

Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa

Fernanda Alves Sanchez

Universidade Estadual Paulista “Júlio Mesquita Filho” – UNESP, Email: feersanchez@gmail.com

Nathalia Britto Pinheiro da Silva

Universidade Estadual Paulista “Júlio Mesquita Filho” – UNESP, Email: nathaliabritto6@gmail.com

Fernando Luiz Vechiato

Universidade Estadual Paulista “Júlio Mesquita Filho” – UNESP; Universidade Federal do Rio Grande do Norte - UFRN, Email: vechiato2008@gmail.com

Resumo

Os padrões de metadados possibilitam que os dados de pesquisas possam ser descritos, obtendo informações de sua proveniência. Objetivou-se identificar os padrões de metadados mais utilizados mundialmente para a representação de dados de pesquisa. A pesquisa documental e exploratória de abordagem qualitativa, utilizou como instrumento metodológico o diretório *Registry of Research Data Repositories - Re3data*, selecionando os três padrões de metadados mais utilizados pelos repositórios de dados de pesquisa, sendo eles: *Dublin Core (DC)*, *Data Documentation Initiative (DDI)* e *ISO 19115 - Geografic information - Metadada.*, Data. O diretório contribuiu ainda para uma escolha de três repositórios que fazem uso dos padrões de metadados. Verificou-se que os padrões de metadados representam seus dados e as informações, de modo que auxiliam na veracidade das informações sobre um determinado dado de pesquisa representado, bem como permite uma descrição, assim tornando-se no formato de dados e informações armazenadas nos repositórios de dados de pesquisa que potencializam a uso, reuso e compartilhamento.

Palavras-chave: Metadados. Repositório de Dados de Pesquisa. Padrão de Metadados.

1 INTRODUÇÃO

A crescente geração de dados de pesquisa vem tornando o desafio da Ciência da Informação de representar e organizar informação cada vez maiores, no que condiz a eficiência e eficácia do ambiente informacional digital e sucesso para os sujeitos informacionais em relação a recuperação e encontrabilidade das informações desejadas.

Nesse sentido, os Repositórios de Dados de Pesquisa vêm ganhando espaço no cenário acadêmico-científico, estimulando e potencializando a cultura de partilha de dados, para que assim, evite-se a duplicação de dados em pesquisas relacionadas ou de mesma temática, além

Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa de instigar novas hipóteses de pesquisa.

Desse modo, com a grande quantidade de dados e informações gerados pelas instituições de ensino, pesquisa e fomento os ambientes que propiciam o armazenamento, disseminação, visibilidade e preservação das produções intelectuais e os dados de pesquisa se torna relevante para a comunidade.

Nesse contexto, estabelecer padrões de metadados para que a representação desses dados e informações sejam adequadas é de suma importância para os Repositórios de Dados de Pesquisa.

Sendo assim, nosso objetivo foi identificar e descrever os padrões de metadados mais utilizados mundialmente para a representação dos dados de pesquisa armazenados nos repositórios dessa característica.

Para atingir o objetivo, foi utilizado como instrumento metodológico o diretório *Registry of Research Data Repositories - Re3data*, em especial, a métrica que apresenta os padrões de metadados mais utilizados por Repositórios de Dados de Pesquisa.

Além disso, o diretório auxiliou na escolha de três repositórios que utilizam os padrões de metadados apresentados nesse texto, sendo possível observá-los como aplicação. Ainda sobre a metodologia, a pesquisa é caracterizada por documental e exploratória de abordagem qualitativa.

Essa ferramenta é importante por identificar quais os repositórios mais apropriados para abrigar os dados, além disso, armazenam informações sobre repositórios em operação, permitindo que sejam localizados facilmente pelas comunidades, potencializando a visibilidade dos resultados de pesquisa (SAYÃO; SALES, 2015).

A escolha dos repositórios foi feita pelo filtro de busca, denominado “*Metadata standards*”, sendo selecionados: *Environmental Information Data Centre (EIDC)*¹, *Indepth Data Repository*² e *Gulf of Mexico Research Initiative Information and Data Cooperative*³.

Como dito anteriormente a seção de métricas contribuiu para a seleção dos três padrões de metadados mais utilizados por repositórios de dados de pesquisa, sendo eles: *Dublin Core (DC)*, *Data Documentation Initiative (DDI)* e *ISO 19115 - Geografic information - Metadada* apontados pelo diretório.

1 <http://eidc.ceh.ac.uk/>

2 <http://www.indepth-ishare.org/index.php/home>

3 <https://data.gulfresearchinitiative.org/>

2. METADADOS

Os metadados potencializam os recursos informacionais para que possam ser localizados a partir da busca realizada pelo sujeito informacional, principalmente por meio do mecanismo de busca de um ambiente informacional digital.

Desse modo, Castro e Santos (2014) expõem que o ambiente informacional digital pode recuperar a informação por meio de metadados, reconhecendo pontos que enfatizam o acesso aos sujeitos informacionais, sendo eles: a sintaxe (a combinação das palavras ou gramática para comunicar semântica e estrutura); a estrutura (que pode ser lida por humanos e pode ser validada por meio das máquinas); o vocabulário (que seria a correspondência entre os símbolos e os conceitos) e a semântica (o significado dos elementos ou do vocabulário).

De acordo com Gilliland (2016, tradução nossa), a definição de que metadados são “dados sobre dados” é amplamente utilizada, no entanto, os metadados são utilizados em diversos contextos, dessa forma, a compressão sobre metadados em diversas comunidades mostra-se diferente na medida que protege, cria, descreve, preserva e usa sistemas e recursos de informação.

Alves (2010), compreende que os metadados são atributos que representam um objeto no mundo real em um sistema de informação, por sua vez, favorecendo que a descrição sobre o objeto possibilite que seja identificado de forma única. Desse modo, no contexto da Ciência da Informação, proporciona diversas formas de acesso ao recurso informacional “[...] com a finalidade de busca, localização, recuperação, intercâmbio de dados e interoperabilidade entre sistemas.” (ALVES, 2010, p.52)

Segundo Zeng e Qin (2008, tradução nossa), os metadados descrevem qualquer documento ou objeto em formatos digitais e tradicionais. O formato ou o tipo de documento pode ser descrito por metadados, destaca-se que para Gilliland (2016) a descrição é independente da condição física ou intelectual, possuindo as seguintes características: conteúdo; contexto e estrutura.

Dessa forma, Alves (2010, p. 48, grifo nosso) esclarece em relação aos metadados que:

[...] o termo metadados está intimamente associado com a definição de padrões de metadados. Para que eles possam existir, os **metadados** (*metadata*) devem estar codificados em estruturas padronizadas de descrição, denominadas como **padrões de metadados** (*metadata statement*). O conjunto de metadados ou **elementos de metadados** (*element sets*) irá compor o

Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa

esquema de metadados (*metadata schema*) do padrão de metadados (ALVES, 2010, p. 48, grifo nosso).

Os metadados referem-se à capacidade de troca de dados entre ambientes informacionais digitais diferentes, de modo que os recursos informacionais possam ser interpretados de maneira consistente mesmo que tenham padrões heterogêneos. (CASTRO; SANTOS, 2014)

Os metadados representam todas as informações de um objeto, por exemplo: o título, as descrições textuais, as classificações de assunto para o acesso e as relações contextuais entre o objeto a ser descrito. Neste sentido, Riley (2017, p. 6, tradução nossa) descreve que existem quatro tipos de metadados, os:

- Descritivos (*Descriptive metadata*) - são para encontrar e entender um recurso;
- Administrativos (*Administrative metadata*) – são aqueles referentes aos técnicos, preservação e de direitos autorais e intelectuais do conteúdo;
- Estruturais (*Structural metadata*) - são os relacionamentos entre as partes de recursos entre si;
- Linguagens de Marcação (*Marketing languages*) - são a integração de metadados e sinalizadores com outros recursos estruturais e semânticos do conteúdo.

Neste contexto, os metadados descritivos possibilitam que os recursos possam ser encontrados por meio de mecanismos de busca, bem como compreender se aquele recurso atende ou não as expectativas do sujeito. Referente aos metadados administrativos, são aqueles que fornecem informações de sua proveniência, direitos autorais e intelectuais que um determinado recurso possua. Os metadados estruturais permitem correlação de recursos. Assim, as linguagens de marcação referem-se a atribuir semântica as informações contidas no metadados.

Os metadados propiciam uma representação descritiva das informações de um recurso informacional e/ou objetos digitais. De acordo com Woodley (2016, tradução nossa) os padrões de metadados são reflexos da funcionalidade da informação e do conhecimento que ficam armazenados e são expressados no processamento em sistemas e nos mecanismos de busca, possibilitando a descoberta da informação por meio dos mecanismos de busca.

3 REPOSITÓRIO DE DADOS DE PESQUISA

A massiva geração de dados e informações produzidos no cenário acadêmico-científico impulsionam a criação de ambientes informacionais digitais como os Repositórios de Dados de Pesquisa que se trata de uma “[...] ferramenta necessária para armazenar e gerenciar os dados utilizados e produzidos durante uma pesquisa” (VIDOTTI et al, 2017, p. 8).

Tais ambientes são apropriados para que os dados de pesquisas sejam preservados, recuperados, acessados e citados por outros pesquisadores, atingindo escala mundial em relação à visibilidade de pesquisas enriquecidas por estudos comparativos e com análises mais sofisticadas (SAYÃO, SALES, 2015; KIM, ZHANG, 2015).

Nesse sentido, Sayão e Sales (2015) algumas vantagens referentes à utilização dos Repositórios de Dados de Pesquisa a fim de consolidar e a cultura de partilha de dados entre a comunidade científica. Abaixo, apresenta-se no quadro 1, o que Sayão e Sales apresentam em seu texto como vantagens.

Quadro 1: Vantagens em utilizar os repositórios de dados de pesquisa

Vantagem	Descrição
Qualidade	Assegura que os dados estejam em conformidade com os padrões de qualidade
Preservação Digital	Garante a preservação digital de longo prazo dos dados
Segurança	Mantém os dados em ambiente seguro
Backup	Providência backups regulares dos dados
Descoberta	Viabiliza a descoberta online dos dados via catálogos de dados
Formatos Conhecidos	Acesso aos dados em formatos populares
Direitos	Mantém acordo de licenciamento que reconhece a autoria dos dados
Citação	Disponibiliza mecanismos de citação que reconhece a autoria dos dados
Reuso	Promove o uso e reuso dos dados para muitos usuários
Gestão de Acesso	Gerencia o acesso aos dados e as consultas dos usuários em nome do proprietário

Fonte: Adaptada (SAYÃO; SALES, 2015, p.55).

Dada as vantagens em utilizar um ambiente informacional digital com essas características é fundamental que os objetos e dados digitais armazenados sejam recuperados permitindo a interoperabilidade com sistemas de informação, para isso, se faz necessário que a

Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa descrição seja realizada por meio de metadados que sejam adequados para sua representação, além de estabelecer padrões que facilitem e impulsionam o acesso, uso, recuperação e encontrabilidade da informação. (VIDOTTI et al, 2017, p.9).

Os metadados em Repositórios de Dados de Pesquisa, são importantes no que diz respeito a facilitar a interoperabilidade entre sistemas de informação, além disso, os dados armazenados nesses ambientes podem ser [...] combinados, recombinaados e usados por áreas multidisciplinares, ser unificados, integrados e interoperados em rede” (MONTEIRO; SANT’ANA, 2016, p. 655).

Nesse sentido, independente dos diferentes padrões existentes de metadados, sendo eles, direcionados a descrição de dados de pesquisa de modo geral ou de determinada comunidade científica, como, por exemplo, o *Ecological Metadata Language* (EML) – que é direcionado a descrição de dados e objetos da ecologia (ciência da terra), é possível interoperar os metadados desde que as instituições de ensino, pesquisa e fomento utilizem protocolos adequados para tal processo.

A interoperabilidade é fundamental em ambientes informacionais dessas características, pois é responsável por garantir a capacidade de comunicação entre os repositórios, tornando “[...] possível a pesquisa em simultâneo com significados compartilhados nos vários repositórios, bem como a relação automática entre os resultados dessas pesquisas.” (BAPTISTA, 2010, p. 72).

A necessidade de estabelecer a interoperabilidade é proporcionar o entendimento entre sistemas, padrões, protocolos, formatos heterogêneos para que possam trocar informações e compartilhamento respeitando suas diferenças, de modo que, as informações sejam convertidas de acordo com suas características para que não se perca semântica e sintaxe. O conceito de interoperabilidade concentra-se na sua capacidade de:

[...] dois ou mais sistemas de interagir, de intercambiar e de compartilhar dados, entre si, pertinentes às suas necessidades. As informações devem estar organizadas de maneira eficiente para que essas interações aconteçam. (SILVA, 2013, p. 62)

Sendo assim, por meio do diretório *Re3Data* utilizamos a métrica que disponibiliza os padrões mais utilizados por Repositórios de Dados de Pesquisa espalhados pelo mundo, selecionando o top 3 do *ranking* a ser explorado nessa pesquisa, sendo eles: *Dublin Core (DC)*, *Data Documentation Initiative (DDI)* e *ISO 19115 - Geographic information - Metadata*.

Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa

Sendo assim, os Repositórios de Dados de Pesquisa, podem utilizar tipos de padrões de metadados diferentes para que possam atingir determinadas comunidades científicas específicas, ou seja, padrões que possam representar dados e informações em determinadas temáticas do conhecimento. A seguir, serão apresentados os resultados dessa pesquisa.

4 DISCUSSÃO E RESULTADOS

No *Registry of Research Data Repositories*¹ foi possível recuperar as métricas que identificam quais padrões de metadados estão sendo mais utilizados pelos repositórios de dados de pesquisa.

No tabelo 1, foi adaptado a partir dos dados apresentamos no Re3data apresenta-se na primeira coluna a posição em relação a quantidade de repositórios de dados de pesquisa que utilizam determinado padrão de metadados; na segunda coluna é exposto os nomes de cada padrão de metadados, na terceira coluna exposto a quantidade.

Tabela 1 -Padrões de metadados em repositórios de dados de pesquisa

POSIÇÃO	PADRÃO DE METADADOS	REPOSITÓRIOS DE DADOS
1	Dublin Core (DC)	173
2	Data Documentation Initiative (DDI)	116
3	ISO 19115	114
4	DataCite Metadata Schema	77
5	Federal Geographic Data Committee Standard For Digital Geospatial Metadata (FGDC/CSDGM)	68
6	Other	58
7	Repository – Developed Metadata Schemas	41
8	Directory Interchange Format	30
9	Climate and Forecast (CF) Metadata Conventions	27
10	Ecological Metadata Language (EML)	22
11	RDF Data Cube Vocabulary	19
12	Darwin Core	19
13	ISA-Tab	8
14	Flexible Image Transport System (FITS)	8
15	Open Archives Initiative – Object Reuse and Exchange (OAI-ORE)	7
16	Minimum Information for Biological and Biomedical Investigations (MIBBI)	7
17	Access to Biological Collection Data (ABCD)	7
18	Data Catalog Vocabulary (DCAT)	6
19	Crystallgraphic Information Framework	5
20	International Virtual Observatory Alliance Technical Specifications	4

Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa

21	Statiscal Data and Metadata Exchange (SDMX)	2
22	PROV	1
23	Genome Metadata	1
24	Core Scientific Metadata Model (CSDM-CCLRC)	1
25	Common Infomation Model (CIM)	1
26	Astronomy Visualization Metadata (AVM)	1

Fonte: Adaptada do Re3Data

Disponível: <https://www.re3data.org/metrics/metadataStandards>

Acesso em: 27 nov. 2017.

A partir dos resultados apresentadas na métrica no diretório Re3Data apresentados na tabela 1 acima, selecionou-se os três primeiros padrões de metadados mais utilizados. O primeiro padrão é o *Dublin Core*, são 173 repositórios de dados de pesquisa utilizando-o, seguido do *Data Documentation Initiative* (DDI) com 116 repositórios e terceiro mais utilizado é a ISO **19115** com 114 repositórios.

Definiu-se buscar um registro de dados de pesquisa entre os repositórios que utilizam os padrões de metadados supracitados. Com o filtro de busca do diretório, “*Metadata Standards*” para auxiliar na escolha. Sendo assim, elaborou-se os quadros 2, 3 e 4 com fragmentos de determinados registros, de modo que, foram descritas as estruturas dos metadados.

O *Dublin Core* surgiu em 1995 a partir de uma reunião ocorrida em Dublin - Ohio, que teve como enfoque identificar o conjunto básico de recursos comuns à maioria dos tipos de informações digitais. Essa reunião resultou na identificação dos 13 elementos principais, que em seguida, foi complementado com mais dois elementos. (RILEY, 2017)

Sendo assim, os 15 elementos do DC *são*: colaborador, cobertura, criador, data, descrição, formato, identificador, idioma, editor, relação, direitos, fonte, assunto, título e tipo. (RILEY, 2017). Esse padrão de metadados propicia que sejam criados mais elementos de acordo com os requisitos que cada objeto digital necessite para sua descrição.

No quadro 2, apresenta-se o padrão de metadados *Dublin Core* em um dos registros que diz respeito sobre dados mensais e diários do potencial de evapotranspiração⁴ no Reino Unido entre o período de 1891 a 2015 (TANGUY *et al.* 2017) do repositório *Environmental Information Data Centre (EIDC)* que contempla dados de pesquisa sobre ecologia e hidrologia.

Quadro 2 - Padrão de Metadados Dublin Core

4 Evaporação de água para a atmosfera ocasionando a conseqüente perda de água de um ecossistema. Dicionário Michaelis, 2015

```
dct:title "Historic Gridded Potential Evapotranspiration (PET) based on temperature-based equation  
McGuinness-Bordne calibrated for the UK (1891-2015)" ;  
dct:description "Monthly and daily 5km gridded Potential Evapotranspiration (PET) data for the  
UK. PET was derived using temperature-based equation from McGuinness-Bordne. The units are  
mm/day for daily PET and mm/month for monthly PET. The dataset covers the period from 1891-  
2015.  
For both subsets (daily and monthly), a set of performance metrics were calculated, which are  
provided together with the PET grids. The list of metrics provided is: Mean Absolute Percent Error  
(MAPE), Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), Correlation Coefficient,  
Variability Ratio (VR), Bias Ratio and monthly MAPE.  
";  
dct:spatial "POLYGON((-8.648393 49.86382, -8.648393 60.86099, 1.767549 60.86099, 1.767549  
49.86382, -8.648393 49.86382))"^^geo:wktLiteral ;  
dct:temporal "1891-01-01/2015-12-31"^^dct:PeriodOfTime ;-
```

Fonte: Extraído e Adaptado de (TANGUY *et al.* 2017)

Disponível em: <https://catalogue.ceh.ac.uk/documents/17b9c4f7-1c30-4b6f-b2fe-f7780159939c>.

Acesso em: 26 nov. 2017.

As informações descritas sobre os dados da pesquisa no fragmento exposto, possibilita verificar a presença dos metadados que indicam o *dct:title*: título, *dct:description*: descrição sobre o conteúdo, *dct:spatial*: especificações relacionadas a espaço, *dct:temporal*; delimitação de tempo de pesquisa.

Além disso, os metadados estão estruturados em um modelo de dados em *Resource Description Framework (RDF)*, sendo expressados com *Turtle - Terse RDF Triple Language*.

O *Data Documentation Initiative (DDI)*⁵ é um padrão de metadados composto por elementos das ciências sociais, comportamentais e econômicas. A estrutura sustenta o pesquisador na descrição e no ciclo de vida de dados, bem como no planejamento do gerenciamento de dados.

No quadro 3, apresenta-se o registro do repositório *Indepth Data Repository*⁶, esse refere-se à avaliação demográfica sobre saúde de populações de baixa e média renda (*LMICs*, na sigla em Inglês) (SANKOH *et al.* 2014).

Quadro 3 - Padrão de Metadados DDI

5 <https://www.ddialliance.org/>

6 *Demographic Evaluation of Populations and Their Health (INDEPTH)*. Disponível em: <<http://www.indepth-ishare.org/index.php/home>>. Acesso em: 28 nov. 2017.

```
<?xml version="1.0" encoding="UTF-8"?>
<codeBook xsi:schemaLocation="http://www.icpsr.umich.edu/DDI
http://www.icpsr.umich.edu/DDI/Version1-2-2.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.icpsr.umich.edu/DDI" xml-lang="en"
ID="INDEPTH.GH003.COD2014.v1" version="1.2.2"><docDscr><citation><titlStmt><titl>
INDEPTH Network Cause-Specific Mortality - Release 2014 </titl><IDNo>
DDI.INDEPTH.GH003.COD2014.v2 </IDNo></titlStmt><prodStmt><producer role="Study
documentation" affiliation="INDEPTH Network" abbr="KHe"> Kobus Herbst
</producer><producer role="Study documentation" affiliation="INDEPTH Network"
abbr="PBy"> Peter Byass </producer><prodDate date="2014-08-13"> 2014-08-13
</prodDate><software version="4.0.9" date="2013-04-23"> Nesstar Publisher
</software></prodStmt><verStmt>
```

Fonte: Extraído e adaptado de (SANKOH *et al.* 2014).

Disponível em: <http://www.indepth-ishare.org/index.php/catalog/48>

Acesso em: 27 out. 2017.

No quadro 3, o registro aponta descrição de uma pesquisa que obteve dados sobre a causas específicas de mortalidade nos continentes da África, Ásia e Oceania (SANKOH *et al.* 2014).

Nesse fragmento, obtivemos resultados quanto ao “*titl*” (título) atribuído a pesquisa, bem como um “*IDNo*” (número identificação), a “*prodDate*” (data de publicação) e outros, os metadados encontram-se estruturados em estruturado em *EXtensible Markup Language* (XML).

O DDI auxilia no processo de pesquisa como coleta, normalização, análise, compartilhamento e arquivamento dos dados. Possui esquema em XML, possibilitando que os dados possam ser vinculados com outros dados. (RILEY, 2017, tradução nossa)

O padrão de metadados ISO 19115 - *Geografic information - Metadada* contém conjuntos de elementos de metadados que descrevem informações geográficas, podendo ser aplicadas em dados e recursos como: mapas, gráficos, documentos textuais, possuindo metadados obrigatórios e adicionais.

Essa norma foi publicada em 2003, sendo atualizada e substituída pela ISO 19115:2014, na primeira publicação teve como objetivo contemplar metadados sobre a “[...] identificação, extensão, qualidade, o esquema espacial e temporal, a referência espacial e a distribuição de dados geográficos digitais.” (ISO, 2003, tradução nossa).

Na ISO 19115:2014, além dos elementos citados acima, contempla informações referentes aos “[...] aspectos espaciais e temporais, o conteúdo, a representação, e outras propriedades dos dados e serviços geográficos digitais.” (ISO, 2014, tradução nossa). Esse

Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa padrão é utilizado em *datasets* que lidam com dados geográficos.

No quadro 4, apresenta o padrão de metadados da ISO 19115 de um registro sobre “[...] conjunto de dados laboratoriais de degradação de dispersante de derramamento de óleo” (CHOYKE; FERGUSON, 2017, tradução nossa) disponibilizado no repositório de dados de pesquisa *Gulf of Mexico Research Initiative Information and Data Cooperative*.

Quadro 4 - Padrão de Metadados ISO 19115

Fonte: *Gulf of Mexico Research Initiative Information and Data Cooperative*.

```
<gmd:fileIdentifier>
  <gco:CharacterString>R2.x218.000-0004-metadata.xml</gco:CharacterString>
</gmd:fileIdentifier>
<gmd:language>
  <gco:CharacterString>eng; USA</gco:CharacterString>
</gmd:language>
<gmd:characterSet>
  <gmd:MD_CharacterSetCode codeList="http://www.isotc211.org/2005/resources/Codelist/g
mxCodelists.xml#MD_CharacterSetCode" codeListValue="utf8" codeSpace="004">utf8</gmd:
MD_CharacterSetCode>
</gmd:characterSet>
<gmd:hierarchyLevel>
  <gmd:MD_ScopeCode codeList="http://www.isotc211.org/2005/resources/Codelist/gmxCo
delists.xml#MD_ScopeCode" codeListValue="dataset" codeSpace="005">dataset
</gmd:MD_ScopeCode>
```

Disponível em: <https://data.gulfresearchinitiative.org/data/R2.x218.000:0004#identificationInfo>

Acesso em: 26 out. 2017.

No registro supracitado, composto por metadados com a identificação do arquivo “*fileIdentifier*”, a linguagem do conteúdo “*language*”, possuindo também *namespace* e estrutura do XML.

Como apresentado nas figuras anteriores, os repositórios de dados de pesquisa utilizam padrões de metadados, para descrever, administrar e apresentar a proveniência dos dados de pesquisa, além de permitir a interoperabilidade entre sistemas diversos.

As informações que os padrões de metadados oferecem são de extrema relevância para o sujeito informacional. Percebe-se que cada padrão dispõe de metadados, por exemplo, para assuntos e palavras-chaves que permitem aos mecanismos de buscas recuperarem os dados de pesquisas e objetos referentes a eles a partir do seu contexto.

Neste sentido, a descrição dos dados de pesquisa por meio dos metadados leva o sujeito a compreender se tal objeto é determinante ao seu contexto ou não, possibilitando assim, o reuso e compartilhamento entre a comunidade acadêmico-científico.

Os padrões de metadados nos processos informacionais no que tange as representações afetam a organização do ambiente Informacional digital onde será disponibilizado os dados de pesquisa. A representação refere-se ao fato de que os dados podem e devem ser descritos da melhor forma possível, por meio dos metadados, quanto a organização os metadados referem-se à apresentação dos dados de pesquisa no ambiente informacional digital.

5 CONSIDERAÇÕES FINAIS

Nota-se que os padrões de metadados podem auxiliar na representação e organização de qualquer ambiente informacional digital. Tratando-se de repositórios de dados de pesquisa a descrição dos objetos digitais e dos dados que compõem a pesquisa são determinantes para a recuperação e encontrabilidade de cada um deles.

Nesse sentido, a representação de cada elemento permite que o sujeito informacional possa fazer uso desses dados, sendo possível o reaproveitamento de dados, como por exemplo, de coletas extensivas com sujeito informacionais, dados referentes ao meio ambiente e etc. Contribuindo assim, para amenizar a duplicação de dados dentro da comunidade acadêmico-científica, além da possibilidade de novas hipóteses de pesquisa a partir desse reuso e compartilhamento.

Sobre os diferentes tipos de metadados, podemos observar que atualmente o *Dublin Core* é o mais utilizado, possivelmente por conta do *software* de desenvolvimento e implementação dos Repositórios Digitais, que é o *Dspace* - no diretório o *software* é considerado como o mais utilizado com 44,2%.

Nota-se que há veracidade da informação por conta das diferentes comunidades científicas existentes e que essas necessitam de metadados adequados que representam seus dados e informações, como são o caso dos padrões destacados aqui, específicos para dados geográficos (ISO) ou ecologia (EML).

Por fim, os metadados são relevantes aos Repositórios de Dados de Pesquisa de modo que permitem que a descrição seja realizada por meio de padrões diversos de metadados e além disso, que seja adequada aos tipos de dados e informações armazenadas neles para potencializar a uso, reuso, compartilhamento. Os metadados podem potencializar a interoperabilidade entres diferentes sistemas de informação, e se tratando do uso e reuso desses dados, os metadados vão permitir referenciar todos os processos de uma pesquisa.

Metadata standards for representation and organization of information in research data repositories

ABSTRACT

Metadata standards allow research data to be queried and described, gathering provenance information. The aim of this study was to identify and describe the most widely used metadata patterns worldwide for the representation of research data. A documentary and exploratory research with a qualitative approach, using a methodological instrument of the Directory of Repositories of Research Data - Re3data, selecting the three metadata patterns most used by the research data repositories, namely: Dublin Core (DC), Initiative of Data Documentation (DDI) and ISO 19115 - Geographic Information - Metadata, Data. The directory also contributed to a choice of three repositories that make use of metadata standards. It has been found that the metadata patterns represent their data and information, so that they aid in the veracity of the information on a given research data represented, as well as allows a description, thus becoming the format of data and information stored in the repositories of research data that leverages the use, reuse and sharing.

Keywords: Metadata. Search Data Repository. Metadata Standard.

REFERÊNCIAS

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2010. Disponível em: <<http://hdl.handle.net/11449/103361>>. Acesso em: 28 nov. 2017.

BAPTISTA, A. A. A falar nos entendemos: a interoperabilidade entre repositórios digitais. **Repositórios institucionais: democratizando o acesso ao conhecimento**. Salvador: EDUFBA, p. 71-90, 2010. Disponível em: <<http://bit.ly/2jO5wEE>>. Acesso em: 29 nov. 2017.

CASTRO, F. F.; SANTOS, P. L. V. A. C. Elementos de interoperabilidade na perspectiva da catalogação descritiva. *Informação & Sociedade*. João Pessoa, v. 24, n. 3, p. 13-25, set/dez. 2014. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/16660>> Acesso em: 27 nov. 2017.

CHOYKE, S; FERGUSON, P. L. **COREXIT 9500 biodegradation in seawater, experimental data**: Characterizing the Composition and Biogeochemical Behavior of Dispersants and Their Transformation Products in Gulf of Mexico Coastal Ecosystems. 2017. DOI: 10.7266/N7DJ5D2D. Disponível em: <<https://data.gulfresearchinitiative.org/data/R2.x218.000:0004#metadataMaintenanc>>. Acesso em: 28 nov. 2017.

Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **19115: 2003**: Geographic information -- Metadata. [s.i]: ISO, 2003. Disponível em: <<https://www.iso.org/standard/26020.html>>. Acesso em 28 out. 2017.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **19115: 2014**: Geographic information -- Metadata. [s.i]: ISO, 2014. Disponível em: <<https://www.iso.org/standard/53798.html>>. Acesso em 28 out. 2017.

GILLILAND, A. J. Setting the Stage. In: BACA, Murtha (Ed.). **Introduction to Metadata**. 3. ed. Los Angeles: Getty Research Institute, 2016. Cap. 5. p. 8-25. Disponível em: <<http://www.getty.edu/publications/intrometadata/setting-the-stage/>>. Acesso em: 27 nov. 2017

KIM, Y; ZHANG, P. Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. **Library & Information Science Research**, v. 37, n. 3, p. 189-200, 2015. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0740818815000584>>. Acesso em: 23 out. 2017.

MONTEIRO, E. C S A; SANT'ANA, R. C. G. Repositórios de Dados Científicos nas Universidades Brasileiras e Portuguesas. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 6., 2016, Londrina. **Anais...** Londrina: UEL. ISBN: 978-85-7846-383-0. 2016. p.166-174. Disponível em: <<http://www.uel.br/eventos/cinf/index.php/secin2016/secin2016/paper/viewFile/338/166>>. Acesso em: 27 nov. 2017.

RILEY, J. **UNDERSTANDING METADATA WHAT IS METADATA, AND WHAT IS IT FOR?** Baltimore: National Information Standards Organization (niso), 2017. 45 p. Disponível em: <http://www.niso.org/apps/group_public/download.php/17446/Understanding_Metadata.pdf>. Acesso em: 26 out. 2017.

SANKOH, O. A. et al. **INDEPTH Network Cause-Specific Mortality - Release 2014**. 2014. Dados de pesquisa. Disponível em: <<http://www.indepth-ishare.org/index.php/catalog/48/study-description>>. Acesso em: 28 nov. 2017.

SAYÃO, L. F; SALES, L. F. **Guia de Gestão de Dados de Pesquisa para Bibliotecários e Pesquisadores**. Rio de Janeiro: CNEN/IEN, 2015. Disponível em: <http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf>. Acesso em: 23 out. 2017.

SILVA, R. E. **As tecnologias da Web Semântica no domínio bibliográfico**. 2013. 134 f. **Dissertação** (mestrado) - Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciências de Marília, 2013. Disponível em: <<http://hdl.handle.net/11449/93653>>. Acesso em 20 set 2016

TANGUY, M.; PRUDHOMME, C.; SMITH, K.; HANNAFORD, J. (2017). Historic Gridded Potential Evapotranspiration (PET) based on temperature-based equation McGuinness-Bordne

Padrões de metadados para representação e organização da informação em repositórios de dados de pesquisa

calibrated for the UK (1891-2015). NERC Environmental Information Data Centre. <https://doi.org/10.5285/17b9c4f7-1c30-4b6f-b2fe-f7780159939c>. Disponível em: <<https://catalogue.ceh.ac.uk/documents/17b9c4f7-1c30-4b6f-b2fe-f7780159939c>>. Acesso em: 28 nov. 2017.

VIDOTTI, S. A. B. G et al. REPOSITÓRIO DE DADOS DE PESQUISA PARA GRUPO DE PESQUISA: um projeto piloto. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília - SP. **Anais...** Marília - SP: PPGCI, UNESP. Disponível em: <<http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/view/388>>. Acesso em: 23 out. 2017.

WOODLEY, M.S. Metadata Matters: Connecting People and Information.” Revised by Murtha Baca. In: BACA, M. (Ed.). **Introduction to Metadata**. 3. ed. Los Angeles: Getty Research Institute, 2016. Cap. 7. p.38 - 53. Disponível em: <http://www.getty.edu/publications/intrometadata/metadata-matters/>. Acesso em: 27 nov. 2017

ZENG, M. L.; QIN, J. **Metadata**. New York: Neal-schuman Publishers, 2008. 365 p.