recursos armazenados em diferentes servidores (MARCONDES; SAYÃO, 2001).

Assim como outros, o termo interoperabilidade carrega um conjunto de definições que o representam de forma mais geral ou mais específica de acordo com o domínio ao qual este se apresenta. Dentre os conceitos utilizados na definição do termo, entende-se que:

[...] refere-se à capacidade de sistemas e organizações trabalhar juntos. Informações de diversos setores podem ser interoperáveis de várias maneiras - através do desenvolvimento de produtos, através de parcerias entre indústria e comunidade, através do acesso propriedade intelectual e tecnologia, e através da implementação de padrões. (ELLIOTT, 2008, p. 10, tradução nossa)

A norma da *International Organization for Standardization (ISO) 25964-2:2013* traz a seguinte definição acerca da interoperabilidade: “[...] é a capacidade de sistemas diversos, com diferentes hardwares e plataformas de softwares, estruturas de dados e interfaces, trocarem dados com a mínima perda de conteúdo e funcionalidade” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013, p. 16, tradução nossa).

Na visão de Miller (2000), tomando por base o dicionário conciso de Oxford, a interoperabilidade é a capacidade de um sistema ou produto de trabalhar com outros sistemas e produtos sem a necessidade de um esforço por parte do cliente, tornando a interoperabilidade de uma qualidade crescente, uma vez que sua rede é a de computadores.

Para Sayão e Marcondes (2008), ao mesmo tempo em que a interoperabilidade pode ser entendida como a capacidade de programas e computadores de fabricantes diferentes comunicarem entre si, levando em consideração uma visão mais “computadorizada”, o conceito pode se tornar mais complexo quando visto pela área da Ciência da Informação, uma vez que, além dessa visão da comunicação entre programas e computadores, o conceito também diz

respeito à comunicação entre os ambientes informacionais digitais, ampliando a capacidade de cada um desses sistemas desempenharem seu papel de disseminar a informação.

Nessa mesma vertente, Innocenti, Vullo e Ross (2010, p. 29, tradução nossa) trazem que:

A interoperabilidade é uma propriedade referente à capacidade dos sistemas e das organizações de trabalharem em conjunto. Hoje, a interoperabilidade é reconhecida como um passo fundamental na mudança de bibliotecas digitais isoladas para um espaço de informação comum que permitirá aos usuários navegar através de diferentes bibliotecas digitais dentro de um único ambiente integrado.

Em resumo, é a interoperabilidade que permite a descoberta de informações pelos usuários entre sistemas heterogêneos, além de facilitar a entrega das informações, suportar a interação entre os usuários e os sistemas de informação e facilitar a comunicação entre as bibliotecas, arquivos e museus. É por tais razões que se faz importante o desenvolvimento contínuo das técnicas e padrões de interoperabilidade, para aprimorar as descobertas, entregas, interações e colaborações permitidas (ELLIOTT, 2008).

Assim, tem-se para esta pesquisa que a interoperabilidade é a capacidade de dois ou mais sistemas compartilharem metadados de forma transparente, com o mínimo de esforços e perdas possíveis, mesmo que entre sistemas heterogêneos.

2.1 Aspectos da interoperabilidade

Dentro dos ambientes informacionais como os repositórios digitais e do ponto de vista dos usuários, uma interação com diversos ambientes para pesquisa através de uma única interface, é algo de muita valia e conforto (MARCONDES; SAYÃO, 2001).

Com diversas formas de promover a interoperabilidade, ou seja, a comunicação entre os sistemas, autores como Arms et al. (2002) e Sayão e Marcondes (2008) apontam diferentes divisões, características, aspectos e níveis sobre a interoperabilidade, conforme compilado no Quadro 1 a seguir.

Quadro 1 - Diferentes aspectos da interoperabilidade tratados na literatura

Autores		Arms et al. (2002)	Sayão e Marcondes (2008)	Miller (2000)	Fusco (2011)	Alves e Souza (2007)	Innocenti, Vullo e Ross (2010)
Aspectos	Interoperabilidade técnica	X	X	X		X	X
	Interoperabilidade estrutural				X		
	Interoperabilidade de conteúdo	X				X	X
	Interoperabilidade semântica		X	X	X		
	Interoperabilidade organizacional	X				X	
	Interoperabilidade internacional		X	X			
	Interoperabilidade sintática				X		
	Interoperabilidade baseada em serviços						X
	Interoperabilidade política/humana		X	X			
	Interoperabilidade baseada no usuário						X
	Interoperabilidade legal		X	X			
	Interoperabilidade intercomunitária		X	X			
	Níveis	Federação	X	X			
<i>Harvesting</i>		X	X				
Agregação		X	X				

Fonte: Elaborado pela autora

No quadro anterior é apontado quais os termos mais utilizados e quais autores utilizam cada termo. Nesta pesquisa, segue-se as divisões e aspectos de interoperabilidade trazidos por Sayão e Marcondes (2008), uma vez que estes tomam por base e complementam os autores mencionados anteriormente para chegarem aos aspectos descritos a seguir.

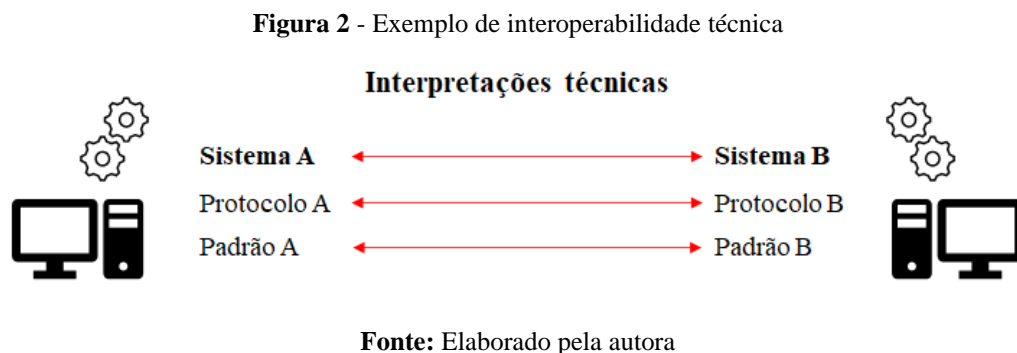
Sendo assim, os aspectos apontados pelos autores Sayão e Marcondes (2008) serão destacados a seguir, com complementações de demais autores: interoperabilidade técnica, interoperabilidade semântica, interoperabilidade política/humana, interoperabilidade intercomunitária, interoperabilidade legal e interoperabilidade internacional.

2.1.1 Interoperabilidade técnica

Traz as configurações sobre os aspectos técnicos, como por exemplo, padrões e protocolos a serem utilizados para assegurar a interoperabilidade (SAYÃO; MARCONDES 2008).

Para Miller (2000), a interoperabilidade técnica é o aspecto mais diretamente ligado à manutenção da interoperabilidade, uma vez que nela existem respostas certas e erradas de forma clara. Nesse caso, estão inclusas questões mais técnicas para a garantia de um envolvimento e um desenvolvimento contínuo dos padrões de transporte, armazenamento e representação de dados e informação.

Uma metáfora utilizada por Chris Rusbridge, ex-programador de bibliotecas eletrônicas da *Joint Information Systems Committee* (JISC), compara a interoperabilidade técnica a encanamentos escondidos atrás de um aplicativo bem-sucedido (MILLER, 2000). Um exemplo de interoperabilidade técnica pode ser visto na Figura 2:



2.1.2 Interoperabilidade semântica

Ao chegar na interoperabilidade semântica, as questões até então pontuais para a interoperabilidade técnica, tornam-se mais profundas à medida em que diz respeito a cada recurso de modo individual (MILLER, 2000).

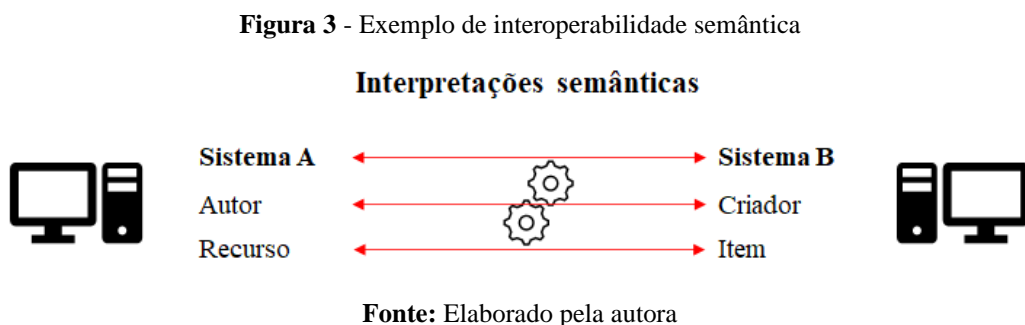
Esse tipo de interoperabilidade está relacionado ao significado das informações, ou seja, dos diferentes recursos aos quais estão relacionadas. Quando semanticamente bem estruturadas, são disponibilizadas por meio de *gateways*, embora não seja raro encontrar diferentes termos descrevendo conceitos semelhantes, por exemplo, “Autor, Criador ou Compositor” (MILLER, 2000; SAYÃO; MARCONDES, 2008).

Uma das soluções para a interoperabilidade semântica se dá através de ferramentas mapeáveis de representação da informação (SAYÃO; MARCONDES, 2008).

Sendo assim, Drame (2008, p. 14, tradução nossa) entende que:

A interoperabilidade semântica refere-se à capacidade de dois ou mais sistemas de computador trocarem e interpretar informações com precisão, sem a necessidade de intervenção humana. É, portanto, crucial para a recuperação, reutilização e compartilhamento de informações - que é o objetivo real de usar ontologias.

Assim, nesta pesquisa entende-se que a interoperabilidade semântica trata de questões mais profundas e individuais de cada recurso informacional de modo a visualizá-lo de forma única e de entendimento ao usuário não máquina. A Figura 3 ilustra um exemplo de interoperabilidade semântica:

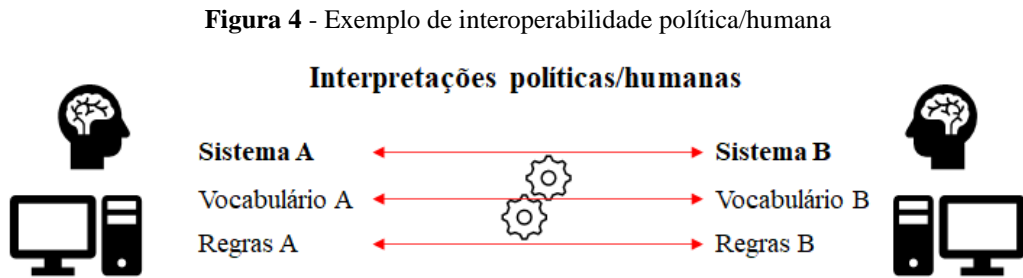


2.1.3 Interoperabilidade política/humana

Além das questões de como as informações serão disseminadas e descritas, deve-se pensar em como tornar os recursos ainda mais acessíveis, levando em consideração as organizações às quais pertencem, seus funcionários e seus usuários finais. Assim, a interoperabilidade política/humana diz respeito à democratização do acesso à informação, independentemente da maneira de como a informação é organizada e disseminada (MILLER, 2000; SAYÃO; MARCONDES, 2008).

As organizações envolvidas podem entender esse tipo de interoperabilidade como uma perda de controle de propriedade e tanto seus funcionários como usuários finais podem não ter as habilidades necessárias para suportar esses sistemas mais complexos, uma vez que mudanças nos processos e treinamentos extensivos raramente são realizados pelas organizações, principalmente para a tomada de decisão de liberar ou não certo recurso. Porém, esse tipo de atividade é essencial para garantir um uso eficaz e prolongado de qualquer recurso.

Além disso, questões como a exclusão social e a disseminação generalizada de informações começam a ser visualizadas e têm sua importância nesse aspecto de interoperabilidade (MILLER, 2000). Um exemplo de interoperabilidade política/humana pode ser visto na Figura 4 a seguir:

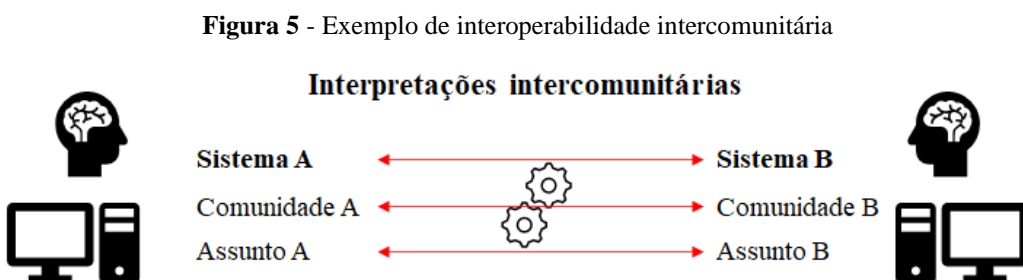


Fonte: Elaborada pela autora

2.1.4 Interoperabilidade intercomunitária

Esse aspecto percorre do acesso à informação interdisciplinar, às diversas informações provenientes de diferentes áreas e diferentes sistemas (SAYÃO; MARCONDES, 2008).

“À medida que as fronteiras tradicionais entre instituições e disciplinas começam a desaparecer, os pesquisadores exigem cada vez mais acesso a informações de uma ampla variedade de fontes, dentro e fora de sua própria área de assunto” (MILLER, 2000, não paginado, tradução nossa). A Figura 5 a seguir ilustra um exemplo de interoperabilidade intercomunitária:



Fonte: Elaborada pela autora

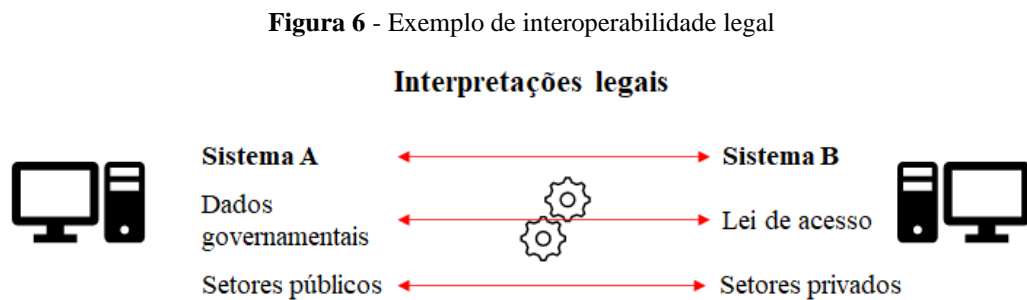
2.1.5 Interoperabilidade legal

A interoperabilidade legal considera as implicações legais de acesso à informação (SAYÃO; MARCONDES, 2008).

Dentro dos requisitos legais acerca do acesso à informação, as instituições tomam, de forma livre, a decisão de tornar ou não algum recurso de livre acesso, levando em consideração a disseminação da informação.

Implicações e impactos como estes podem ser vistos na criação de leis, como por exemplo a criação da Lei de Acesso à Informação, as quais carregam as implicações das divulgações ou não de certas informações.

Tem-se grande participação desse tipo de interoperabilidade em informações, recursos e dados governamentais e de setores públicos, uma vez que mostra interferências na capacidade de troca de dados governamentais e públicas (MILLER, 2000). Tem-se o exemplo de interoperabilidade legal na Figura 6:



Fonte: Elaborada pela autora

2.1.6 Interoperabilidade internacional

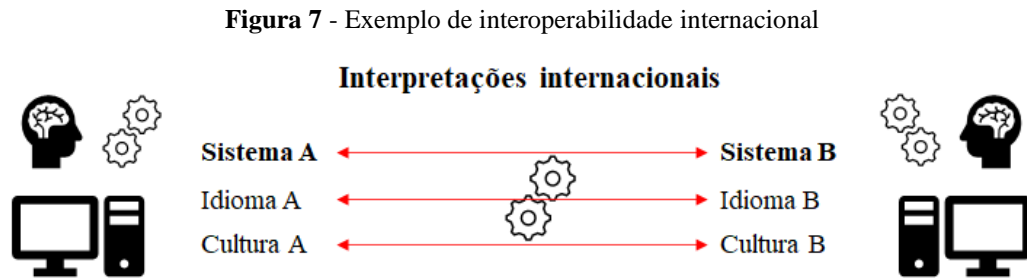
Quando consideradas em escala internacional, cada questão relacionada a cada aspecto de interoperabilidade acima torna-se ampliada, perpassando diferenças de abordagens e práticas de trabalho e organização. Além disso, questões como idioma e diferenças culturais tornam-se ainda mais significativas à medida em que os países se relacionam cada vez mais entre si (MILLER, 2000).

Nesse caso, atuando em escala internacional, deve-se considerar barreiras e dificuldades na comunicação, além da diversidade de normas e padrões (SAYÃO; MARCONDES, 2008).

Assim, entende-se que a interoperabilidade diz respeito à comunicação entre sistemas informacionais heterogêneos, ou seja, de *softwares* e *hardwares* diferentes, focando aqui em repositórios digitais, da forma mais completa e transparente possível (SAYÃO; MARCONDES, 2008).

Mais do que meramente comunicação entre os sistemas, no contexto das bibliotecas e dos repositórios digitais, o conceito de interoperabilidade perpassa pela forma de comunicação,

ficando mais complexo à medida que se percebe que esses ambientes são passíveis de integração e se faz necessário e confortável ao usuário uma única interface que traga resultados além dos esperados, perpassando por diversos sistemas diferentes (SAYÃO; MARCONDES, 2008). Um exemplo de interoperabilidade internacional é ilustrado pela Figura 7 a seguir:



2.2 Níveis de interoperabilidade

Além dos aspectos de interoperabilidade, Sayão e Marcondes (2008), trazem também os níveis de interoperabilidade que podem ser realizados entre os sistemas, são eles: a federação, o *harvesting* e a agregação. Também com base nos autores Sayão e Marcondes (2008), apresenta-se a seguir as principais características desses aspectos.

2.2.1 Federação

É considerado o nível mais alto de interoperabilidade e diz respeito à construção de federações de bibliotecas para interoperarem seus recursos, utilizando os mesmos protocolos e padrões para manterem seus sistemas integrados. O que impede o encontro de grandes federações é a dificuldade por parte das instituições de manterem seus sistemas sempre atualizados e padronizados (SAYÃO; MARCONDES, 2008).

Ainda para os autores, a federação constitui-se de várias bibliotecas independentes que se organizam em torno de um tema comum, formando uma rede de acesso através de uma única interface (SAYÃO; MARCONDES, 2008).

Ao se estabelecer uma federação, o grupo de organizações participante deve concordar que seus serviços deverão estar de acordo com determinadas especificações, geralmente especificadas a partir de padrões formais (ARMS et al., 2002).

Um exemplo de federação pode ser encontrado na *Alexandria Digital Research Library*, da Universidade da Califórnia em Santa Bárbara. Essa biblioteca é uma das parceiras do projeto Smete.org, que disponibiliza materiais educacionais de diversas bibliotecas (ARMS et al., 2002).

Além disso, as bibliotecas que compartilham seus registros de catálogos online através do protocolo Z39.50 podem ser um exemplo de federação, uma vez que todas necessitam estar de acordo para que o compartilhamento aconteça da melhor forma (ARMS et al., 2002).

Instituições em um nível de Interoperabilidade de federação podem promover a busca federada. Para Ferreira e Souto (2006), com base em Wadham (2004), a busca federada pode ser entendida como

[...] uma técnica pela qual as fontes de informação múltiplas podem ser pesquisadas ao mesmo tempo incluindo catálogos de bibliotecas, bases de dados comerciais, mecanismos de busca Web e outros metadados ou bases de dados digitais, sendo que o mecanismo de busca federada, usando um protocolo de busca tal como o Z39.50 por exemplo, recolhe a informação a partir de várias fontes e fornece um resultado de busca integrado ao usuário. (WADHAM, 2004¹ apud FERREIRA; SOUTO, 2006, p. 27-28)

Como já mencionado, em se tratando de federação, o desafio principal se encontra na formação desta, uma vez que todos os sistemas devem estar de acordo e manter-se atualizados. Como o custo pela participação em uma federação é relativamente alto, as federações normalmente têm pequenas associações, porém dedicadas (ARMS et al., 2002).

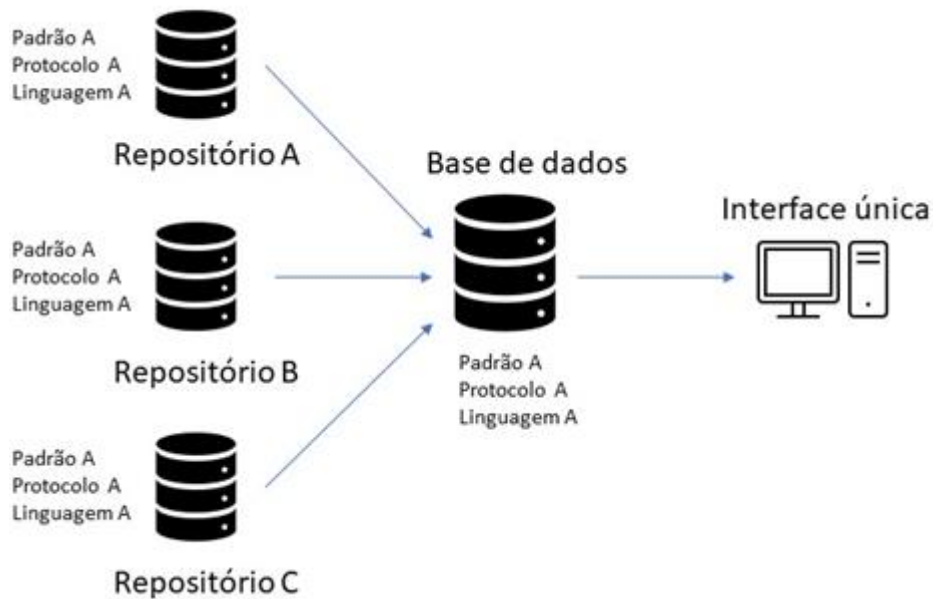
Em se tratando da própria operacionalização, segundo Ferreira e Souto (2006), a federação é mais adequada quando existe um pequeno número de integrantes, porém com coleções numerosas e variadas, uma vez que ao contrário disso, a busca seria praticamente impossível em ambientes maiores e distribuídos, como as buscas online, por exemplo.

Nessa mesma linha, para Martins e Ferreira (2012, p. 434), “A busca em diferentes servidores é recomendada em situações em que há poucos integrantes e com grandes coleções de dados - do contrário, problemas de escalabilidade poderiam ocorrer”.

A Busca Federada embora vá ao encontro da integração de informações no esforço de tornar e alcançar interoperabilidade, ela ainda é limitada quando se trata do número de participantes da federação (PECEGUEIRO, 2019).

¹ WADHAM, Rachel L. Federated searching. **Library Mosaics**, Culver City, CA, v. 15, n. 1, p. 20, Jan./Feb. 2004.

Figura 8 - Exemplo de federação



Fonte: Elaborado pela autora

Na Figura 8 tem-se um exemplo de federação onde todos os repositórios, bases de dados e sistemas em geral das instituições envolvidas possuem um mesmo padrão, protocolo e linguagem e seguem as mesmas regras de atualizações de seus dados.

2.2.2 *Harvesting*

Um primeiro sinal sobre a coleta de metadados foi desenvolvida em meados dos anos 90 por um projeto denominado *harvesting*. A partir de 1998, o conceito foi utilizado em um protótipo denominado *Universal Preprint Server*, que foi concluído como uma estratégia para a criação de sistemas federados de sistemas de pré-impressão heterogêneos (ARMS et al., 2002).

Atualmente, o conceito de *harvesting* é amplamente utilizado pelo protocolo OAI-PMH e conseqüentemente o mais utilizado na interoperabilidade entre os sistemas. Nesse conceito, que seria o conceito de “coleta”, os sistemas utilizam-se do protocolo para exporem e coletarem seus metadados de forma assíncrona, ou seja, a atualização ou não de um sistema não interfere em outro (SAYÃO; MARCONDES, 2008).

Os esforços para manter as federações faz com que a proposta do *harvesting* seja mais atrativa, uma vez que a demanda de esforços se torna menor.

Criado juntamente com o protocolo OAI-PMH, o esquema de *harvesting* consiste em coletar dos sistemas informacionais, tanto provedores de dados quanto provedores de serviços. Nessa perspectiva, os metadados dos diferentes provedores de dados ficam visíveis de forma padronizada para que sejam coletados de forma periódica e automática pelos provedores de serviços, que disponibilizam as informações coletadas dos provedores de dados.

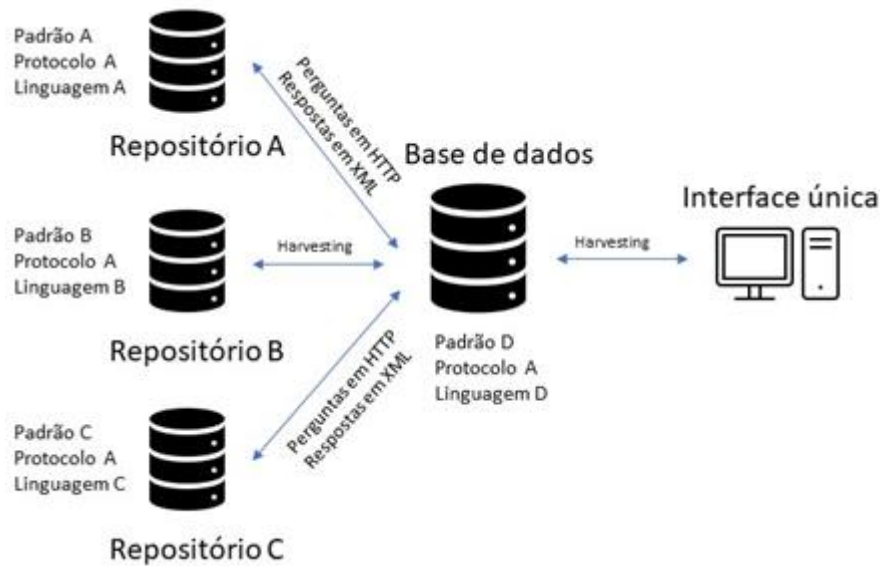
Cada biblioteca digital disponibiliza metadados sobre suas coleções em um formato simples de troca. Esses metadados podem ser coletados por provedores de serviços e incorporados a serviços como descoberta de informações ou vinculação de referência. Dois membros da equipe do Site for Science atuam como executivos da OAI, e o Site for Science foi um dos locais de teste alfa do protocolo de coleta de metadados. (ARMS et al., 2002, não paginado, tradução nossa)

Esse esquema de comunicação traz vantagens como a inclusão de novos provedores de dados, desde que estejam de acordo com o padrão utilizado e uma melhoria no desempenho da consulta aos materiais, uma vez que esta é realizada na base de metadados dos provedores de serviços (MARCONDES; SAYÃO, 2001).

Embora os serviços criados pela coleta de metadados sejam geralmente menos poderosos do que os fornecidos pelas federações, o ônus de participar é muito menor. Como resultado, é provável que muitas outras organizações participem e mantenham seus sistemas atualizados. Isso é confirmado pela rápida aceitação do protocolo de coleta de metadados da OAI. (ARMS et al., 2002, não paginado, tradução nossa)

Por mais que esse tipo de nível de interoperabilidade seja o mais utilizado e exija menos esforços das instituições, elas ainda precisam manter seus dados atualizados e zelar pela sincronização dos metadados mantidos pelos provedores de dados e os metadados coletados pelos provedores de serviços (LIU et al., 2001).

Figura 9 - Exemplo de *harvesting*



Fonte: Elaborado pela autora

Na Figura 9 é ilustrado um exemplo de *harvesting*, na qual, os repositórios A, B e C que são provedores de dados, expõem seus metadados com o OAI-PMH cada um a seu modo, com padrões, protocolos e linguagens diferentes e a base de dados de um outro sistema provedor de serviços coleta esses metadados também através do OAI-PMH. Esses metadados coletados serão armazenados no banco de dados e são esses que serão consultados no momento da pesquisa.

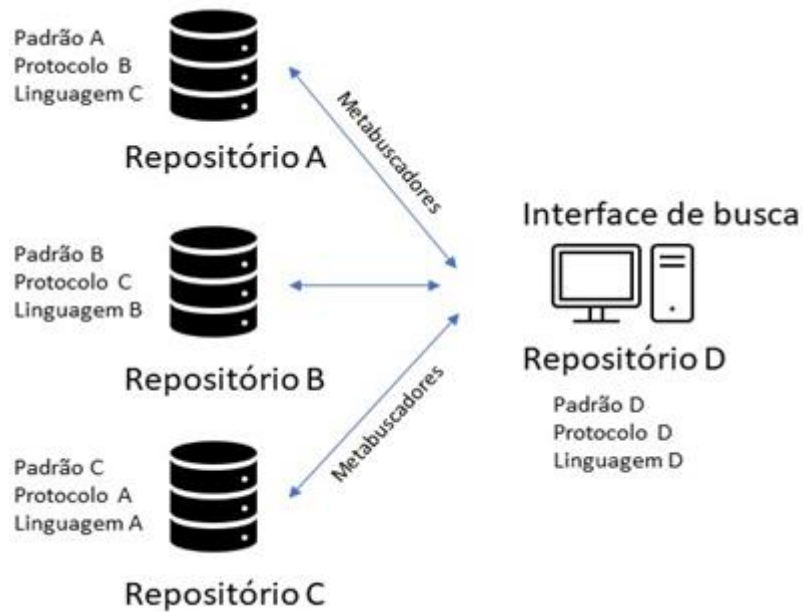
2.2.3 Agregação

Um dos níveis de interoperabilidade. Ainda que os sistemas não tenham padrões ou protocolos iguais, um mínimo sobre eles se pode interoperar através de metabuscadores, estando assim, no nível de agregação (SAYÃO; MARCONDES, 2008).

Nesse nível, o uso dos protocolos utilizados para facilitar a exposição e a coleta de metadados compatíveis são utilizados também fora de seu ambiente local, geralmente caracterizado por uma interface em comum e vários acervos parceiros (SANTAREM SEGUNDO; SILVA; MARTINS, 2019).

Arms et al. (2002, não paginado) entendem que na agregação, “[...] mesmo que as várias organizações não cooperem de maneira formal, ainda é possível um nível básico de interoperabilidade, reunindo informações de acesso aberto usando um rastreador da Web”.

Figura 10 - Exemplo de agregação



Fonte: Elaborado pela autora

Tem-se na Figura 10 um exemplo de agregação, na qual, ao ser realizada uma pesquisa pela interface de busca do Repositório D, que possui o padrão de metadados D, o protocolo D e a linguagem D, por mais que nada dele se comunique diretamente com os repositórios A, B e C que possuem padrões, protocolos e linguagem totalmente diferentes do seu e entre si, seus metabuscadores conseguem resgatar um mínimo possível de cada descrição dentro desses repositórios.

3 SERVIÇOS DE METADADOS

A organização dos materiais informacionais está intrinsecamente ligada à área da Biblioteconomia, Ciência da Informação e áreas afins desde seus primórdios, alterando-se com o tempo para acompanhar as mudanças da sociedade.

Para uma organização eficaz, que permita uma fácil localização da informação, as bibliotecas utilizam-se dos metadados para a realização desta tarefa que parece simples à primeira vista (ZENG; QIN, 2008).

Anterior à era da internet, a organização da informação presente em instituições como bibliotecas, arquivos e museus seguiam regras apresentadas em instrumentos de representação de cunho manual. No caso das bibliotecas, instrumentos como o *Anglo-American Cataloguing Rules*, segunda edição (AACR2r) e o padrão *Machine-Readable Cataloging* (MARC21), guiados semanticamente pelos códigos de classificação, como a Classificação Decimal de Dewey (CDD), para organizar e tornar “encontráveis” os materiais, predominantemente físicos (ZENG; QIN, 2008).

O termo “*metadata*” (metadado) foi cunhado por Jack E. Myers no ano de 1969, porém, popularizou-se apenas na década de 1990, com a era da internet (PEREIRA; RIBEIRO JÚNIOR; NEVES, 2005; ZENG; QIN, 2008).

De modo simples e geral, o termo “metadado”, usado no âmbito digital primeiramente pela Ciência da Computação para especificar informações sobre objetos em uma base de dados, carrega a definição “dados sobre dados” ou “informação sobre informação” (ZENG; QIN, 2008).

A definição “dados sobre dados”, embora não incorreta, recebeu refinamentos, principalmente pela Ciência da Informação, ficando com definições como: “[...] metadados são estruturas informacionais que descrevem, explicam, localizam, ou, em outras palavras, tornam mais fácil a recuperação, uso ou o gerenciamento do recurso informacional” (ZENG; QIN, 2008, p. 7, tradução nossa).

Para Alves (2010, p. 47), os metadados são

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Embora peças-chave para a descrição de recursos, os metadados simples não retornam uma recuperação eficaz dos recursos. Para tanto, se faz necessária a estruturação desses metadados, em estruturas capazes de garantir a recuperação eficaz dos recursos informacionais, além de permitir a interoperabilidade entre os sistemas (ARAKAKI, 2016).

Essas estruturas que organizam os metadados de modo a padronizá-los e permitir qualquer comunicação entre diferentes sistemas, são chamados padrões de metadados.

Zeng e Qin (2008, p. 15, tradução nossa) definem padrões de metadados como estruturas

[...] de descrição constituídas por um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados. O objetivo do padrão de metadados é descrever uma entidade gerando uma representação unívoca e padronizada que possa ser utilizada para recuperação da mesma.

Para as autoras, ainda que se faça o uso dos padrões de metadados, se um recurso não estiver descrito de forma adequada, o usuário não conseguirá fazer a sua recuperação (ZENG; QIN, 2008).

O desenvolvimento dos metadados vinculados às Tecnologias de Informação e Comunicação (TIC) foi amplamente difundida na década de 90, pois, a internet rapidamente se tornava uma imensa base de informações.

Da década de 1990 até os dias atuais, a crescente utilização das tecnologias, em especial as TIC, em meio à novas variedades de recursos nesse novo ambiente informacional (ambiente digital), fez surgir novas formas de representação, com o objetivo de melhorar a capacidade de sistemas de recuperar a informação da melhor forma (ALVES, 2010).

Essas novas formas de representação ganharam enfoque e foram chamadas de “metadados”, embora as formas anteriores de representação também possam ser entendidas com metadados, pois meta significa que algo se aplica a ele mesmo, ou seja, metadados são dados representacionais que podem representar ou até substituir a própria informação (MÉNDEZ RODRÍGUEZ, 2002).

São diversos os tipos, funções e características dos metadados. De forma geral, eles podem ser gerados automaticamente ou manualmente; podem ser internos ou externos em relação ao objeto; ser gerais ou específicos; estáticos ou dinâmicos; podem estar ligados à preservação ou não, assim como estruturados em um padrão ou não (CAMPOS, 2007).

Entende-se que em um ambiente tão extenso como o ambiente *web*, os metadados são de suma importância, pois, são esses que vão descrever cada recurso, tornando-o recuperável de forma inequívoca (ALVES, 2010).

Assim, o uso dos metadados se dá para a descrição dos recursos informacionais, o que minimiza os problemas de localização e recuperação dentro dos ambientes *web* (ARAKAKI, 2016).

O termo “*metadata service*” (serviço de metadado) é utilizado para descrever os sistemas e as ferramentas que suportam a criação e a manutenção de esquemas de metadados (ZENG; QIN, 2008).

Nesse caso, diz respeito diretamente à criação dos metadados, ou seja, desde uma simples fotografia com metadados automáticos como data e hora, até documentos tratados e estruturados em ambientes informacionais. O que acontece aqui, dentro dos serviços de metadados, é uma preocupação para além da criação e utilização dos metadados, é preciso saber também identificar quais metadados e/ou padrões utilizar para cada situação (ZENG; QIN, 2008).

De modo geral, tem-se a seguinte classificação de metadados:

1. Funções a desempenhar:

- **Administrativos:** São os metadados utilizados na administração da informação, fornecendo informações como formas de acesso, tipos de arquivos, informações sobre registros legais etc.
- **Descritivos:** metadados que são utilizados para descrever e identificar recursos informacionais. Fornecem informações ligadas à catalogação como título, autor, *hiperlinks*, resumo etc.
- **Preservação:** utilizados com relação às informações sobre conservação e preservação dos recursos, trazendo informações sobre as ações tomadas para a conservação desses recursos.
- **Técnicos:** trazendo informações sobre *hardware* e *software*, esses metadados portam informações a respeito do funcionamento dos sistemas.
- **Uso:** são os metadados que informam o nível e tipo de uso da informação, com dados sobre os registros de exibição, controles de acesso, logs de acesso etc.

2. Origem:

- **Metadados internos:** gerados no momento em que o recurso é criado (como o nome do arquivo).

- **Metadados intrínsecos:** “[...] internos ao item ou obra (exemplo: título e subtítulo na folha de rosto de um livro impresso)” (ALVES; SANTOS, 2013, p. 45).
- **Metadados externos:** gerados após a criação do recurso.

3. Métodos de criação dos metadados:

- **Metadados automáticos:** gerados automaticamente por máquinas.
- **Metadados manuais:** gerados por um indivíduo.

4. Natureza dos metadados:

- **Metadados não especializados:** criados por não especialistas da área de informação.
- **Metadados especializados:** criados por especialistas da área de informação.

5. Status:

- **Metadados estáticos:** são os metadados que não se alteram após sua criação.
- **Metadados dinâmicos:** são aqueles que se modificam conforme a manipulação dos documentos.
- **Metadados de longa duração:** são usados para assegurar o acesso aos recursos.
- **Metadados de curta duração:** são de natureza temporária.

6. Estrutura:

- **Metadados estruturados:** são estruturados conforme um padrão previsível.
- **Metadados não estruturados:** não possuem uma estrutura previsível e não são compatíveis com qualquer estrutura.

7. Semântica:

- **Metadados controlados:** estão estruturados com algum tipo de padrão de conteúdo.
- **Metadados não controlados:** são aqueles que não estão estruturados com algum padrão de conteúdo.

8. Nível:

- **Metadados nível de coleção:** que indicam coleções de itens originais ou objetos.
- **Metadados nível de item:** indicam itens individuais geralmente pertencentes a coleções (ALVES; SANTOS, 2013).

Pode-se entender que essa classificação é geral, uma vez que dentro de cada categoria ainda existem outros metadados para descrições mais específicas (ALVES, 2010).

A utilização dos metadados na descrição dos recursos informacionais, principalmente quando estruturados em algum padrão, como o *Dublin Core*, torna muito mais fácil a recuperação da informação e a troca informacional entre os diferentes sistemas (PEREIRA; RIBEIRO JÚNIOR; NEVES, 2005).

Ao garantir a padronização das descrições, a utilização dos padrões de metadados por parte dos sistemas garante uma interoperabilidade muito mais satisfatória entre eles, já que não há a necessidade de “traduzir” os metadados utilizados de diferentes formas. Existe a possibilidade de sistemas utilizarem padrões diferentes, porém, ao realizar um acordo, essa chance se torna mínima ou inexistente. Ainda assim, ao ocorrer a interoperabilidade entre sistemas que utilizam padrões diferentes, as perdas dos dados são menores (LOURENÇO, 2005).

O padrão mais citado nessa pesquisa será o padrão *Dublin Core*, pois, além de ser o padrão utilizado e recomendado pela OAI, é um padrão geral e extensível, conseguindo abranger sistemas com descrições completamente diferentes e ainda assim, tornar possível a comunicação entre eles.

De modo geral, o padrão *Dublin Core* foi criado de acordo com um consenso entre os membros da OCLC e da NCSA, que definiu elementos funcionais para tornar o padrão mais acessível, extensível e que abordasse sistemas de modo geral (GRÁCIO, 2002).

O padrão *Dublin Core*, que pode ser definido como um “[...] conjunto de elementos de metadados planejado para facilitar a descrição de recursos eletrônicos [...]” (SOUZA; VENDRUSCULO; MELO, 2000, p. 93), possui 15 elementos principais, sendo esses: Título, Criador, Assunto, Descrição, Publicador, Colaborador, Data, Tipo, Formato, Identificador, Fonte, Idioma, Relação, Cobertura e Direitos autorais.

Esses elementos são considerados principais na descrição de um recurso informacional, porém, o padrão é flexível, tornando possível sua adaptação conforme as necessidades do sistema (SOUZA; VENDRUSCULO; MELO, 2000).

3.1 Provedores de dados e provedores de serviços

Um repositório, de modo geral, pode ser entendido como um sistema ou plataforma de armazenamento de dados e recursos informacionais. Assim, os provedores de dados são os repositórios que armazenam objetos digitais e os metadados que os descrevem individualmente. Já um provedor de serviço refere-se aos repositórios que só armazenam metadados dos objetos. Esses objetos continuam pertencendo aos repositórios originais.

Em um estudo realizado por Café e Fachin (2007), existiam apenas oito repositórios de metadados, ou seja, provedores de dados que podiam ser considerados relevantes para a área da Ciência da Informação, sendo eles: *Librarian's Digital Library* (LDL), *Digital Library of Information Science and Technology* (dLIST), *Caltech Library System Papers and Publications*, *Mémoires de 3^o cycle en Science de l'Information et de la Communication* (MémSIC), *Archive Ouverte em Sciences de l'Information et de la Communication* (@archivesSIC), *E-prints in Library and Information Science* (E-LIS), *CNR Bologna Research Library Eprint Service*, Diálogo Científico (DICI).

Esses exemplos de provedores de dados têm a infraestrutura e a interface de um repositório de dados (institucional ou temático) comum, porém, em se tratando de forma técnica, seus metadados ficam expostos para a fácil coleta pelos provedores de serviços, o que não necessariamente ocorre em repositórios de dados. Entretanto, nada impede que um repositório de dados seja também repositório de metadados, consequentemente provedor de dados e provedor de serviços (CAFÉ; FACHIN, 2007).

Baseados na filosofia de Arquivos Abertos, existe uma lista grande de sistemas cadastrados como provedores de serviços na página da OAI. Como provedores de serviços, os repositórios se prestam à coleta automática dos metadados (CAFÉ; FACHIN, 2007).

Como provedores de serviços especificamente da área da Ciência da Informação, as autoras elencaram os seguintes repositórios como os mais relevantes: ColLib, OASIC, DL-Harvester e METALIS (CAFÉ; FACHIN, 2007).

A Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) desenvolvida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) em parceria com as instituições brasileiras de ensino e pesquisa é um dos primeiros exemplos brasileiros de provedor de serviços, porém, não há um registro oficial da BDTD na página da OAI.

3.1.1 OAI-PMH

A falta de padronização dos recursos informacionais digitais fez com que, em meados dos anos 90, surgisse a *Open Archives Initiative* (OAI), um movimento iniciado por pesquisadores a fim de buscar soluções para integração de publicações online e para promover o livre acesso às publicações científicas de todo o mundo (FRANÇA; RAMALHO, 2013).

Atualmente, é possível encontrar a seguinte missão da OAI:

A Open Archives Initiative desenvolve e promove padrões de interoperabilidade que visam facilitar a disseminação eficiente do conteúdo. A OAI tem suas raízes em um esforço para aprimorar o acesso aos arquivos de impressão eletrônica como um meio de aumentar a disponibilidade da comunicação acadêmica. (OPEN ARCHIVES, [201-?a], não paginado, tradução nossa)

Em 1999 foi realizada uma convenção em Santa Fé, no Novo México, abordando questões sobre interoperabilidade, repositórios digitais, e-prints e busca integrada. Desses assuntos, as atenções voltaram-se para o *Open Access* (Acesso Aberto), mais evidenciada em bibliotecas, arquivos e museus.

Assim, a convenção de Santa Fé deixou especificações técnicas para interoperabilidade, publicadas em 2001, que deu origem ao *Protocol for Metadata Harvesting*. Em 2002 foi publicada a sua versão 2.0, que vem sendo utilizada e melhorada até o presente momento.

Como um dos protocolos que popularizou o formato de provedores de dados e provedores de serviços, o protocolo OAI-PMH, como ficou conhecido, funciona conforme descrito a seguir.

Ao basear-se no esquema de provedores de dados e provedores de serviços, os participantes que utilizam o protocolo podem realizar um cadastro definindo se o sistema é um provedor de dados, um provedor de serviços ou ambos. Para tanto, os provedores de dados são repositórios de documentos digitais que expõe os metadados de seus documentos, enquanto os provedores de serviços realizam a busca pelos metadados expostos pelos provedores de dados (GARCIA; SUNYE, 2003). Cabe ressaltar que o cadastramento é recomendado, porém voluntário e não obrigatório, assim, a OAI não possui o número exato de sistemas que utilizam o protocolo OAI-PMH (OPEN ARCHIVES, [201-?b]).

Com foco no compartilhamento e transporte dos metadados, para Martins e Ferreira (2012), o OAI-PMH utiliza de três camadas básicas de padrões tecnológicos:

- ✓ A primeira camada é a camada de semântica da descrição dos dados, a qual dá preferência para a utilização do padrão *Dublin Core*, porém não impede a utilização de outro padrão para a saída dos dados;
- ✓ A segunda camada é a de linguagem de descrição dos dados, sendo escolhida a linguagem de marcação XML para descrever os dados. A XML foi escolhida para tentar padronizar a forma como os metadados seriam descritos, além de ser uma recomendação oficial da *World Wide Web Consortium* (W3C);
- ✓ A terceira camada diz respeito ao transporte dos dados entre os provedores de dados e provedores de serviços. Para isso, foi escolhido o padrão *HyperText Transfer*

Protocol (HTTP), uma vez que é baseado na arquitetura cliente-servidor, servindo de base para a *web* (MARTINS; FERREIRA, 2012).

Ao utilizar o protocolo HTTP e a linguagem de marcação XML, o OAI-PMH conta com seis verbos de solicitação, sendo esses:

- Verbo *Identify*: é o verbo utilizado para a recuperação de dados e informações sobre o provedor de dados;
- Verbo *ListMetadataFormat*: verbo que recupera informações acerca dos formatos de metadados disponíveis no repositório;
- Verbo *GetRecord*: o qual recupera apenas um registro no repositório e deve-se especificar o formato de metadados, sendo o padrão *oai_dc*;
- Verbo *ListRecords*: é o verbo que identifica os metadados do repositório. Pode-se estabelecer uma coleta seletiva (*data-based*) ou uma coleta em conjunto (*set-based*);
- Verbo *ListIdentifiers*: uma abreviação do verbo *ListRecords*, recupera informações dos cabeçalhos dos registros. Também pode-se utilizar as coletas seletiva ou em conjunto;
- Verbo *ListSet*: verbo que recolhe uma lista da estrutura de todo o conjunto de um repositório, geralmente utilizada para *harvesting* seletivo (GARCIA; SUNYE, 2003).

Assim, observa-se que o OAI-PMH funciona através de “perguntas e respostas” conduzidas pelos seis verbos apresentados, pelas linguagens HTTP e XML, onde, um provedor de serviços faz sua solicitação, uma “pergunta” através do protocolo HTTP utilizando um dos seis verbos e um provedor de dados “responde” em linguagem XML com toda a estrutura de descrição solicitada em HTTP (GARCIA; SUNYE, 2003).

3.2 Resource Description Framework (RDF)

Como um produto da W3C, o RDF é uma estrutura principalmente aplicada em XML, permitindo a codificação, troca e reutilização de dados estruturados.

Anteriormente ao RDF, foi desenvolvido pela *Internet Engineering Task Force* (IETF) um mecanismo para exibir classificações de páginas da *web* dos servidores para seus clientes, conhecido como *Platform for Internet Content Selection* (PICS) (NEEDLEMAN, 2001).

Criado para ser um mecanismo geral de criação de classificação, as classificações, ou etiquetas trazidas pelo PICS eram definidas pelas próprias comunidades que desejavam utilizá-

lo. Nesse caso, diferentes organizações podiam avaliar seus conteúdos com base em seus próprios objetivos, ainda que tal mecanismo pretendesse seguir uma classificação generalizada.

Apesar disso, uma das principais motivações para o desenvolvimento do PICS foi para desenvolver um autorregulador de sistemas na internet, para tentar minimizar os esforços governamentais para impor limites de conteúdo (NEEDLEMAN, 2001).

Porém, algumas limitações no mecanismo PICS foram identificadas, fazendo com que se desenvolvesse um funcional mais geral para tentar resolver a questão da associação de metadados descritivos com os recursos, além da definição de que o PICS poderia ser utilizado em diferentes domínios em aplicativos adicionais.

Foi a partir dessas constatações que a W3C desenvolveu um grupo de trabalho RDF (NEEDLEMAN, 2001).

RDF é o resultado de várias comunidades de metadados se unindo e contribuindo com suas necessidades e requisitos para fornecer uma arquitetura flexível e robusta para suportar metadados na Web em uma ampla variedade de formulários. Embora fortemente influenciado pelo PICS, e ser uma estrutura geral de metadados e conhecimento mecanismo de representação, RDF não é realmente o produto de qualquer influência específica, veio através das contribuições de muitos esforços de metadados que vinham acontecendo, incluindo o trabalho de várias empresas membros do W3C. (NEEDLEMAN, 2001, p. 59, tradução nossa)

No contexto da Web Semântica, a semântica (o significado) dos metadados é definida por cada comunidade de descrição de recursos, cada uma buscando atender às suas necessidades específicas. A sintaxe, ou seja, a linguagem computacional em que os metadados são expressos e sua a disposição sistemática facilita a troca, seu uso e reuso entre vários sistemas. Já a estrutura pode ser entendida como uma restrição formal sobre a sintaxe para uma representação consistente da semântica (MILLER, 1998; FERREIRA; SANTOS, 2013).

Diferentemente da HTML e da XML, a principal intenção não é exibir os recursos corretamente, mas permitir o posterior processamento e a recombinação das informações que estão contidas nesses recursos, ou seja, os dados que compõem as suas representações. Dessa forma, o RDF é consequentemente e frequentemente visto como o modelo básico de representação para o desenvolvimento da Web Semântica (HITZLER; KRÖTZSCH; RUDOLPH, 2009², p. 19 apud FERREIRA; SANTOS, 2013, p. 15).

² HITZLER, Pascal; KRÖTZSCH, Markus; RUDOLPH, Sebastian. **Foundations of Semantic Web technologies**. Boca Raton, FL: CRC Press, 2010.

Assim, o RDF fornece um modelo para descrever os recursos, uma vez que cada um desses possui suas propriedades (NEEDLEMAN, 2001).

Ao utilizar o modelo de triplas para realizar as representações, o RDF faz o uso de “sujeito-predicado-objeto” (WITT, 2010), que pode também ser representados por recurso-propriedade-valor, estritamente nesta ordem, representados em grafos (FERREIRA; SANTOS, 2013).

No modelo RDF, um “recurso” pode ser qualquer objeto que possa ser identificado exclusivamente por um *Uniform Resource Identifier* (URI). As “propriedades” associadas aos recursos são identificadas pelos tipos de propriedades (como por exemplo título, autor, criador) e cada uma tem seu “valor” correspondente, nesse caso, o tipo de propriedade expressa o relacionamento de cada valor associado ao recurso (NEEDLEMAN, 2001).

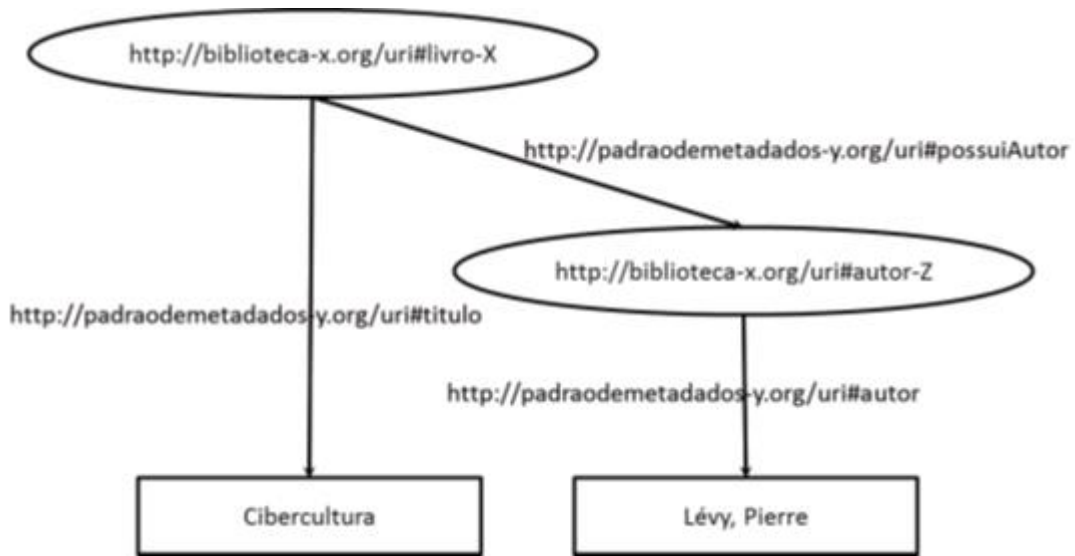
Quanto aos valores, esses podem ser de natureza atômica, como uma sequência de números, letras ou caracteres, ou podem ser outros recursos que podem ter suas próprias propriedades. À essa coleção de propriedades que se referem a um recurso é dado o nome de descrição (NEEDLEMAN, 2001).

Independente de um ambiente de sintaxe, um documento RDF é estruturado através de grafos (grafos estruturados, orientados, grafos dirigidos, dígrafos ou *quiver*), que significa um conjunto de nós (ou vértices) ligados pelas arestas direcionadas (ou setas, flechas), sendo tanto os nós quanto as arestas rotuladas com identificadores para distingui-los (FERREIRA; SANTOS, 2013).

Nesse caso, cada grafo expressa uma tripla (recurso, propriedade, valor), não necessariamente seguindo uma forma hierárquica, mas uma forma de ligação entre os grafos.

Assim, de forma a exemplificar, tem-se na Figura 17 um grafo simples que descreve o relacionamento entre um livro e seu autor. Nesse grafo, o livro X possui o seu autor Z, tendo então, o livro como “recurso”, sua “propriedade” autor e seu “valor” sendo o nome do autor do livro. Vale lembrar que cada um deles, o recurso, a propriedade (que também pode ser um recurso) e o relacionamento entre eles devem ser identificados através de uma URI (FERREIRA; SANTOS, 2013).

Figura 11 - Grafo RDF com recurso-propriedade-valor



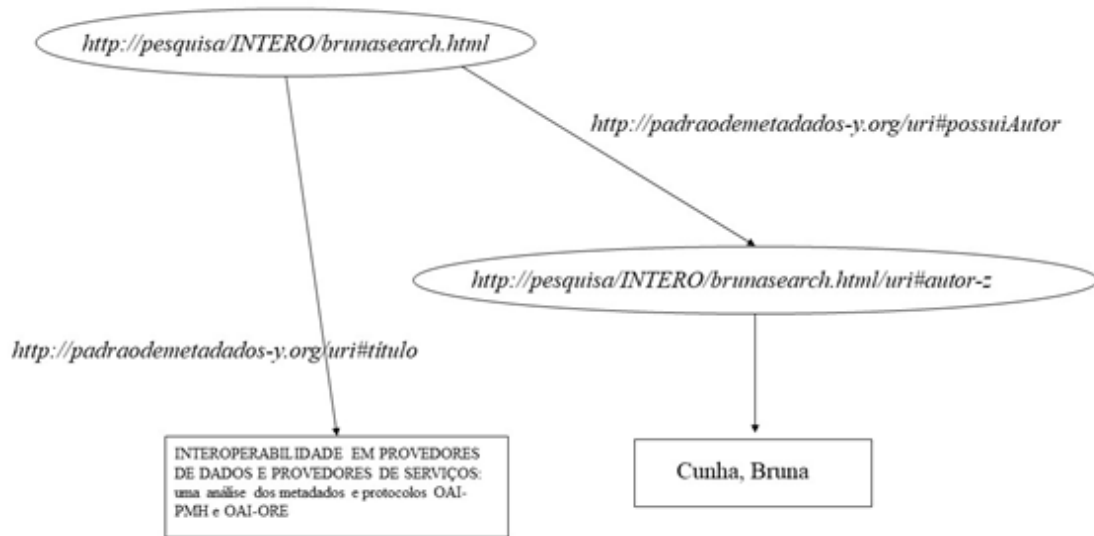
Fonte: Ferreira e Santos (2013, p. 19)

Embora livre de sintaxe e construído para não carregar um modelo hierárquico (ou modelo de árvore), o RDF é representado por meio da XML, de modo a codificar o modelo, assim, pode-se assumir o seguinte exemplo adaptado de Needleman (2001) para esta pesquisa.

Assume-se que temos um documento que representa essa pesquisa, representado pelo URI (fictício) *http://pesquisa/INTERO/brunasearch.html*.

Esse recurso possui suas propriedades, entre elas a propriedade “Autor”. Essa propriedade, nesse caso, possui o valor “Cunha, Bruna”, podendo essa propriedade ser repetida pela quantidade total de autor, ou seja, pode-se ter uma ou mais ocorrências dessa propriedade, assim como uma ou mais ocorrências de valor. Outra propriedade é a propriedade “Título”, que possui como valor “INTEROPERABILIDADE EM PROVEDORES DE DADOS E PROVEDORES DE SERVIÇOS: uma análise dos metadados e protocolos OAI-PMH e OAI-ORE”, podendo ter quaisquer outras propriedades como idioma, número de páginas, editor (se for o caso), entre outros, podendo até mesmo essas propriedades serem criadas especificamente para uma certa área, como por exemplo, a propriedade “Cor”, para algum uso específico. Needleman (2001, p. 59, tradução nossa) afirma que “Eu mesmo poderia definir esses recursos, do meu vocabulário particular ou optar por usar definições (e, portanto, a semântica inerente a eles) criada por outras comunidades de metadados”.

Figura 12 - Exemplo de grafo RDF aplicado a esta pesquisa



Fonte: Elaborado pela autora

A figura 18 mostra como é um grafo simples em RDF desta pesquisa. Temos a primeira pesquisa iniciada por um usuário (máquina ou não máquina) pelo prenome da autora. A autora está relacionada a esta pesquisa e também tem uma relação com uma página com seus dados, no caso apresentado, seu sobrenome, prenome.

Porém, no exemplo citado, as propriedades definidas foram preenchidas com valores atômicos (valores simples), que são os caracteres. Mas, o valor “Autor”, por exemplo, no lugar de uma sequência simples de caracteres, poderia ser preenchido com outro recurso representado por uma URI, como (também fictício) `http://me/bib/camposbruna`, direcionando para outro recurso com descrições apenas do autor, como nome, endereço, e-mail, telefone e até a comida favorita e também poderia ter vários níveis (NEEDLEMAN, 2001).

Não há limite para o número de níveis que podem ser usados, nem existe uma maneira certa ou errada de fazê-lo. A estrutura e níveis de codificação dependerão do que for sendo descritos e os objetivos da organização ou indivíduo que fornece a descrição. (NEEDLEMAN, 2001, p. 59, tradução nossa)

Assim, pode-se ter descrições mais simples e descrições mais elaboradas, as quais dependem inteiramente de quem está realizando essa descrição e de suas necessidades.

Como já mencionado, o RDF não necessita de uma sintaxe definida, porém, a forma mais comum de representá-lo se dá através da utilização da linguagem XML, uma vez que assim estruturados, os metadados ficam ordenados e tornam possível o intercâmbio.

Essa linguagem, que também é um produto da W3C, é bastante flexível, principalmente na questão do idioma, o que permite a definição do conjunto de *tags* à necessidade da comunidade, além de permitir uma estrutura formal para permitir uma representação semanticamente consistente (NEEDLEMAN, 2001).

Uma das vantagens do XML é tornar possível o uso de *namespaces*, que são uma maneira de associar uma autoridade à uma *tag*, de maneira que esta não fique ambígua.

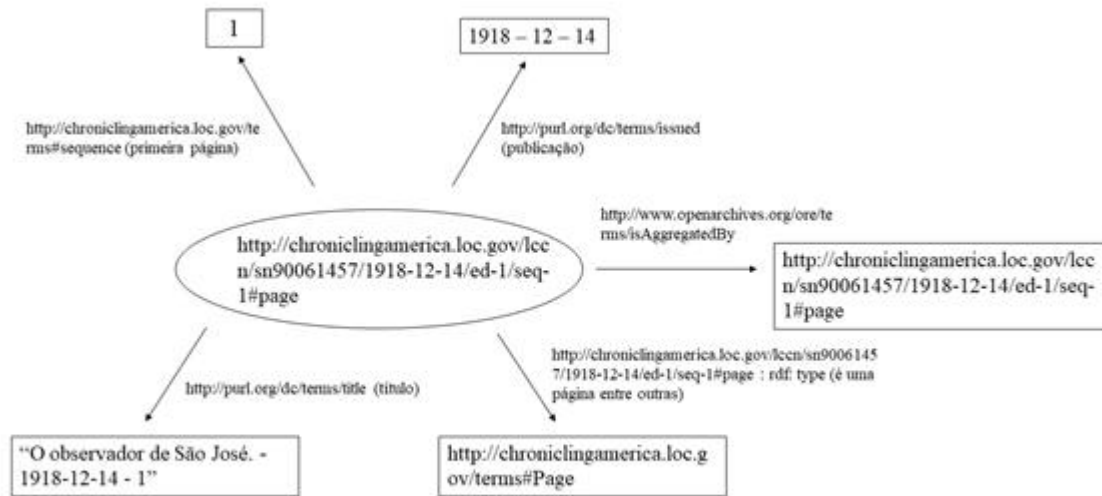
Usando XML Namespaces, uma capacidade é acessível para associar um Autoridade com propriedade e distinção entre propriedades que podem ter sintaxe semelhante, mas semântica bastante diferente (NEEDLEMAN, 2001, p. 60, tradução nossa).

Aqui, o modelo faz uso do padrão *Dublin Core*, uma vez que este define um conjunto de elementos de metadados principais para descrever um recurso.

Ao lidar com recursos da *web* usando RDF, o sujeito é indicado por um URI. O predicado também é indicado por um URI. O objeto pode ser indicado por um URI ou por um literal (uma sequência de texto). Se a página de um jornal, por exemplo, foi digitalizada e disponível na Internet, nossas declarações podem ser:

1. *O documento foi emitido em 14 de dezembro de 1918.*
 Sujeito: <http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page>
 Predicado: <http://purl.org/dc/terms/issued>
 Objeto: 1918- 12-14
2. *O documento é intitulado "O observador de São José. - 1918-12-14 - 1"*
 Sujeito: <http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page>
 Predicado: <http://purl.org/dc/terms/title>
 Objeto: "O observador de São José. - 1918-12-14 - 1"
3. *O documento é pedido primeiro.*
 Sujeito: <http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page>
 Predicado: <http://chroniclingamerica.loc.gov/terms#sequence> Objeto: 1
4. *O documento pertence a algo específico.*
 Sujeito: <http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page>
 Predicado: <http://www.openarchives.org/ore/terms/isAggregatedBy>
 Objeto: <http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1#issue>
5. *O documento é uma página de jornal.*
 Sujeito: <http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page> : rdf: type
 Objeto: <http://chroniclingamerica.loc.gov/terms#Page>
 (WITT, 2010, p. 12-13, tradução nossa).

Figura 13 - Página de jornal descrita em grafos RDF



Fonte: Elaborada pela autora

A Figura 19 representa todo o esquema de descrição RDF acima citado, indicando para o usuário não-máquina, onde está cada elemento do grafo representado em XML para a máquina.

Em outras palavras, algo (o sujeito ou recurso) é descrito ou relacionado a (predicado ou propriedade) outra coisa (um objeto ou valor).

4 O LINKED DATA E O OAI-ORE

Em 2006, Tim Berners-Lee identificou o termo *Linked Data* como sendo uma abordagem de ligação de dados. Diferente da primeira versão *web*, a Web Semântica se utiliza da abordagem do *Linked Data*, explorando a *web* dos dados através de ligações que permitem que um usuário (máquina e não-máquina) encontrem dados relacionados a quaisquer outros dados dentro desse ambiente (BERNERS-LEE, 2006).

Baseados na proposta de Berners-Lee, vários autores procuram identificar o termo. Portanto, nesta pesquisa serão abordadas algumas definições consideradas importantes dentro desse tema. É importante frisar que essa análise não fará distinção entre os termos “*Linked Data*” e “*Linked Open Data*”, levando em consideração das definições do termo “*Linked Data*”, uma vez que ele também engloba o termo “*Open*”.

A partir da definição de Berners-Lee (2006), Bizer, Heath e Berners-Lee (2009) expõem *Linked Data* como o jeito de usar a *web* para criar ligações entre diferentes dados de diferentes fontes. Para os autores, os dados publicados de modo que possam ser legíveis pelas máquinas, o significado passa a ser mais definido, conectado diretamente à ligação de conjunto de dados externos, ligados a outros conjuntos de dados e assim sucessivamente (ARAKAKI, 2016).

Para Baker et al. (2011), *Linked Data* refere-se aos dados publicados com o propósito de facilitar as ligações entre os conjuntos de dados, elementos e vocabulários de valor.

Heath e Bizer (2011) entendem que o termo *Linked Data* diz respeito a um conjunto de melhores práticas para a interligação entre os dados inseridos de forma estruturada no ambiente *web*.

O relacionamento do conceito de *Linked Data* não com a ligação entre dados entre si, mas com o entendimento de melhores práticas para esse relacionamento foi base para muitos outros autores.

Bizer, Heath e Berners-Lee (2011, p. 205, tradução nossa) definiram que: “O termo *Linked Data* refere-se a um conjunto de práticas recomendadas para a publicação de dados estruturados na Web”.

Para Baker (2012), o conceito se relaciona com ferramentas para a ligação dos dados. Para o autor, *Linked Data* é um corpo de dados em crescimento dentro da rede mundial, ligados através do RDF com a utilização de identificadores de recursos, como por exemplo o URI, identificando assim, não só os recursos descritos, como também os termos usados para a descrição.

Em 2013, a W3C trouxe a definição do *Linked Data* como “[...] uma coleção de conjunto de dados inter-relacionados na Web [...]” (WORLD WIDE WEB CONSORTIUM, 2013³, não paginado apud ARAKAKI, 2016, p. 24).

É interessante notar que, a relação do termo com melhores práticas e práticas recomendadas para a publicação de dados estruturados na *web* foi percebida por Heath e Bizer (2011), porém, as melhores práticas as quais os autores mencionam também foram cunhadas por Tim Berners-Lee no ano de 2006, sendo elas:

- Use os URIs para nomear as coisas;
 - Use os URIs e o HTTP para que possam procurar esses nomes;
 - Forneça informações úteis ao pesquisar um URI;
 - Inclua *links* para outros URIs para ser possível descobrir mais coisas.
- (BIZER; HEATH; BERNERS-LEE, 2011, p. 206, tradução nossa)

Silva (2013) enfatiza que o termo está relacionado com a ligação entre os dados dentro da *web* e não aos documentos em si.

Pomerantz (2015) entende que os princípios propostos por Tim Berners-Lee devem ser vistos como sugestões de como os dados podem ser estruturados para que dessa forma, tecnologias com infraestrutura em constante transição possam ser interoperáveis.

Embora com várias definições, é possível identificar que o *Linked Data* foi criado para tentar resolver questões de recuperação das informações armazenadas no ambiente *web* e desfazer os silos, ou seja, as “ilhas” informacionais que foram sendo criadas ao longo da disponibilização da informação nesse ambiente sem a preocupação com a recuperação (JESUS; CASTRO, 2019).

O avanço das tecnologias *Open Source* e do movimento de democratização da informação colocam os repositórios digitais em evidência, uma vez que derivados desses movimentos, proporcionam recursos baixos para investimentos e manutenção para o desenvolvimento de pesquisas. A interoperabilidade, também propiciada por tecnologias acessíveis dão aos repositórios e seus usuários um acervo maior para as pesquisas, com a comodidade de uma interface única (SOUZA; VIERA, 2012).

Para tanto, foi necessária a criação de um novo protocolo, que pudesse ir além do já consolidado *Harvesting*. É nesse contexto que surge o protocolo OAI-ORE, que traz uma

³ WORLD WIDE WEB CONSORTIUM. Semantic Web. **W3C**, 2013. Disponível em: <http://www.w3.org/standards/semanticweb/>. Acesso em: 25 ago. 2019.

proposta diferente do OAI-PMH, embora também criado pela mesma Iniciativa e com características semelhantes para o acesso aberto.

Iniciado em 2006, financiado por *Andrew W. Mellon Foundation*, *National Science Foundation* e *Microsoft*, sendo publicada sua primeira versão (beta) em 2008 (ORDUÑA-MALEA, 2009), o OAI-ORE transforma a concepção de repositórios de simplesmente armazenamento de documentos para algo além do intercâmbio desses recursos, permitindo intercambiar objetos digitais compostos (objetos digitais e suas agregações), ou seja, páginas da *web* que geralmente são descritas como um recurso único. Com o uso do OAI-ORE, as páginas *web* e os recursos nela inseridos ganham uma descrição (SOUZA; VIERA, 2012).

O protocolo OAI-ORE proporciona a descrição de agregações na *web*, isso significa que ele permite que uma página tenha sua própria descrição, assim como os recursos dentro dela e as relações entre eles, não sendo mais tratados como uma coisa única, porém objetos diferentes que estão relacionados. O protocolo, permite a descrição tanto do próprio recurso, quanto de relações que esse recurso possa ter com outros recursos dentro de um determinado repositório, ou de outro repositório, alavancando os resultados de pesquisas e a visibilidade desses ambientes (SOUZA; VIERA, 2012).

A OAI explica o funcionamento do OAI-ORE

[...] define padrões para a descrição e troca de agregações de recursos da Web. Essas agregações, às vezes chamadas de objetos digitais compostos, podem combinar recursos distribuídos com vários tipos de mídia, incluindo texto, imagens, dados e vídeo. O objetivo desses padrões é expor o conteúdo rico nessas agregações a aplicativos que oferecem suporte à criação, depósito, troca, visualização, reutilização e preservação. (OPEN ARCHIVES, [201-?c], não paginado, tradução nossa)

Para Brooking et al. (2009, p. 149, tradução nossa), “[...] o OAI-ORE visa apoiar a criação, gestão e disseminação de novas formas de recursos digitais compostos sendo produzidos e pesquisados, tornando as informações inseridas neles (aqui entendidas como metadados) detectáveis, legíveis, interoperáveis e reutilizáveis”.

O protocolo OAI-ORE define padrões para descrever e trocar recursos na Web, refere-se à exportação de recursos especiais representados em RDF que codificam objetos compostos. Esses novos objetos descritos pelo protocolo OAI-ORE são denominados de “agregações” (LA BRUZZO; MANGHI; BARDI, 2013).

Ao todo, o OAI-ORE captura os recursos e suas representações, criando quatro tipos de recursos, sendo esses:

Agregação: recurso que agrupa outros recursos, chamados ‘recursos agregados’;

Recurso agregado: um recurso que pertence à uma agregação, ou seja, a representação de um objeto informacional em um objeto composto;

Mapa de recursos: descrição em série de uma agregação [...]. Um mapa de recursos lista os recursos agregados e possui propriedades sobre a agregação e seus recursos agregados;

Proxy: recurso que permite firmar relacionamentos entre os recursos agregados no contexto de uma agregação específica.
(LA BRUZZO; MANGHI; BARDI, 2013, p. 38, tradução nossa)

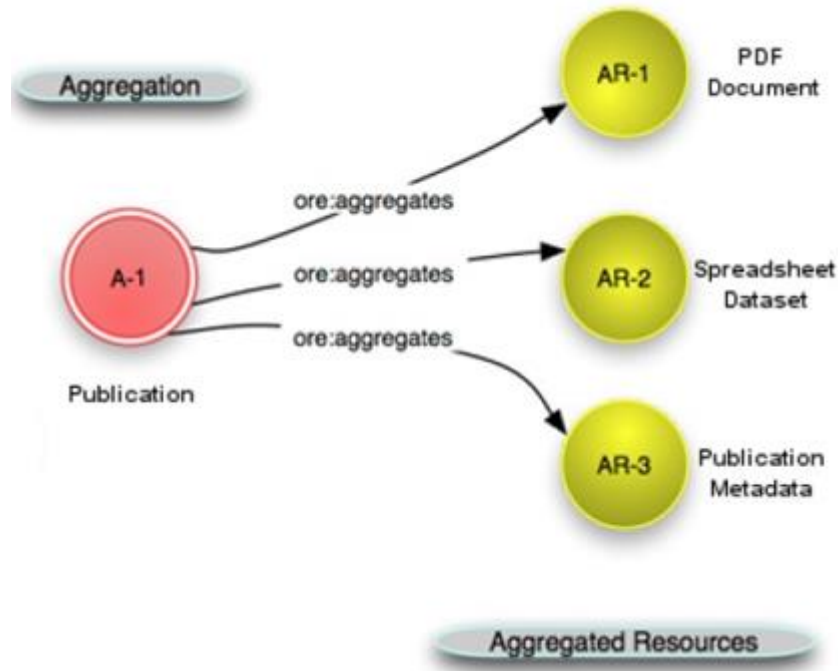
Para a implementação do OAI-ORE em um sistema, são necessários, no mínimo, as seguintes tecnologias:

- Ontologia: um vocabulário para descrever relacionamentos, definindo seu domínio e escopo, avaliando ontologias já existentes e sua aplicabilidade, enumerando os termos e conceitos importantes, definindo as classes e suas hierarquias, definindo os atributos (ou propriedades) de cada classe, definindo os tipos e valores permitidos para cada propriedade. Uma vez estruturados esses passos, a ontologia pode ser codificada em uma linguagem de ontologia apropriada, como a *Web Ontology Language* (OWL), um padrão W3C;
- Ferramenta de validação de ontologia: essa ferramenta permitiria que os arquivos verificassem a implementação da ontologia e seus metadados descritivos de arquivamentos em OAI-ORE;
- Modelos de mapas de recursos: um mapa de recurso identifica os objetos em uma agregação e define os limites dessa agregação, além disso, fornece as diretrizes para a construção de mapas de recursos;
- Diretrizes de melhores práticas: Essa documentação forneceria orientações para criar e implementar de maneira mais eficiente os recursos para pesquisa OAI-ORE. “Um dos principais componentes dessas diretrizes e documentação seria a orientação para a construção de mapas de recursos. O mapa de recursos, juntamente com a ontologia, formaria o núcleo de um auxílio à descoberta baseado em OAI-ORE” (KAPLAN; SAUER; WILCZEK, 2011, não paginado, tradução nossa).

Por ser baseado na arquitetura *web*, o OAI-ORE utiliza-se do modelo RDF para descrever os objetos e as relações entre eles.

Paralelamente, é introduzido o conceito de “Agregações” e “Recursos agregados”, sendo uma agregação um conjunto de recursos agregados, todos representados por uma URI. Aplicando esse conceito à um recurso de um repositório, tem-se a Figura 11:

Figura 14 - Exemplo de agregação do OAI-ORE

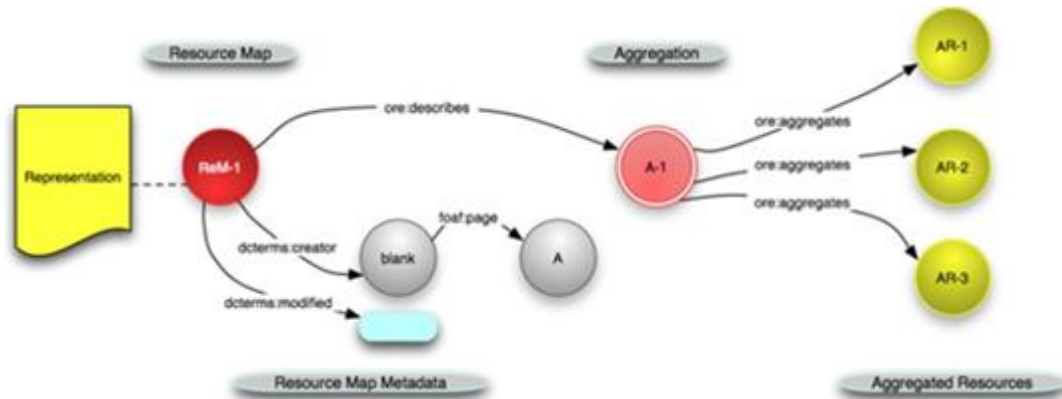


Fonte: Tarrant et al. (2009, não paginado)

Na Figura 11, é representado um único registro de publicação, porém, o conceito é mais abrangente e permite o “aninhamento” de agregações. Segundo os autores, uma agregação do nível mais alto poderia ser o próprio repositório que, por sua vez, contém seus próprios recursos agregados (TARRANT et al., 2009).

No nível mais alto do modelo OAI-ORE estão os mapas de recursos. Um mapa de recursos também possui uma URI e este é usado para descrever uma única agregação, assim como só pode ser vinculado à essa mesma agregação. Assim, um modelo completo, pode ser representado na Figura 12:

Figura 15 - Modelo completo do OAI-ORE



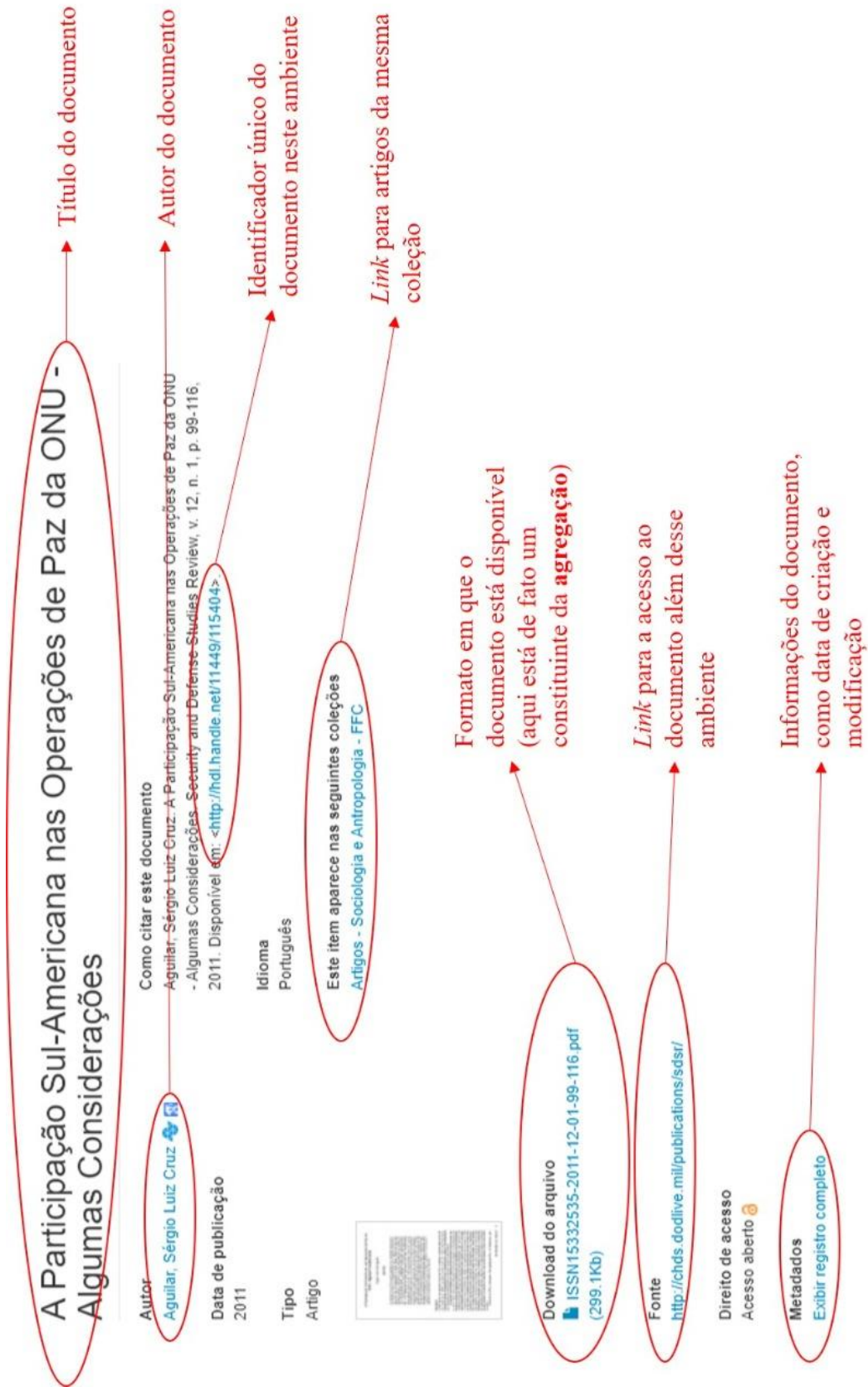
Fonte: Tarrant et al. (2009, não paginado)

Na Figura 12, com o modelo completo, está representado o mapa de recurso (ReM-1) em vermelho, à esquerda da agregação com alguns metadados expressos usando o *Dublin Core*.

Ao basear-se no RDF, o mapa de recurso se torna capaz de suportar quaisquer *namespace* e metadados que o usuário deseja adicionar (TARRANT et al., 2009).

O objetivo aqui é fornecer um padrão para identificar as agregações dos recursos e descrever tanto os constituintes como os limites das agregações (WITT, 2010). A Figura 13 ilustra um exemplo de descrição do OAI-ORE.

Figura 16 - Exemplo de descrição do OAI-ORE



Fonte: Elaborado pela autora

A Figura 13 aponta como ficariam os elementos de uma página iniciada do Repositório Institucional Unesp se fosse descrito com o OAI-ORE. Nesse caso, cada elemento, desde o URI, até o recurso em texto a qual a página se refere teria um mapa de recurso próprio, contendo sua descrição e relação com cada um dos outros elementos, além da descrição de cada relação, como por exemplo, o documento intitulado “A Participação Sul-Americana nas Operações de Paz da ONU - Algumas Considerações” pertence ao autor “Aguilar, Sérgio Luiz Cruz”. Vale destacar aqui que os sistemas atuais não conseguem fazer uma relação com esta pesquisa e a pesquisa do exemplo citado.

A grande motivação por trás do uso do OAI-ORE se dá a partir do momento em que este consegue se aproximar um pouco mais ao ideal de *web* semântica, uma vez que configura uma maneira de relacionar objetos (reais ou abstratos) da mesma maneira que se faz no mundo real, além de sobrepor esses objetos.

Um exemplo, no mundo físico (ou mundo real), podemos coletar fotos de diferentes lugares e guardá-las em um único álbum, do mesmo modo em que, ainda que essas fotos estejam espalhadas, é fácil lembrar destas, uma vez que estiverem relacionadas, como tiradas na mesma data ou pertencerem ao mesmo grupo familiar, por exemplo.

Estendendo essa prática para dentro da *web*, também é possível agrupar fotos, ou adicionar *Uniform Resource Locators* (URLs) em marcadores como listas ou “favoritos”. Apesar disso, embora exista o uso dessas agregações, sua existência na *web* se torna rápida, uma vez que não existe uma maneira padrão de identificar uma agregação e o uso da URI é feita para identificar a agregação inteira.

Por exemplo, usamos o URI da primeira página de um documento da Web com várias páginas para identificar o documento inteiro ou o URI da página HTML que fornece acesso a um conjunto do Flickr para identificar o conjunto inteiro de imagens. Mas esses URIs realmente apenas identificam essas páginas específicas, e não a união de páginas que compõe o documento inteiro, ou a união de todas as imagens em um conjunto do Flickr, respectivamente. Em essência, o problema é que não existe uma maneira padrão de descrever os constituintes ou os limites de uma agregação, e é isso que o OAI-ORE pretende fornecer. (LAGOZE et al., 2008, não paginado, tradução nossa)

Em se tratando de ambientes informacionais dentro do ambiente *web*, como bibliotecas e repositórios digitais, seu gerenciamento é focado em torno do sistema que abrigam suas coleções, onde um bibliotecário decide quais recursos digitais irão agrupados e dentro desses recursos. Nesse caso, é gerado uma série de páginas da *web* que devem ser navegadas através dos links dos recursos até que se atenda a necessidade informacional do usuário e, embora esses

ambientes possam expor e coletar metadados, os próprios recursos estão em seus silos proprietários (WITT, 2010).

Assim, a proposta do OAI-ORE é quebrar esses silos, expondo a semântica dos recursos, fornecendo ganchos para recuperá-los sem a necessidade de um link para acesso. Desse modo, a liberação do conteúdo dos silos dessas bibliotecas, além de permitir a reutilização e troca, pode expandir a construção de “coleção”, uma vez que cada usuário (humano e máquina) seria capaz de montar suas próprias coleções (WITT, 2010).

Pode-se considerar que o mapa de recurso é o elemento mais importante do OAI-ORE, uma vez que é este o elemento que descreve, de forma “palpável” os recursos e suas agregações. Os documentos do próprio OAI-ORE descrevem três formatos diferentes para a serialização dos mapas, ou seja, formatos diferentes de escrevê-los, transmitidos e lidos por máquinas, sendo esses formatos o RDF/XML, RDFa e Atom XML (WITT, 2010).

Para exemplificar como se dá essa serialização, será utilizado o formato RDF/XML, uma vez que este é o formato mais utilizado.

No exemplo de Witt (2010), temos uma agregação para uma página de jornal. Vale ressaltar que as mesmas diretrizes para a aplicação gráfica do RDF (recurso, propriedade, valor) são aplicadas ao estruturar um mapa de recursos. Acontece que os predicados e o assunto da tripla representada são os URI da agregação.

1. *A agregação é uma página de jornal.*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page> rdf: tipo rdf: resource =
<http://chroniclingamerica.loc.gov/terms#Page>
2. *Seq-1.rdf é um mapa de recursos.*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.rdf> rdf: tipo rdf: resource =
 “<http://www.openarchives.org/ore/terms/ResourceMap>”
3. *A página Agregação é descrita por um Mapa de Recursos, seq-1.rdf.*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page> minério: isDescribBy rdf: resource =
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.rdf>
4. *O Mapa de Recursos, seq-1.rdf, descreve a página Agregação.*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.rdf> minério: descreve
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page>
 (WITT, 2010, p. 18, tradução nossa)

Em um mapa de recursos, expressa-se alguns metadados básicos como o criador e a data de modificação.

1. *O Mapa de Recursos, seq-1.rdf, possui um criador do dlc (o símbolo da OCLC da Biblioteca do Congresso).*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.rdf> dcterms: creator rdf: resource =
<http://chroniclingamerica.loc.gov/awardees/dlc#premiado>
2. *O Mapa de Recursos, seq-1.rdf, foi modificado pela última vez em 13 de fevereiro de 2010 às 13:06.*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.rdf> dcterms: modificado rdf: datatype =
 "http://www.w3.org/2001/XMLSchema#dateTime" 2010-02-13T13:06:31-04:00
3. *O Mapa de Recursos, seq-1.rdf, foi criado em 13 de fevereiro de 2010 às 13:06.* <http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.rdf> dcterms: created rdf: datatype =
<http://www.w3.org/2001/XMLSchema#dateTime> "2010-02-13T13:06:31-04:00
 (WITT, 2010, p. 20, tradução nossa)

Também podem ser expressos metadados sobre a agregação descrita.

1. *A página Agregação foi publicada em 14 de dezembro de 1918.*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page> dcterms: emitido rdf: datatype=
 "http://www.w3.org/2001/XMLSchema#date" 1918-12-14
2. *A página Agregação é intitulada "O observador de São José. - 1918-12-14 - 1" (o título do jornal, sua data e o número da página).*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page> dcterms: title "O observador de São José. - 1918-12-14 - 1"
3. *A página Agregação possui um número de sequência 1 (ou seja, é a primeira página do jornal).*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page> ndnp: sequência rdf: datatype =
 "http://www.w3.org/2001/XMLSchema#long" 1
4. *A página Agregação é representada por thumbnail.jpg (uma imagem em miniatura pode servir como metadado sobre um objeto).*
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page> foaf: representação rdf: resource =
<http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1/thumbnail.jpg>
 (WITT, 2010, p. 20, tradução nossa)

Além dessas descrições, temos o relacionamento entre a página Agregação e seus recursos agregados. Nesse caso, cada página consiste em uma imagem JPG, um JPG em miniatura, um PDF, o texto em si (considerando os caracteres) e sua saída estruturada em XML. O mapa de recursos descreve que todos esses arquivos são agregados em uma página:

1. *A página Agregação agrega um arquivo JPEG 2000, seq-1.jp2.*
[http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource = http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.jp2](http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource=http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.jp2)
 2. *A página Agregação agrega um arquivo PDF, seq-1.pdf.*
[http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource = http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.pdf](http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource=http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1.pdf)
 3. *A página Agregação agrega um arquivo de imagem em miniatura, thumbnail.jpg.* [http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource = http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/thumbnail.jpg](http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource=http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/thumbnail.jpg)
 4. *A página Agregação agrega um arquivo de texto OCR, ocr.txt.*
[http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource = http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/ocr.txt](http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource=http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/ocr.txt)
 5. *A página Agregação agrega um arquivo XML de OCR, ocr.xml.*
[http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource = http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/ocr.xml](http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/seq-1#page_minério:agrega_rdf:resource=http://chroniclingamerica.loc.gov/lccn/sn90061457/1918-12-14/ed-1/ocr.xml)
- (WITT, 2010, p. 20, tradução nossa)

Além desses, os mapas de recursos também podem descrever as páginas, volumes e títulos em lote, emitindo páginas agregadas e volumes agregados (WITT, 2010).

Figura 17 - Exemplo de descrição em RDF e OAI-ORE



Fonte: Elaborado pela autora

A Figura 14 traz de modo mais visual o exemplo da descrição em RDF e OAI-ORE. Nesse exemplo, temos uma primeira página qualquer (Seq-1) que tem seus metadados estruturados em Dublin Core (dc.) e em RDF (rdf) em um outro recurso que seria o mapa de recuso. Além de a primeira página, todos os recursos a ela agregados possuem seus próprios mapas de recurso com as suas descrições e seus relacionamentos (e é aí que encontramos o OAI-ORE).

No Brasil, não foram encontrados sistemas que utilizem o OAI-ORE, o mais próximo do tema, foi o Sistema Aberto e Integrado de Informação em Agricultura (SABIIA), que é um provedor de serviços, coletando os metadados dos provedores de dados do tema de interesse, porém, nada com o próprio OAI-ORE.

Fora do Brasil, vários lugares implementaram o protocolo, como por exemplo a *Texas Digital Library*, o Observatório Virtual Nacional dos EUA e o *Australian Literature Resource* (uma colaboração entre a Biblioteca Nacional da Austrália e doze bibliotecas universitárias australianas), além de outros.

Pode-se citar a Europeana como o sistema exemplo que utiliza de forma mais completa as ferramentas aqui abordadas.

A Europeana (projeto emblemático de biblioteca digital da União Europeia) pode ser considerada um provedor de serviços, uma vez que reúne os dados de diversos outros repositórios, tornando-se um ponto de acesso para milhares de recursos informacionais de mais de 15 países. Em fevereiro de 2012 foi lançado a primeira versão do conjunto de dados representados no *Europeana Data Model* (EDM) de acordo com os princípios do *Linked Data* (ISAAC; HASLHOFER, 2013)

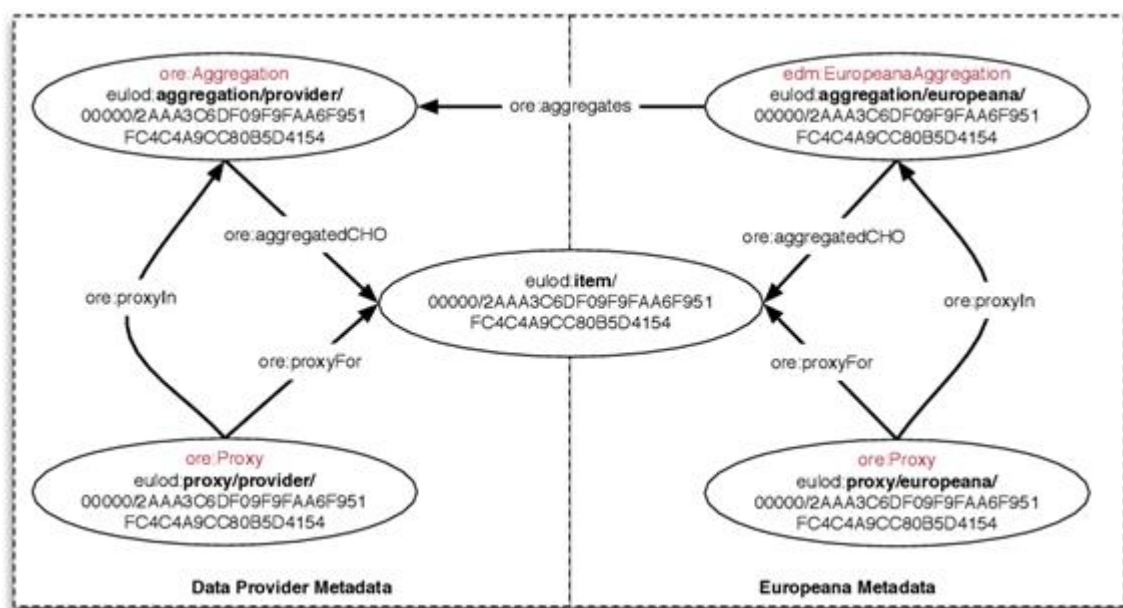
Esse modelo de dados desenvolvido pela própria Europeana é um modelo flexível e preciso para anexar os recursos ao sistema. Os principais requisitos na estrutura básica do EDM incluem:

- Distinção entre os objetos fornecidos (pintura, livro, etc.) e representações digitais;
- Distinção entre um item e seus metadados de descrição;
- Permissão de vários registros de um mesmo item, contendo declarações contraditórias, ou seja, representar de diferentes perspectivas um determinado objeto cultural;

- Permite representar informações contextuais, na forma de entidades (lugares, agentes, períodos).

Na utilização do OAI-ORE originam-se *proxies*, ou seja, assuntos de declarações descritivas (criador, assunto, data de criação) para cada item, fornecidos pela própria Europeiaana. Esses *proxies* permitem a separação de diferentes visualizações para um mesmo recurso no contexto de diferentes agregações, permitindo distinguir os metadados originais para o objeto dos metadados criados pela Europeiaana.

Figura 18 - Estrutura básica do EDM e identificação do *proxy*



Fonte: Isaac e Haslhofer (2013)

Na Figura 15, um item identificado como “00000/2AAA3C6DF09F9FAA6F951”, agregado ao sistema da Europeiaana recebeu o *proxy* com esse mesmo nome, fazendo com que, tanto o sistema provedor de dados, quanto o sistema provedor de serviços (a Europeiaana) identifiquem o item.

Cada recurso e cada item dentro dos recursos teriam suas descrições em RDF o OAI-ORE, escritas de forma a identificá-los e interligá-los;

O recurso pode ou não pertencer ao sistema, mas interage com sistemas exteriores a partir de suas descrições e dos protocolos, por exemplo, se esse sistema “abriga” um recurso de outro sistema, além de receber um tratamento semelhante aos dos itens do próprio sistema de descrição, também deve poder conduzir o usuário ao recurso em seu sistema original, mesmo que esse recurso esteja dentro de um item;

Cada item e cada recurso ser descrito e interligado através dos mapas de recurso.

De forma geral, um sistema construído para suportar a utilização dessas ferramentas funcionaria da seguinte forma:

Figura 19 - Exemplo de um sistema informacional



Fonte: Elaborado pela autora

O exemplo da Figura 16 ilustra um sistema configurado para suportar um padrão de metadados, os protocolos OAI-PMH, OAI-ORE, assim como RDF e a linguagem XML, tido como "Meu sistema". Nesse caso, qualquer página iniciada neste sistema, assim como qualquer item e itens dentro deste tenham suas próprias descrições e relações através de mapas de recursos, ou seja, um item pesquisado X possui sua descrição estruturada nas ferramentas citadas, e este mesmo item trás com ele outros elementos, que podem ser uma imagem (recurso y) ou um autor (recurso Z), tendo cada um a sua própria descrição, estando todos relacionados entre si.

5 CONSIDERAÇÕES FINAIS

A interoperabilidade é um requisito importante a ser estabelecido entre sistemas e ambientes informacionais heterogêneos, mas carrega uma complexidade para ser estabelecida. Assim, o objetivo inicial desta pesquisa era “Estudar as principais características e níveis de interoperabilidade estabelecidos na literatura”. Com isso, foi possível verificar na literatura os diferentes níveis e aspectos da interoperabilidade.

Os tipos e níveis de interoperabilidade são nomeados de modo diferente por vários autores e referem-se aos níveis de trocas de dados e metadados. Vão desde um nível básico, trocando apenas metadados simples através da agregação, até níveis mais complexos, como o *harvesting* por meio da implementação do OAI-PMH e a federação, onde todos os sistemas usam os mesmos padrões e protocolos estabelecidos em uma rede de ambientes informacionais. Esses níveis de interoperabilidade apresentam aspectos mais específicos para o seu estabelecimento, esses aspectos se referem às questões técnicas, semânticas, estruturais, legais entre outras. O estabelecimento desses aspectos está interrelacionado e sua consolidação contribui para uma interoperabilidade mais efetiva.

Encontra-se na literatura três níveis de interoperabilidade: a “agregação” é utilizada para referir-se a um nível mais básico de interoperabilidade, trocando metadados básicos, ainda que os sistemas não sejam de uma estrutura complexa. O “*harvesting*”, que se trata da coleta de metadados em provedores de dados e provedores de serviços e a Federação, que é o nível mais alto de interoperabilidade estabelecido por meio de acordos internacionais em rede. Entretanto, foi identificado na pesquisa que o termo “agregação” apresenta um sentido diferente quando está no contexto do protocolo OAI-ORE, pois trata-se do conjunto de recursos e suas descrições agregados e disponibilizados na *Web*.

Para que a interoperabilidade ocorra é necessário o uso de protocolos, por isso, um dos objetivos da pesquisa era de “Analisar as características dos protocolos OAI-PMH e OAI-ORE e sua aplicação em provedores de dados e serviços”, pois constituem-se como os protocolos mais utilizados para a interoperabilidade na *Web*.

O protocolo OAI-PMH é considerado o mais utilizado para a troca de metadados em provedores de dados e provedores de serviços. A OAI disponibiliza em seu *site* uma lista dos sistemas que utilizam o OAI-PMH, separando os provedores de dados e os provedores de serviços, os quais, os provedores de dados expõem os seus metadados e os provedores de serviços os coletam.

Com uma proposta diferente, o protocolo OAI-ORE não se preocupa com o transporte dos metadados, mas com a descrição de todos os recursos compostos e relacionados, criando uma Agregação de descrições e objetos. Para o protocolo OAI-ORE não há diferença entre provedores de dados ou provedores de serviços, mas se cada recurso apresenta uma descrição adequada que possa ser.

O OAI-PMH é utilizado em diversos ambientes informacionais, uma vez que facilita a coleta de metadados de outros sistemas e proporciona uma gama maior no momento da busca. No Brasil, pode-se citar como exemplo de provedor de dados diversas revistas, como a Revista Brasileira de Ciências da Comunicação (Intercom), Revista Brasileira de Herbicidas (RBH), entre outras; já como provedor de serviços, na lista da própria OAI, temos apenas o Clio-i, do Laboratório Liber da Universidade Federal de Pernambuco (UFPE).

No site do OAI existe uma lista de alguns ambientes informacionais que utilizam o protocolo OAI-PMH. Já o protocolo OAI-ORE não apresenta uma lista, por isso não há como saber quantos ambientes informacionais utilizam esse protocolo, mas foi possível averiguar na pesquisa que a Biblioteca Digital do Texas (*Texas Digital Library*) e a Europeana utilizam o protocolo OAI-ORE para a agregação de recursos.

Quanto aos objetivos de “Estudar os aspectos atuais para o uso e reuso dos dados disponibilizados na *web* e no contexto *do Linked Data*” e “Analisar as contribuições dos protocolos citados para os ambientes informacionais digitais na *web* e no *Linked Data*”, pode-se ressaltar que tanto o protocolo OAI-PMH quanto o protocolo OAI-ORE possibilitam o uso e reuso dos dados na Web e no *Linked Data* de diferentes formas.

Os resultados mostram que embora o OAI-ORE traga uma proposta enriquecedora por descrever recursos e suas agregações, o uso do protocolo ainda é inicial entre a maioria dos ambientes informacionais no Brasil.

A utilização do protocolo OAI-PMH é bem maior seja com provedores de dados ou provedores de serviços, porém, a ampla utilização do OAI-ORE traria vantagens maiores e ainda não exploradas pela maioria dos sistemas e ambientes informacionais, em relação ao reuso dos dados, a não duplicidade de trabalhos e a amplificação dos resultados de pesquisas.

Com base nos resultados apresentados é possível considerar que a interoperabilidade se constitui como algo crucial diante da heterogeneidade dos ambientes informacionais digitais, além de peça-chave na concepção de *Linked Data*.

Considera-se que o uso conjunto dos protocolos OAI-PMH e OAI-ORE seriam de significativa importância para o estabelecimento da interoperabilidade e do uso e reuso dos dados, pois o primeiro protocolo é eficaz por expor e coletar metadados e o segundo protocolo

utiliza os metadados para uma descrição mais detalhada proporcionando agregações e conexões mais enriquecidas entre os recursos na *Web* e no *Linked Data*.

A atual construção e consolidação do *Linked Data* é mais um passo para alcançar uma *Web Semântica* de qualidade, na qual, o protocolo OAI-ORE seria um complemento significativo para a efetivação do *Linked Data*.

A Europeia pode ser destacada como o ambiente que melhor utiliza os recursos dos protocolos OAI-PMH, OAI-ORE e da linguagem RDF.

Atualmente algumas pesquisas estão abordando a questão das publicações ampliadas, ou seja, as publicações digitais complexas que combinam recursos heterogêneos de diversos tipos e ambientes, agregando os dados de pesquisa que dependem das possibilidades das ligações da web.

Assim, acredita-se que o protocolo OAI-ORE seria uma ferramenta de grande valia na interação entre as publicações ampliadas, uma vez que sua forma de descrição e o conceito de agregação estabelecido pelo protocolo parecem se encaixar às propostas das publicações ampliadas. Entretanto, por não ser objeto de pesquisa, o uso do protocolo OAI-ORE nas publicações ampliadas constitui-se como um tema futuro a ser estudado.

Assim, considera-se que o uso dos protocolos estudados nesta pesquisa possibilita uma maior relação entre os recursos informacionais, suas relações com objetos complexos, corroborando ainda mais com a conectividade, o compartilhamento, o uso e o reuso das informações por parte das comunidades científicas, na *Web* e no *Linked Data*.

REFERÊNCIAS

- ALVES, Maria das Dores Rosa; SOUZA, Marcia Izabel Fugisawa. Estudo de correspondência de elementos metadados: DUBLIN CORE e MARC 21. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 5, n. 1, p. 20-38, jan./jun. 2007. <https://doi.org/10.20396/rdbci.v4i2.2019>
- ALVES, Rachel Cristina Vesú. **Metadados como elementos do processo de catalogação**. 2010. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.
- ALVES, Rachel Cristina Vesú; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013.
- ARAKAKI, Felipe Augusto. **Linked Data: ligação de dados bibliográficos**. 2016. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016.
- ARAKAKI, Felipe Augusto; SANTOS, Plácida Leopoldina Ventura Amorim da Costa; ALVES, Rachel Cristina Vesú. Panorama das pesquisas sobre o padrão de metadados Dublin Core no Brasil. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 20, n. 1, p. 86-97, jan./abr. 2015. Disponível em: <https://revista.acbsc.org.br/racb/article/view/983>. Acesso em: 24 maio 2018.
- ARMS, William Y.; HILLMANN, Diane; LAGOZE, Carl; KRAFFT, Dean; MARISA, Richard; SAYLOR, John; TERRIZZI, Carol; VAN DE SOMPEL, Herbert. A spectrum of interoperability: the site for science prototype for the NSDL. **D-Lib Magazine**, Reston, VA, v. 8, n. 1, Jan. 2002. Disponível em: <http://www.dlib.org/dlib/january02/arms/01arms.html>. Acesso em: 7 mar. 2019.
- BAKER, Thomas. Libraries, languages of description, and linked data: a Dublin Core perspective. **Library Hi Tech**, Bingley, v. 30, n. 1, p. 116-133, 2012. <https://doi.org/10.1108/07378831211213256>
- BAKER, Thomas; BERMÈS, Emmanuelle; COYLE, Karen; DUNSIRE, Gordon; ISAAC, Antoine; MURRAY, Peter; PANZER, Michael; SCHNEIDER, Jodi; SINGER, Ross; SUMMERS, Ed; WAITES, William; YOUNG, Jeff; ZENG, Marcia. Library Linked Data Incubator Group final report. **W3C Incubator Group Report**, 25 Oct. 2011. Disponível em: <https://www.w3.org/2005/Incubator/ld/XGR-ld-20111025/>. Acesso em: 17 maio 2019.
- BARDIN, Laurence. **Análise de conteúdo**. Lisboa: Edições 70, 2011.
- BERNERS-LEE, Tim. Linked Data. **W3C**, Design issues, 27 jul. 2006. [Last change: 18 jun. 2009]. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 20 jun. 2019.
- BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked Data: the story so far. In: SHETH, Amit (ed.). **Semantic services, interoperability and web applications: emerging concepts**. Hershey, PA: Information Science Reference, 2011. p. 205-227.

BROOKING, Charles; SHOULDICE, Stephen R.; ROBIN, Gautier; KOBE, Bostjan; MARTIN, Jennifer L.; HUNTER, Jane. Comparing METS and OAI-ORE for encapsulating scientific data products: a protein crystallography case study. *In: IEEE INTERNATIONAL CONFERENCE ON E-SCIENCE*, 5., 2009, Oxford, UK. **Proceedings** [...]. Piscataway, NJ: IEEE, 2009. p. 148-155. <https://doi.org/10.1109/e-Science.2009.29>

CAFÉ, Lígia; FACHIN, Gleisy Regina Bóries. Provedores de dados, provedores de serviços e periódicos em Ciência da Informação, Biblioteconomia e áreas afins. **Encontros Bibli**, Florianópolis, n. esp., p. 59-76, 2007. <https://doi.org/10.5007/1518-2924.2007v12nesp1p59>

CAMPEROS REYES, Jacquelin Teresa. **Metadados nas instruções de governos para publicadores de dados**. 2018. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2018.

CAMPOS, Claudinei José Gomes. Método de análise de conteúdo: ferramenta para a análise de dados qualitativos no campo da saúde. **Revista Brasileira de Enfermagem**, Brasília, DF, v. 57, n. 5, p. 611-614, set./out. 2004. <https://doi.org/10.1590/S0034-71672004000500019>

CAMPOS, Luiz Fernando de Barros. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Encontros Bibli**, Florianópolis, v. 12, n. 23, p. 16-46, 2007. <https://doi.org/10.5007/1518-2924.2007v12n23p16>

DRAME, Anja. Can machines be smart? **ISO Focus**, Genève, v. 5, n. 6, p. 13-14, Apr. 2008.

DZIEKANIAK, Gisele Vasconcelos; KIRINUS, Josiane Boeira. Web semântica. **Encontros Bibli**, Florianópolis, v. 9, n. 18, p. 20-39, 2004. <https://doi.org/10.5007/1518-2924.2004v9n18p20>

ELLIOTT, Alison. Facilitating discovery, interaction and collaboration of knowledge banks. **ISO Focus**, Genève, v. 5, n. 6, p. 10-12, Apr. 2008.

FERREIRA, Jaidir Andrade; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. **Informação & Sociedade: Estudos**, João Pessoa, v. 23, n. 2, p. 13-23, maio/ago. 2013. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/15436>. Acesso em: 2 fev. 2019.

FERREIRA, Sueli Mara S. P.; SOUTO, Leonardo Fernandes. Dos sistemas de informação federados à federação de bibliotecas digitais. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 2, n. 1, p. 23-40, jan./jun. 2006. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/3>. Acesso em: 8 set. 2018.

FRANÇA, Fabiana da Silva; RAMALHO, Francisca Arruda. Informação científica em repositórios digitais: processo de busca e recuperação da informação. *In: ALBUQUERQUE, Maria Elizabeth Baltar Carneiro; NEVES, Dulce Amélia de Brito; OLIVEIRA, Bernardina Maria Juvenal Freire de; MELO, Ana Virginia Chaves de; FRANÇA, Fabiana da Silva (org.). Representação da informação: um universo multifacetado*. João Pessoa: UFPB, 2013. p. 207-235.

FUSCO, Elvis. **Aplicação dos FRBR na modelagem de catálogos bibliográficos digitais**. São Paulo: Cultura Acadêmica, 2011. Disponível em: <http://hdl.handle.net/11449/109186>. Acesso em: 03 dez. 2020.

GARCIA, Patrícia de Andrade Bueno; SUNYE, Marcos Sfair. O protocolo OAI-PMH para interoperabilidade em bibliotecas digitais. *In*: CONGRESSO DE TECNOLOGIA PARA GESTÃO DE DADOS E METADADOS DO CONE SUL, 1., 2003, Ponta Grossa. **Anais [...]**. Ponta Grossa: UEPG, 2003. Disponível em: <http://conged.deinfo.uepg.br/~iconged/>. Acesso em: 17 set. 2017.

GRÁCIO, José Carlos Abbud. **Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade**. 2002. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2002.

GUAJARDO SALINAS, Aldo. Z39.50 y OAI-PMH: protocolos de transferencia y recuperación da información. *In*: CONFERENCIA INTERNACIONAL DE BIBLIOTECOLOGIA, 15., 2010, Santiago. **Anales [...]**. Santiago: Colegio de Bibliotecarios de Chile, 2010. Disponível em: http://www.bibliotecarios.cl/descargas/2010/11/guajardo_z3950.pdf. Acesso em: 20 maio 2017.

HEATH, Tom; BIZER, Christian. **Linked Data: evolving the web into a global data space**. San Rafael, CA: Morgan & Claypool, 2011. (Synthesis lectures on the Semantic Web: theory and technology, v. 1).

INNOCENTI, Perla; VULLO, Giuseppina; ROSS, Seamus. Towards a digital library policy and quality interoperability framework: the DL.org Project. **New Review of Information Networking**, Abingdon, v. 15, n. 1, p. 29-53, 2010. <https://doi.org/10.1080/13614571003751071>

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 25964-2:2013: information and documentation: thesauri and interoperability with other vocabularies: part 2: interoperability with other vocabularies**. Geneva: ISO, 2013.

ISAAC, Antoine; HASLHOFER, Bernhard. Europeana Linked Open Data – data.europeana.eu. *Semantic web*, v. 4, n. 3, p. 291-297, 2013

JESUS, Ananda Fernanda de; CASTRO, Fabiano Ferreira de. Dados bibliográficos para o Linked Data: uma revisão sistemática de literatura. **Brazilian Journal of Information Science**, Marília, v. 13, n. 1, p. 45-55, 2019. <https://doi.org/10.36311/1981-1640.2019.v13n1.08.p45>

KAPLAN, Deborah; SAUER, Anne; WILCZEK, Eliot. Archival description in OAI-ORE. **Journal of Digital Information**, Austin, TX, v. 12, n. 2, 2011. Disponível em: <https://journals.tdl.org/jodi/index.php/jodi/article/view/1814/1769>. Acesso em: 13 set. 2019.

LA BRUZZO, Sandro; MANGHI, Paolo; BARDI, Alessia. OAIzer: configurable OAI exports over relational databases. *In*: GAROUFALLOU, Emmanouel; GREENBERG, Jane (ed.). **Metadata and semantics research: 7th Research Conference, MTSR 2013, Thessaloniki**,

Greece, November 19-22, 2013. Proceedings. Heidelberg: Springer, 2013. (Communications in Computer and Information Science, v. 390).

LAGOZE, Carl; VAN DE SOMPEL, Herbert; JOHNSTON, Pete; NELSON, Michael; SANDERSON, Robert; WARNER, Simeon. ORE user guide: primer. **Open Archives**, 17 Oct. 2008. Disponível em: <http://www.openarchives.org/ore/primer>. Acesso em: 10 mar. 2020.

LIU, Xiaoming et al. Arc: an OAI service provider for digital library federation. *D-Lib Magazine*, Reston, VA, v. 7, n. 4, Apr. 2001. Disponível em: <http://www.dlib.org/dlib/april01/liu/04liu.html>. Acesso em: 25 jul. 2001.

LOURENÇO, Cíntia de Azevedo. **Análise do padrão brasileiro de metadados de teses e dissertações segundo o modelo entidade-relacionamento**. 2005. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

MARCONDES, Carlos Henrique; SAYÃO, Luís Fernando. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. **Ciência da Informação**, Brasília, DF, v. 30, n. 3, p. 24-33, set./dez. 2001. Disponível em: <http://revista.ibict.br/ciinf/article/view/909>. Acesso em: 2 fev. 2016.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Metodologia do trabalho científico: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos**. 6. ed. São Paulo: Atlas, 2006.

MARTINS, Dalton Lopes; FERREIRA, Sueli Mara Soares Pinto. Protocolo OAI-PMH e Sistemas Federados de Informação: fundamentos de arquitetura da informação para análise de dados do portal de produção científica da área de Ciências da Comunicação *Univsciencia.org*. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 431-447, set. 2012. <https://doi.org/10.18617/liinc.v8i2.479>

MÉNDEZ RODRÍGUEZ, Eva. **Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales**. [Gijón]: Trea, 2002.

MILLER, Paul. An introduction to the Resource Description Framework. **D-Lib Magazine**, Reston, VA, v. 4, n. 5, May 1998. Disponível em: <http://www.dlib.org/dlib/may98/miller/05miller.html>. Acesso em: 25 maio 2018.

MILLER, Paul. Interoperability: what is it and why should I want it? **Ariadne**, Bath, n. 24, 2000. Disponível em: <http://www.ariadne.ac.uk/issue/24/interoperability/>. Acesso em: 14 maio 2018.

MORAES, Roque. Análise de conteúdo. **Revista Educação**, Porto Alegre, v. 22, n. 37, p. 7-32, 1999.

NEEDLEMAN, Mark H. RDF: the resource description framework. **Serials Review**, Abingdon, v. 27, n. 1, p. 58-61, 2001. <https://doi.org/10.1080/00987913.2001.10764639>

OPEN ARCHIVES. Open Archives Initiative: organization. **Open Archives**, [201-?a]. Disponível em: <https://www.openarchives.org/organization/>. Acesso em: 10 mar. 2020.

OPEN ARCHIVES. Open Archives Initiative: Protocol for Metadata Harvesting. **Open Archives**, [201-?b]. Disponível em: <https://www.openarchives.org/pmh/>. Acesso em: 11 mar. 2020.

OPEN ARCHIVES. Open Archives Initiative: Object Exchange and Reuse. **Open Archives**, [201-?c]. Disponível em: <https://www.openarchives.org/ore/>. Acesso em: 6 abr. 2020.

ORDUÑA-MALEA, Enrique. Reutilización e intercambio de objetos digitales compuestos en la Web: el proyecto OAI-ORE. **Anuario ThinkEPI**, [S. l.], v. 3, p. 45-48, 2009. Disponível em: <http://www.thinkepi.net/anuario-thinkepi/anuario-thinkepi-2009>. Acesso em: 20 jun. 2020.

PAYETTE, Sandra; BLANCHI, Christophe; LAGOZE, Carl; OVERLY, Edward A. Interoperability for digital objects and repositories: the Cornell/CNRI experiments. **D-Lib Magazine**, Reston, VA, v. 5, n. 5, May 1999. <https://doi.org/10.1045/may99-payette>

PECEGUEIRO, Cláudia Maria Pinho de Abreu. Os desafios da recuperação da informação na era digital. **Biblionline**, João Pessoa, v. 15, n. 2, p. 47-55, 2019. <https://doi.org/10.22478/ufpb.1809-4775.2019v15n2.45145>

PEREIRA, Ana Maria; RIBEIRO JÚNIOR, Divino Ignácio; NEVES, Guilherme Luiz Cintra. Metadados para a descrição de recursos da internet: as novas tecnologias desenvolvidas para o padrão Dublin Core e sua utilização. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 10, n. 1, p. 241-249, 2005.

POMERANTZ, Jeffrey. **Metadata**. Cambridge, MA: The MIT Press, 2015. (The MIT Press Essential Knowledge series).

SANTAREM SEGUNDO, José Eduardo; SILVA, Marcel Ferrante; MARTINS, Dalton Lopes. Revisitando a interoperabilidade no contexto dos acervos digitais. **Informação & Sociedade: Estudos**, João Pessoa, v. 29, n. 2, p. 61-84, abr./jun. 2019. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/38107>. Acesso em: 14 set. 2019.

SANTOS, Daiane Barrieli dos; ROCKEMBACH, Moisés. Publicações ampliadas: aspectos da integração de dados de pesquisa. **Informação & Sociedade: Estudos**, João Pessoa, v. 28, n. 2, p. 257-269, mai./ago. 2018. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/38558>. Acesso em: 07 jan. 2021.

SAYÃO, Luís Fernando. Padrões para bibliotecas digitais abertas e interoperáveis. **Encontros Bibli**, Florianópolis, n. esp., p. 18-47, 2007. <https://doi.org/10.5007/1518-2924.2007v12nesp1p18>

SAYÃO, Luís Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **TransInformação**, Campinas, v. 20 n. 2, p. 133-148, maio/ago. 2008. <https://doi.org/10.1590/S0103-37862008000200002>

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. 23. ed. São Paulo: Cortez, 2007.

SILVA, Renata Eleutério da. **As tecnologias da web semântica no domínio bibliográfico**. 2013. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013.

SOUZA, Joel de; VIERA, Angel Freddy Godoy. A tecnologia Open Archives Initiative, Object Reuse and Exchange: histórico e perspectivas. **DataGramZero**, Rio de Janeiro, v. 13, n. 2, artigo 04, abr. 2012.

SOUZA, Marcia Izabel Fugisawa; VENDRUSCULO, Laurimar Gonçalves; MELO, Geane Cristina. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, Brasília, DF, v. 29, n. 1, p. 93-102, jan./abr. 2000. <https://doi.org/10.1590/S0100-19652000000100010>

TARRANT, David; O'STEEN, Ben; BRODY, Tim; HITCHCOCK, Steve; JEFFERIES, Neil; CARR, Leslie. Using OAI-ORE to transform digital repositories into interoperable storage and services applications. **Code4Lib Journal**, [S. l.], n. 6, Mar. 2009. Disponível em: <https://journal.code4lib.org/articles/1062>. Acesso em: 20 jun. 2020.

WITT, Michael. Object Reuse and Exchange (OAI-ORE). **Library Technology Reports**, Chicago, IL, v. 46, n. 4, May/June 2010.

ZENG, Marcia Lei; QIN, Jian. **Metadata**. New York, NY: Neal-Schuman, 2008.